Data Management Plan

For California's Statewide MPA Monitoring Program

CALIFORNIA
OCEAN
SCIENCE
TRUST

Version 1: March 2016

About this document

California Ocean Science Trust stewards data from State funded marine protected area (MPA) monitoring. As the steward, Ocean Science Trust is charged with ensuring the open-access, durability, and longevity of public trust data. In fulfillment of that mandate, this data management and stewardship plan for California's MPA monitoring data was developed in collaboration with State, federal, and academic partners and provides a clear analysis of needs, requirements, and recommendations for a pragmatic solution. To assist Ocean Science Trust in the creation of this plan, a data management advisory team was formed, comprised of members with expertise in scientific monitoring, data and metadata management, information systems and storage, data informatics and semantics, and web portals for data distribution.

This data management plan:

- Recommends partnerships and collaborations for long-term data storage that effectively engage the technological and data expertise and capacity within the University of California system as well as standards based data-initiatives, such as DataONE[1];

- Identifies the high-level data architecture including where to store data, in what format, and how data will be integrated into the complex data management landscape to support analysis and synthesis, drawing on the experience and potential for partnership with the California Department of Fish and Wildlife (CDFW), Sanctuary Integrated Monitoring Network (SIMoN), Central & Northern California Ocean Observing System (CeNCOOS), and Partnerships for Interdisciplinary Studies of Coastal Oceans (PISCO)[2];

- Evaluates the range of potential data sources (e.g., state-funded, citizen-generated) and recommend any necessary additional approaches for data stewardship and management that align with these different sources;

- Identifies the role of OceanSpaces.org in the data cycle, and articulates productive relationships between OceanSpaces and complementary online data assets such as the California Geoportal, West Coast Ocean Data Portal, and My Water Quality portal[3];

- Identifies key data and metadata standards to fully describe, build consistency into, and promote durability of monitoring data; and

---

[1] DataONE is a community driven project providing access to data across multiple member repositories, supporting enhanced search and
[2] Partnerships for Interdisciplinary Studies of Coastal Oceans (PISCO) is a long-term ecosystem research and monitoring program. http://www.piscoweb.org/
[3] California Geoportal (http://portal.gis.ca.gov/geoportal/catalog/main/home.page); West Coast Ocean Data Portal (http://portal.westcoastoceans.org/); California My Water Quality Portal (http://www.mywaterquality.ca.gov/)

- Provides a holistic estimate of the expertise, capacity and resources needed to implement and maintain the recommendations in the plan.

This document presents the findings, recommendations, and expert analysis of the needs for stewarding open-access public trust MPA monitoring data. This document serves as a high-level planning guide for the digital systems required to ingest, store, discover, and distribute these data. This document will be appended to the Statewide MPA Monitoring Program Guidance document. An implementation plan will be created based on these recommendations to address the technical requirements, use case development, data workflows, and timelines for the implementation of the data management system. Ocean Science Trust generated the following recommendations based on input received throughout the plan development process from the Advisory Team (see below), Workshop Participants, and additional external partners.

Advisory Team

- Tanya Haddad, West Coast Governor's Alliance; Oregon Department of Land Conservation and Development
- Dr. Matt Jones, National Center for Ecological Analysis & Synthesis; University of California Santa Barbara
- Dr. Steve Lonhart, Monterey Bay National Marine Sanctuary; National Oceanic and Atmospheric Administration
- Jennifer Patterson, Central & Northern California Ocean Observing System
- Paulo Serpa, California Department of Fish & Wildlife

Executive Summary

Purpose

This document serves as a high-level planning guide that brings together guidance and recommendations from experts in MPA monitoring and data management. The recommendations within this document seek to address the needs of the State of California to improve the return on investment of MPA monitoring data by improving the discoverability, relevance, and usability of MPA monitoring data. By implementing a new data management system, we will ensure public trust MPA monitoring data are housed within a durable and safe technical infrastructure, in addition to being provided to the State.

Assumptions

The data management plan:

- Does not address specific implementation requirements, such as technology use cases, workflow specifications, or hardware and software requirements. An implementation plan will be created to finalize the technical details once this document is presented to the Ocean Protection Council (OPC) and the California Division of Fish and Wildlife (CDFW).
- Presents an estimated operating budget of one-time and on-going costs for which funding is not yet fully committed.
- Relies on the continued funding of the OceanSpaces program within the Ocean Science Trust. During this planning process the State emphasized the importance of OceanSpaces in providing a rich engagement experience for MPA monitoring data (see budget requirements). OceanSpaces currently satisfies the State's need for a simple data catalogue for MPA monitoring data. The State would like to enhance the current system to improve the discoverability, relevance, and usability of MPA monitoring data.
- Relies on the longevity and availability of the Knowledge Network for Biocomplexity (herein referred to as KNB) data system.
- Budgets presented in this plan present the total cost of ownership for a data management system and include costs associated with relevant, existing systems. Assumes multiple sources of funding will be sought and secured to implement this plan.

Recommendations for Minimum Viable Product

The data management infrastructure described herein includes recommendations for a minimum viable product to meet the State's needs of making State-funded MPA monitoring data and results publicly accessible, including drawing connections to other management-relevant data sources. These recommendations were informed by the Advisory Team, Workshop Participants, and state information needs. The data systems chosen for implementation of this plan should be based on open-source technology standards. OceanSpaces will act as the user-facing primary access point. The KNB data management system will be the primary repository (backend) for MPA monitoring data. KNB is an international repository that utilizes rich metadata to facilitate the sharing and integration of ecological and environmental research. Using this repository builds on an existing open source technology, and thus satisfies the State's requirements for making MPA monitoring data publicly available in a cost-effective manner and leverages a system that is already widely used by the academic community and linked with DataONE. Data discovery options will include, thematic and keyword search, coupled with an interactive map featuring spatial data selection tools within the OceanSpaces framework. This document provides a set of recommendations based on the current working knowledge around data management and State needs. This document will remain a living document requiring review and adaption based on the future needs of MPA monitoring data management.

Table of Contents

Introduction

Situation Analysis, Needs Assessment, Justification

Long-term secure stewardship and management of monitoring data are often a central challenge for organizations overseeing monitoring programs. Here, stewardship encompasses data storage, including technology infrastructure, provenance, description, and dissemination. Over the course of the California's Baseline Marine Protected Area (MPA) Monitoring Program, and in collaboration with our state partners, Ocean Science Trust has adopted a minimalist approach to data stewardship and management, on the basis that experience and learning would illuminate where and how to invest – strategically and efficiently – in long-term data management capacities.

OceanSpaces.org serves as the primary platform for baseline data storage and dissemination. However, Ocean Science Trust adopted a basic approach to fulfill public accessibility requirements of bond-funded monitoring projects. Thus, data files are accepted in their native formats and stored as independent packages. Ocean Science Trust did not invest in an integrated database structure because to do so would have incurred expense prior to full knowledge of technical and user requirements. During this initial phase we prioritized the durability of data files – ensuring data can be understood and used in the future. Adopting a model used by large research consortia[4], we developed data and metadata standards to require rigorous description of submitted data packages. These serve as a basis for statewide standards for the transition from baseline to long-term monitoring data.

To handle long-term monitoring in of California's MPA network, we need to develop a comprehensive and scalable strategy for data management. A number of factors are likely to change over time, including the diversity of data types, the requirements for database integration for synthetic analyses, a complex suite of data accessibility and dissemination requirements linked to diverse funding sources, among others. We have gained experience managing a wide variety of data types and formats, including data products, and have a better understanding of State requirements and user needs for a data management system. Moreover we can learn from and utilize the available technology and data-driven advances in California encompassing data catalogues and portals, to distributed data networks with remote and cloud storage, and integrated databases with linked web visualizations.

Defining Data

This document covers the management of data, information derived from data, and imagery assets (see below). To represent this holistic approach, we will refer to data using a range of levels that span raw measurements to synthesized products.

---

[4] For example, Long Term Ecological Research Network (LTER), Partnerships for Interdisciplinary Studies of Coastal Oceans (PISCO), Multi-Agency Rocky Intertidal Network (MARINe)

- Level 1: Unprocessed field measurements, such as ecological data, fisheries landings data, and environmental data that are time-referenced, and annotated with ancillary information and georeferencing parameters.

- Level 2: Derived variables at the same resolution as the source Level 1 data (e.g., salinity from conductivity)

- Level 3: Variables that have been aggregated and are displayed or represented on a uniform scale (e.g., temporally or spatially binned data, mapped datasets, x-y plots, histograms)

- Level 4: Results from analyses (e.g., of lower level data, from within reports) or model outputs (e.g., variables derived from multiple measurements, ANOVA tables, plots showing means and indicating which are significantly different)

Data archiving, for the purposes of this document, means data files are uploaded, stored with a Digital Object Identifier (DOI), and available to download through the web interface. Archiving requirements for most MPA monitoring projects include Level 1 and Level 2 data. For data governed by confidentiality and non-disclosure agreements data archiving requirements also include Level 3 and/or Level 4. However, data from all projects will contribute to at least one report that includes Level 4 data.

Imagery

Imagery assets are not included in the four data levels because they are not data *per se*. Imagery assets, such as photographs and videos, require processing to yield data. During the MPA baseline monitoring phase, imagery assets for MPA monitoring projects were archived by those who generated them. For long-term monitoring, submission and archiving requirements will change – see Data Discovery and Integration section.

Defining Interoperability

For the purposes of this document, interoperability will is defined via two terms:

- Syntactic: A system-to-system handshake for understanding numbers, usually representing communications between digital systems, such as servers and databases.

- Semantic: Systems to distinguish among meanings, which include, but are not limited to, defining the terms in data tables, site lists, and monitoring regions. Semantic taxonomies are designed to eliminate confusion based on using a set vocabulary.

Interoperability between data storage systems will allow us to automate, share, and integrate datasets, while at the same time reducing duplicative effort, the resources required to integrate datasets, and streamlining the QA/QC process.

Considering User Needs – User Stories

Based on the need to create a system that provides maximum benefit with limited resources and with guidance from our Advisory Team and Workshop Participants (see Planning Process section), we plan to initially focus on the following two personas that represent our primary users. These user groups were defined and their needs were assessed as part of the *MPA Monitoring User Needs Assessment (Appendix B).*

- Resource Managers[1] & Decision Makers[2]: *Provide these users multiple data discovery pathways that characterizes the what, when, where, and by whom*

  This user group is most concerned with understanding the what, when, where, and by whom with respect to management and policy decisions. To support these users, the system will maximize discoverability of data and synthesis products through an interactive map and thematic, temporal, and taxonomic filters. The system will also draw clear connections between among data products (e.g., summary reports, visualizations), and raw the data, and technical reports. Demonstrating these connections, as well as the standards and QA/QC protocols, ensures that the system communicates the foundation of rigorous science behind data products. The system will point to additional relevant data and identify the appropriate next steps to learn more about MPA monitoring results (e.g., contact information for data contributor, link to related data portals for advanced visualizations).

- Data Contributors[3] & Users[4]: *Provide these users an efficient and easy pathway to understand the data upload process, and provide access to supplemental and ancillary data in one location*

  The system will provide a rich data engagement experience to link together raw data, technical reports, synthesis reports, publications, and visualizations, which help illustrate the breadth, depth, and rigor of the science created by data contributors. Where technologically possible, the system will link data with publications and other products to illustrate the scientific contributions of MPA monitoring participants. Data contributors would also benefit from a system that offers contextual and ancillary data to which they may not normally have easy access. Additionally, they would benefit from a system that allows for direct data access and can output data in enhanced formats (e.g., Representational State Transfer Application Program Interface, Web Mapping Services, and visualizations).

Persona Definitions:

1. ***Resource Managers****: Local, state, or federal agency staff responsible for management and/or regulatory decisions about MPAs or issues directly affected by MPAs. Organizational affiliations of this persona include the California Department of Fish and Wildlife (CDFW), California Ocean Protection Council (OPC), and NOAA's Office of National Marine Sanctuaries (ONMS).*

2. ***Decision Makers****: People with decision-making authority related to MPAs. Examples included individuals that make decisions at both state and federal levels, including members of the*

California Fish and Game Commission (FGC) and legislators who make high-level management or policy decisions about MPAs or related issues and staff who advise those decision-makers.

3. ***Data Contributors**: Any person or group that collect data as part of a structured MPA monitoring project. For current MPA monitoring data contribution, this requires official support as a State-funded monitoring project. In the future, data from non-State-funded MPA monitoring may be added to the system based upon relevance, scientific rigor, and contributor adherence to data and metadata standards.*

4. ***Users**: People that are informed and interested in the management or regulations related to MPAs. Examples include: Commercial and Recreational fishermen, environmental non-governmental organizations, and concerned citizens.*

Goals for MPA Monitoring Data

The Statewide MPA Monitoring Program aligns with the goals and objectives within the Marine Life Protection Act (MLPA) and MLPA Master Plan for MPAs.[5] This program is led by Ocean Science Trust and implemented through a collaboration amongst Ocean Science Trust, CDFW, and OPC. Specific goals for MPA monitoring data are guided by the priorities and needs of the State, including uptake and synthesis of raw monitoring data:

- Ensure the sustained use of MPA monitoring data by making them accessible, discoverable, durable, and appropriately described; and
- Ensure interoperability of data to allow for higher-level synthesis across organizations that maximizes the return on investment and widespread use of the data.

In addition to the use of raw data, MPA monitoring results support science-informed decision making. These goals can be stated in terms of the following desired science and management outcomes related to data and information use:

- MPA monitoring data and results are published in scientific papers;
- Data integration, synthesis and uptake into products (e.g., scientific publications) help demonstrate the broad relevance and applicability of the data and results;
- MPA monitoring results are seen as rigorous and credible, particularly by resource managers and policy makers, and serve as the foundation for understanding MPA management effectiveness and answering network evaluation questions.

Planning Process: who, why, and how

To accomplish the stated MPA monitoring goals, we need to leverage existing resources, and construct a plan that draws from the knowledge of data management held by our partners. Therefore, we engaged

[5] The goals of the Statewide MPA Management Program align with the goals of the Marine Life Protection Act (MLPA), and program objectives are detailed in the MLPA Master Plan for MPAs. Research and monitoring is one component of the management program.

in a collaborative planning process using informational interviews with over 15 government, nonprofit, and for-profit scientific organizations. As a result of these calls we decided to create an Advisory Team, made up of professionals with expertise in scientific monitoring, data and metadata management, information systems and storage, data informatics and semantics, and web portals for data distribution. Iterative engagement with the Advisory Team provided a depth of engagement and knowledge that would not have been available through a simple series of calls and interviews with individuals.

After our initial meetings with the Advisory Team we developed a limited a scope for the data management framework. To vet our current thinking we convened a workshop made-up of the Advisory Team and partners from the State and the MPA monitoring community (see Appendix xxx for the Workshop Summary). The main goals for the workshop were to examine the real-world use cases for monitoring data, budget line items, and technology requirements. As a result of the workshop, we determined the infrastructure requirements for a data management system will likely require us to utilize existing technology, or an off-the-shelf solution to be cost effective and sustainable. This system will likely need to be pre-configured with certain functionality, but also allow for customization on both the back and front-ends. The level of interest exhibited by the Workshop Participants represents opportunities to create linkages between our data system and other established systems, providing us with a greater opportunity for making data interoperable thereby maximizing use and return on investment.

This document went through solicited review with the Advisory Team, the Workshop Participants, MBARI (Monterey Bay Aquarium Research Institute), and Google.

Advisory Team Members

- Tanya Haddad, West Coast Governor's Alliance; Oregon Department of Land Conservation and Development
- Dr. Matt Jones, National Center for Ecological Analysis & Synthesis; University of California Santa Barbara
- Dr. Steve Lonhart, Monterey Bay National Marine Sanctuary; National Oceanic and Atmospheric Administration
- Jennifer Patterson, Central & Northern California Ocean Observing System
- Paulo Serpa, California Department of Fish & Wildlife

Overview of MPA Monitoring Data

California invested $16 million to support baseline MPA monitoring across all four coastal MLPA regions. Each of the four, regional baseline monitoring programs produce comprehensive sets of data, metadata, and other assets, including reports, visualizations, and other products. Taken together, these programs have generated the most comprehensive statewide dataset to-date, characterizing eight ecosystems and human uses at or near the time of regional MPA network implementation. As of fall 2015, this dataset extends from the California-Mexico border north to Point Arena, and comprises tens of thousands of data points, bundled into over 250 data packages across 27 projects. By 2018, when baseline monitoring in the North Coast region is complete and data from long-term monitoring in the other three regions are coming online, this dataset is likely may double in size. In addition to the data packages, the monitoring programs also produce other assets, such as imagery and maps, as well as technical reports, project summaries, and state of the region reports (digital and print). Demonstrating the rigorous science that lays the foundation for these information products is a vital component of data management in support of decision-making.

California's MPA Monitoring Program includes both condition and trends monitoring, and management effectiveness evaluations under the umbrella of ten Ecosystem Features, including ecosystems and human uses.[6] As such, the program employs a vast array of data collection methods that produce a wide breadth and diversity of data, crossing multiple scientific disciplines, including ecological, environmental[7], geospatial, and social sciences. Data and metadata files produced vary widely in format and size across these disciplines. For example, ecological research typically produces tabular data from field surveys that are bundled into .CSV and .TXT files and/or imagery (e.g., .JPG, .MPG4, .RAW), while geospatial research produces shapefiles (e.g., SHP, SHX, DBF), placemark files (e.g., KML, KMZ), and data visualizations often in the form of maps (e.g., PDF). In addition to diverse data types and varying spatial extents, these data also span varying temporal scales. Considering the breadth of ecosystems and disciplines involved, condition and trends monitoring to date has involved data collection at discrete temporal scales that vary by project, bringing time-series to bear in analyses when and where available. For example, baseline monitoring projects involved one to three years of data collection. As we transition to long-term monitoring in 2016, data collection will likely proceed at more regular intervals.

The State has taken a partnerships-based approach to MPA monitoring to not only bring together the diverse ocean community in California, but also to develop and implement a cost-effective and sustainable program in support of MPA management. Through this approach, we are leveraging existing capacity across the State, including State and federal agencies, tribal governments, academic and citizen scientists, fishermen, NGOs, and the private sector, including over 70 organizations engaged in baseline MPA monitoring across the state. As we look ahead to long-term monitoring, there may be opportunity to expand to yet another burgeoning source of data: crowdsourcing. The diverse network of organizations involved in producing data has a direct impact on the complexity of data management. For example, organizations may vary in their capacity for data processing, packaging, and technical support, and produce data with a wide range of geographic and temporal scope.

---

[6] California's MPA Monitoring Framework identifies ten Ecosystem Features, including eight ecosystems and two human use categories (i.e., consumptive and non-consumptive). For more information, see the framework guide: http://oceanspaces.org/sites/default/files/regions/files/monitoring_framework.pdf.

[7] Within the context of California's Statewide MPA Monitoring Program, environmental data includes oceanographic data, such as biochemical parameters (e.g., chlorophyll, dissolved oxygen) and physical parameters (e.g., current velocity, wave height, sea surface temperature), and atmospheric data (e.g., air temperature, precipitation, wind velocity).

Data and Metadata Standards

Data and metadata standards were developed to guide data submission at the close of regional baseline monitoring programs. Ocean Science Trust is currently updating these regional standards to establish a single set of Statewide Data and Metadata Standards for long-term MPA monitoring. The standards aim to supplement data coordination and integration, facilitate data submission to the Ocean Science Trust and the State at (or before) the completion of projects, and to increase consistency and utility of MPA monitoring data. The standards outline required individual metadata fields in plain language, for example project descriptions (e.g., what was measured, where, by whom), which enables both user discovery as well as machine-to-machine automated discovery. All MPA monitoring projects funded by the State are required to submit data and metadata, in accordance with the standards, with the understanding that these data will be made publicly accessible because MPA monitoring data collection to-date has been funded by the State.

As outlined previously, MPA monitoring data span a diversity of disciplines and thus require guidelines that match the specific data types produced. As such, corresponding metadata standards are employed for each data type with a metadata crosswalk employed for communication between the two standards:

- Ecological Metadata Language (EML) is the metadata standard used for ecological and socioeconomic data, which was developed by the Knowledge Network for Biocomplexity (KNB) and offers compatibility with the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM)
- ISO-19115 is the metadata standard for geospatial data, which was developed by the International Organization for Standardization, and an ISO-EML crosswalk already exists (see below)

Program-specific standards are under development for traditional knowledge, local knowledge, and imagery, and will be added to the standards document. Pre-existing, long-term monitoring programs (e.g., CeNCOOS) focused on collecting solely environmental data[7] are outside the scope of the MPA Monitoring Program. Thus, specific standards for these data are not outlined in the standards document. However, some environmental data are often collected as part of ecological monitoring projects (e.g., PISCO) and as such are governed by EML. As new data types are collected, additional data and metadata standards will be developed and incorporated into the standards document. In recognition that not all monitoring program collaborators have established metadata practices or technical experience with metadata standards, metadata templates may be developed and provided to data contributors. During data management implementation planning, we will assess the ontologies and vocabularies necessary to ensure interoperability among data providers and add those to the data and metadata standards.

Handling sensitive data

There are two main types of sensitive data: (1) those that are governed by nondisclosure or confidentiality agreements and (2) those that include information about protected species or culturally sensitive sites. Data that are governed by nondisclosure or confidentiality agreements are typically submitted as data level 3 or 4 (e.g., data produced through interviews and some boat or port-specific fisheries data). Data that include information about protected species, such as species listed in the Endangered Species Act, are not shared publicly. For these data, they are submitted in two packages. The first includes the data on the sensitive species but is shared only with the MPA monitoring management agencies team (i.e., CDFW, OPC, Ocean Science Trust). The second includes data that are scrubbed of all information about the sensitive species, which is shared publicly.

Quality Assurance/Quality Control

The bulk of the responsibility for quality assurance/quality control (QA/QC) of data and metadata lies with the Data Contributor. The data upload workflow is based on the standards to easily guide users through the process. The process culminates with the Data Contributor indicating that they have read and understood the standards and that data being submitted have been through a QA/QC process. We are beginning to work with KNB to investigate whether KNB, as a system, can offer additional QA/QC layers to help automate the process of checking for errors or deviation from the required standards.

Developing metadata crosswalks

To improve the ability of other systems to interact with contributed data, we plan to utilize metadata crosswalks. The main crosswalk of concern is the translation of EML metadata to ISO. The majority of MPA monitoring data falls in the ecology category and is adhere to by EML standardized metadata. By investing time in standardizing our metadata requirements and providing a mechanism for translating EML to ISO we can increase the interdisciplinary use of data and the ability to view the data based on geographic interest.

The EML to ISO crosswalk already exists, and is included as part of the data workflow within this plan. In the future, new metadata created that conforms to the Statewide Data and Metadata Standards attached to this plan will be ready for translation. The task remains to update, and in some cases generate, metadata associated with MPA monitoring data collected under the older standards (i.e., North Central Coast, Central Coast, and South Coast). These data will need translation to new EML or ISO to be updated to adhere to the statewide standards for inclusion in the new system. This process will mostly focus on metadata that need to be translated or created. Central Coast baseline data packages do not include metadata on the project or data table level. Most, but not all, North Central Coast data packages have EML metadata, but the XML metadata files may require formatting changes to be machine-readable. South Coast data packages include metadata for all but the geospatial data.

National Center for Ecological Analysis and Synthesis (NCEAS) is continuing to work on improving metadata crosswalks. We will work in close partnership with NCEAS to build crosswalks that improve data sharing and interoperability. By attaching standardized metadata to packages and ensuring the XML is machine readable on those with metadata, we will increase data availability, discovery, durability, and uptake by other data systems.

Data System Architecture

The data system architecture will focus on meeting the needs of our two primary user personas and will be implemented in phases, - beginning with the minimum viable product (MVP). OceanSpaces, as a data management system, provides the core set of features for building out the initial data system. The following section will outline the basic framework for operationalizing our user's needs, data management requirements, and workflows necessary to provide a system that builds on the successes of OceanSpaces, while looking to the future needs of data sharing, interoperability, and visualizations. Some areas addressed will remain necessarily high-level as the technical requirements for creating server-to-server linkages or dynamic visualizations will require case-by-case development. They are included here to illustrate the intent of leveraging existing resources and our goal to create a collaborative data sharing system.

Proposed Data Workflow

The data system framework illustrated here below is meant to show the basic workflow surrounding MPA monitoring data from the upload process through to the user exiting the system. External connections with services like DataONE, CeNCOOS, and PISCO, and added features like data visualizations, are included to illustrate additional functionality that can be developed to move us above our minimum viable product (MVP). A full-sized version of this framework can be found in Appendix F.

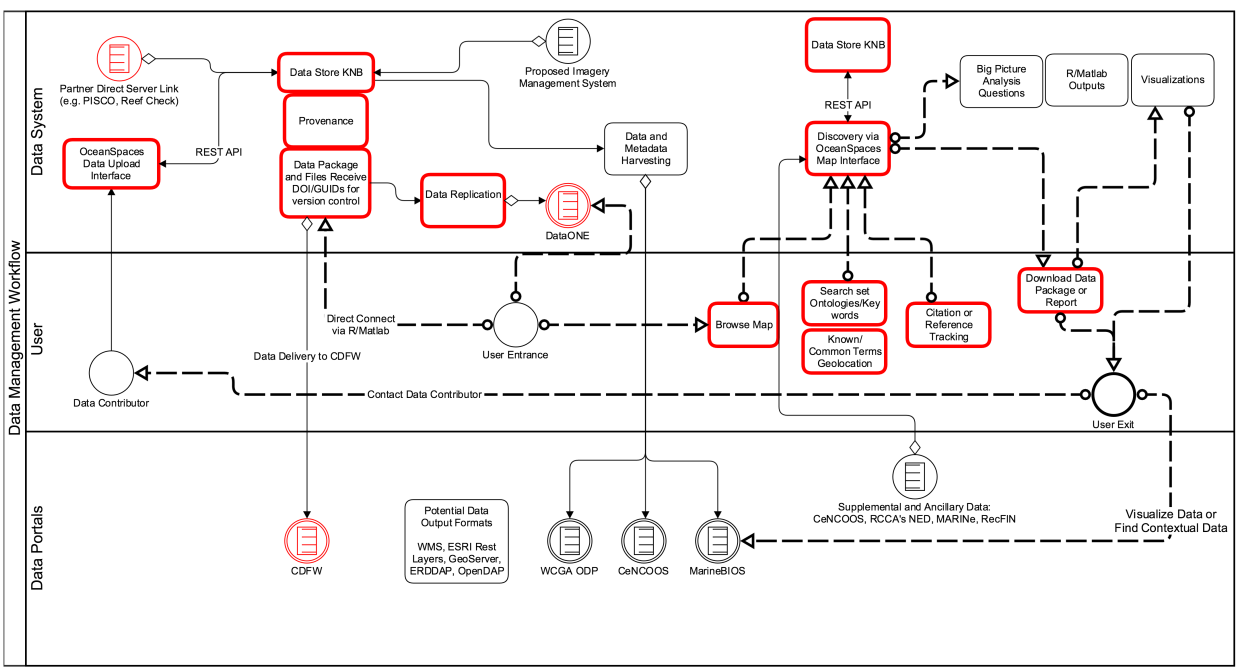


Figure 1: Components colored red are part of the initial MVP phase.

Data System Features

1. ***Data Upload:*** One of the primary functions of this data system will be to ingest data through a streamlined and easy to use workflow. Two methods of ingesting data are required:

Manual upload user interface: This manual upload option is important as many of our monitoring partners do not have advanced data management systems and use OceanSpaces as their primary vehicle for submitting data.

Server-to-server: Automated workflow for partners with advanced data management capabilities will be created to reduce the man-hours required to upload data.

KNB has the ability to accept data through a customized manual data upload process, - similar to the process currently available on OceanSpaces. KNB also has the ability to accept data through a server-to-server automated workflow. For partners with advanced data management systems such as PISCO, we can dramatically reduce the man hours required to upload data by creating an automatic push/pull workflow.

2. ***User Validation:*** The system will include a user management framework for authentication and permission control.
    a. Currently all users contributing data must be members of OceanSpaces. The OceanSpaces CMS software controls authentication and permission for data upload and management.

    KNB also offers an authentication and permissions-based user management system. The KNB uses both the InCommon federation and the ORCID registry for providing researcher identities. KNB allows for a Single Sign-on (SSO) connection between OceanSpaces and KNB, thereby reducing the complexity for users with data contribution permissions.

3. ***Quality Assurance and Quality Control:*** Due to the wide range of heterogeneous data created as part of MPA monitoring, the burden of QA/QC is held by data contributors. The Ocean Science Trust provides detailed data and metadata standards to help data contributors adhere to the schema required for data to remain accessible and usable far into the future. Data contributors are required to confirm that they have QA/QC'd their data at the end of the data upload process, - while also following the guidelines set forth in the Statewide Data and Metadata Standards. Ocean Science Trust and OceanSpaces, as a data distribution platform, do not currently have the capabilities to check data beyond basic data package completeness.

    Once the basic functionality of the new data management system is completed, we would like to explore adding a layer of QA/QC to the data upload process. KNB allows for custom QA/QC layers to exist within the data upload process.

4. ***Version Control and Provenance***:

    Version Control: The data system needs to incorporate the ability for data providers to upload multiple versions of a dataset and store these versions with a version control system. For example, this allows us to store a raw version and a transformed version of a dataset. The data provider should be able to describe differences to the data versions in a standard discoverable metadata field.

Provenance: The data system will have a method for maintaining a record of the origin of and processing used to create a dataset. Tracking provenance will help with data discovery, as many data users want to find a dataset that backs a certain finding.

The data system will incorporate a method for preserving the scripts used to create datasets because this is vital for data preservation. As part of the implementation plan development, we will explore the best way to preserve scripts alongside their datasets to ensure users have full access to the methods used by the original researchers. KNB supports archiving code and software products and any relevant analyses and transformational code can be linked directly to the datasets for future use. KNB also supports detailed provenance descriptions for describing data derivation and processing relationships following the W3C PROV model.

To provide version control, KNB assigns digital object identifiers (DOI) for every file uploaded. These identifiers ensure that every time a file is updated its file name is iterated. This is especially important when a user downloads a data package that has been updated due to errors or changes in how derived datasets are calculated. KNB also allows the provenance related to a data package to be shared as part of the metadata within the package. Additional details about data processing, citations, and reports can be displayed here as a form of metadata ensuring the data lineage is properly documented.

5. ***Data Semantics and Ontologies:*** A set of terms and vocabulary used within data packages allows us to create accurate search protocols and integrate datasets because they share a common descriptive language. To maximize the ability for data systems to communicate, we will explore creating these taxonomies for assets such as data tables and site lists. Due to the heterogeneity of MPA monitoring it is difficult to integrate data on the region or statewide scale. Developing data semantics and ontologies will help reduce the effort required to integrate these diverse datasets.

   KNB staff continue to work on the integrating semantic interoperability and improved search accuracy through the use of formal ontologies. As part of the DataONE network advances made by members (such as KNB) will be available to us to increase the opportunities for integration and improved search capabilities.

6. ***Replication:*** To improve the availability and security of MPA monitoring data, it should be stored on a data system that allows for replication to other nodes located at different geo-locations. Replicating the data increases redundancy and improves availability. A primary goal of this plan is to improve access to and the discoverability of MPA monitoring data. Therefore, replication to the DataONE network should be a direct result of the implementation of this plan.

   KNB functions as a member node of the DataONE network. As a member node, data submitted to KNB is made available to the DataONE system and replicated to other DataONE nodes. This provides resilience to our datasets by ensuring multiple copies exist across the country. Data in the KNB are then replicated to geographically dispersed hosts, and an audit process periodically checks the validity of the replicas, and replaces them if any that have been deleted or corrupted.

Additionally, housing data on KNB and DataONE improves data discoverability and creates a higher profile for MPA monitoring data. The State sees this as a high priority and a primary way to ensure a greater return on investments for in MPA monitoring programs.

7. ***Data Services:*** The primary value-add of a data management system is the ability to share data through various services. The State and MPA monitoring partners are very interested in building connections between the MPA monitoring data system and their organization's data systems. To accomplish this task, MPA monitoring data needs to be housed in a system that makes data accessible through automated processes to allow for expanded integration and collaboration.

   KNB is built on an open source REST API[8] that allows for development of tools for direct and seamless access to data via REST, Java, Python, and R. As part of our implementation plan we will explore the creation of tools that will make MPA monitoring data available to other systems. Data owners such as CDFW, CeNCOOS, SIMoN, and PISCO will work closely with staff on creating these tools.

8. ***Security and Access:*** The data management system will be hosted on a platform that provides a secure infrastructure, replication, and backups.

   KNB is run in a physically secure data center at University of California Santa Barbara (UCSB), with fully redundant power, cooling, and network. The security system uses a defense in depth strategy, including firewalls at both the institutional level and at the host level, intrusion detection systems for all network traffic, and periodic audits of software and processes. For backup, all data will be replicated via DataONE to other data centers upon ingestion, and all data are backed up to tape nightly on site, with periodic offsite tape backups. Data are not encrypted in storage, but all data are encrypted over the network using TLS/SSL.

Based on the MVP requirements for this system developed in coordination with the Advisory Team and the Workshop Participants, we are modeling our architecture around OceanSpaces and the Knowledge Network for Biocomplexity (KNB) data system. KNB is a platform developed by the National Center for Ecological Analysis and Synthesis (NCEAS) based at UCSB. OceanSpaces will act as the user interface, and KNB will act as the data management back-end. By basing our technology infrastructure needs within the University of California (UC) and California State University (CSU) systems, we leverage existing State resources, a primary goal of our approach to data management. By doing this, we promote the durability and longevity of the chosen platform and establish an enduring collaboration among Ocean Science Trust, state agencies, and UC/CSUs.

KNB is also an established technology platform for data management that offers flexibility and enhanced features, such as a decoupled REST API for creating custom interfaces, unique digital object identifiers (DOI) to ensure accurate version records, and data management workflows. KNB is also a partner with DataONE. By choosing to work with KNB, MPA monitoring data will automatically replicate to DataONE,

[8] Representational State Transfer Application Program Interface (REST API)

elevating state monitoring data to the national level. With KNB as our core data platform, we can design an interface that lives within OceanSpaces coupling information rich monitoring program and project descriptions with enhanced data ingestion and map-based data discovery.

Data Discovery & Integration

Discovery

Data discovery for MPA monitoring data will come in two forms: map-based and filtered search. Map-based discovery will be available through OceanSpaces and will include options to subset displayed data, and filter displayed data spatially and temporally.

Filtered search discovery will allow users to browse the entire collection of data packages and utilize search with filters. Search and filter taxonomies will be developed to provide the most useful terms for our target user groups, including:

- Thematic
- Data Contributor
- Temporal
- Taxonomic
- Ecosystem
- Keyword

Map based discovery can be enabled in OceanSpaces by utilizing the open-source java-based mapping system currently running on KNB. The map can be customized to display a specific subset of the data available on KNB. The map can also be coupled with filtering options and spatial selection tools to increase search precision. KNB also offers basic keyword search, which can be extended with the use of semantic ontologies to increase the accuracy of search results from the metadata index. The metadata index can be searched directly via Solr query through the REST API.

Data Delivery

Data selected through map-based or filtered search discovery can be downloaded to the users system as discrete data packages. Each data package contains the raw data and accompanying metadata files; methods and lookup tables; and any additional MPA monitoring program files. Each data package will also include SHA-1/256 hash values to verify the integrity of the files within the data package.

Integration

Integrating MPA monitoring data with other systems is a priority for the State and a benefit to the data contributors that also act as a data user. This plan recommends creating additional uptake pathways and syntactic interoperability connections for State-funded MPA monitoring data. Data contributors and the State see value in creating web ready data through advanced services such as:

- REST
- Web Mapping Services
- Esri Rest Layers
- Direct R and Matlab Access

Enabling these services will allow other data portals such as DataONE, CDFW's MarineBIOS, West Coast Ocean Data Portal, SIMoN, CeNCOOS, PISCO, and ArcGIS Open Data to register MPA monitoring data.

Data contributors and the State see value in creating a rich data experience. This requires the contextual information about MPA monitoring programs and projects supplied by OceanSpaces, and the ability to link to and/or display contextual and supplemental data alongside State-funded monitoring data. For example, data users expressed interest in a use case related to the recent sea star wasting event. A system that can bring the Multi-Agency Rocky Intertidal Network (MARINe) sea star wasting syndrome abundance data together with together sea star abundance data, with CeNCOOS conductivity-temperature-depth (CTD) data, and MARINe sea star wasting syndrome counts creates a hub for finding the most up-to-date and relevant data. During implementation of this plan, we will explore the most valuable and effective linkages with other data providers – such as: CDFW, NMS/NOAA, CeNCOOS, PISCO, MARINe, and Reef Check California.

KNB's open source architecture provides the framework to create the tools and linkages necessary to satisfy these requirements. Development costs for work performed by KNB staff are low and the REST architecture allows for seamless access to data and metadata objects. Data owners can easily create automated tools to extract data from the KNB repository or to contribute data.

Reports

Reports are considered level 4 data and a key part of the provenance structure. The data system will integrate reports and ensure persistence and global resolution. Reports will be discoverable and accessible through the map-based and filtered search discovery pathways. Technical and summary reports for the North Central, Central, and South Coasts are currently accessible through the program and project information on OceanSpaces. Reports are a key part of provenance and play a critical role in communicating results, so this plan recommends linking reports directly to the data packages in the data management system. OceanSpaces will act as a communications tool for disseminating reports through program and project communications materials.

KNB is working to integrate reports into the data management system, treating them like data and assigning a unique ID to ensure persistence and global resolution. Reports will be available through the KNB map interface and a filter search for data discovery.

Imagery

Imagery assets require large storage systems, so they are not usually held with the data derived following processing. The State and Data Contributors agree that persistent access to the imagery assets underlying MPA monitoring datasets is important. Most imagery assets related to MPA monitoring are held by their creators, which introduces the risk that assets may become unavailable in the future. For video imagery in particular, this plan recommends that an imagery management system (either physical or cloud-based) be created in collaboration with imagery management experts. The system can be linked with KNB to provide a discovery pathway for imagery via the metadata and the foundation. Ensuring access to imagery for the academic community increases the return on investment by allowing researchers to re-analyze archived footage to address new questions.

Visualizing and Communicating Results

Visualizations

A primary goal of this plan is to increase the return on investment for MPA monitoring data. The State wants to improve the discoverability, relevance, and usability of monitoring data. The previous sections of this plan address access and relevance. The new data management system should improve our ability to access data for the creation of visualizations, which can help answer questions for decision-makers. There are two types of visualizations of interest to our user groups. The first type of visualizations is basic displays of information, such as species counts, trend graphs, or infographics. The second type of visualization is more advanced, including heat maps or time-series graphs. Development of any visualization products should be strictly focused and closely tied with the goal of answering specific questions relevant to MPA monitoring assessments.

The technical development of visualization tools will need to be addressed in phases, with very basic visualizations tools developed first. Visualization tools should prioritize adaptability across multiple spatial and temporal scales, as well as across varying priority management or research questions. Other portals, such as CDFW's MarineBIOS, offer advanced geospatial data display capabilities. Data visualization development partnerships should be created with portals like MarineBIOS and CeNCOOS to reduce the duplication of effort. Working with these partners on interoperability alignment during the implementation phase will reduce the difficulty of displaying data from multiple sources.

Communicating Results

OceanSpaces will continue to fulfill the communications needs around MPA monitoring. Data and product launches will be shared and contextualized through the current OceanSpaces channels. To help ensure data products (e.g., technical reports, summary reports, and articles) are properly catalogued and globally resolvable, they will be housed in the KNB data system to provide a digital object identifier and ensure they are properly documented in data package provenance. As stewards of the data, Ocean Science Trust will continue to support marketing and social media initiatives to share data products with the stakeholder communities and create relevant links on partner websites. Partners such as MBNMS/NOAA and CeNCOOS have voiced their support for bi-directional sharing of data and results. We will continue to work with these and other organizations to expand the reach and availability of our shared data pools.

Implementing this Plan

In collaboration with Knowledge Network for Biocomplexity (KNB) staff and with input from the advisory team (or a portal partner team), Ocean Science Trust will develop an implementation guide plan for the data management infrastructure that operationalizes the requirements identified in this plan. This plan will likely include the following:

- OceanSpaces will act as the primary access point for monitoring data upload and discovery. Working with the OceanSpaces web development team and KNB, new resources will be developed that interface with the KNB data system for uploading and discovering baseline monitoring data. This system will be utilized by North Coast MPA Baseline Program collaborators will submit data and metadata through the new data system for data in fall/winter 2016, following training on the use of the new data upload system.

- Based on the new infrastructure, MPA Monitoring Statewide MPA Monitoring Data and Metadata Standards will be updated to incorporate new metadata requirements and data management workflows. North Coast MPA Baseline Program data contributors will receive training on the use of the new data upload system.

- To bring existing MPA monitoring data already housed on OceanSpaces into compliance with this plan, Ocean Science Trust will create a position description for a Data Manager/Analyst to work with staff on update baseline monitoring data packages for the North Central Coast, Central Coast, and South Coast to work with the new KNB-based data management system.
    - Ocean Science Trust will create a position description for a Data Manager/Analyst who initial focus will be on updating existing data packages.

- To address the needs for housing imagery assets, Ocean Science Trust staff will implement an imagery asset management system. The goal for this system is to provide a single point of contact for accessing imagery assets used to create MPA monitoring data, and ensure the longevity and use of those assets for new future scientific questions.

To accomplish these priorities, below is a summary of the tasks to be supported by the budget in Year 1:

- In collaboration with KNB staff and support from a web development team, develop new data management architecture linking OceanSpaces with KNB - and potentially with other data portals through new tools.
    - Integrate the OceanSpaces data upload tool with KNB systems, and with the map-based data discovery tool for MPA monitoring data.
    - Develop and launch a new map-based data discovery tool for OceanSpaces that includes all protected areas, spatially displays MPA monitoring data packages, and links to other data sources.

- Develop a process for migrating existing baseline monitoring data packages into the new data management system, including bringing them into compliance with the Statewide Data and Metadata Standards.
- Work with state partners and others as appropriate to identify the long-term requirements and plan for data delivery and integration with existing data management systems administered by CDFW and other key partners.
- Develop a shared understanding of how to deliver MPA monitoring data and metadata to CDFW
- Begin scoping the collaborations necessary to implement the imagery asset management system.

Funding Requirements

The following budget describes the funds necessary to optimally meet the requirements outlined in this plan for year 1 and year 2 of implementation with ongoing support. It leverages existing data management resources, in the form of KNB and OceanSpaces, to achieve the goal of improved discoverability, relevance, and usability of MPA monitoring data. The budget table depicts the total yearly allotment of funds to the three major funding areas – Infrastructure, Personnel, and System Development. We have existing relationships with service vendors like Project Ricochet, which will increase the speed of development and lower the costs of onboarding staff. The total for the first two years of implementing this Data Management Plan is $1.23 million. This plan will be implemented with a combination of State and philanthropic funds.

BUDGET	YEAR 1	ONGOING	SOURCE	VENDOR
Infrastructure	**$5,000**	**$100,000**		
Data & metadata storage	$5,000	$0	Outsourced	KNB
Imagery storage & access	$0	$50-100,000	Outsourced	TBD
OST Personnel Costs*	**$376,500**	**$504,500**		OST
Program Manager, Technology & Information Systems	$84,000	$84,000	OST	30% time FTE
Program Manager, MPA Monitoring	$42,000	$42,000	OST	15% time FTE
Communications Coordinator	$27,500	$27,500	OST	10% time FTE
Data & Imagery Manager**	$128,000	$256,000	OST	100% time FTE
Associate Scientist	$95,000	$95,000	OST	35% time FTE
System Development	**$165,000**	**$80,000**		
OceanSpaces maintenance & other operating Costs*	$90,000	$30,000	Outsourced	Ricochet $165/hr
Data Management Architecture Development	$35,000	$0	Outsourced	Ricochet $165/hr; KNB $80/hr
Map-Based Data Discovery Tool	$40,000	$0	Outsourced	Ricochet $165/hr; KNB $80/hr
Ongoing data management system development	$0	$50,000	Outsourced	Ricochet $165/hr; KNB $80/hr
GRAND TOTAL	**$546,500**	**$684,500**		

** Indicates recurring costs and fully loaded staff rates approved by the State*
*** Data & Imagery Manager will be hired by Q3 in year 1.*

Adapting this plan

This plan provides recommendations based on the current working knowledge around managing MPA monitoring data. To ensure the durability and optimal usability of these data there should be a review process to update this plan to stay in line with current data management and technology standards. The recommended schedule for full review of this plan is to align with the statewide MPA network ten-year management review cycle. In addition, Ocean Science Trust will consider emerging technologies and best practices, as necessary, outside the full review cycle.

In addition, Ocean Science Trust will conduct two additional reviews during the first two years of implementation of this plan (2016-2018):

1. Review and update after North Central Coast, Central Coast, and South Coast data are transferred to KNB.
2. Review and update after the North Coast baseline data are uploaded to KNB.
3. Review and update in advance of 10-year management review in 2022.
    a. This would be a six-year review of this plan as it will have been in place since 2016.

Participants: Representatives from the Advisory Team and each of the user personas described above should be included in the review process.

Supporting Materials

All supporting materials and Appendices are available online

1. NASA Earth Science Data Processing Levels - http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/
2. Knowledge Network for Biocomplexity - https://knb.ecoinformatics.org/
3. DataONE - https://www.dataone.org/

Appendices

A. A Comparison and Recommendation of Metadata Standards for MLPA Baseline Monitoring Data

B. Marine Protected Area Monitoring Enterprise - User Needs Assessment (2010)

C. Data Portal Inventory

D. Workshop Summary

E. Visualizations Meeting Summary

F. Data Management Workflow

G. Organization Conversation Log

Appendix A: A Comparison and Recommendation of Metadata Standards for MLPA Baseline Monitoring Data

Will Tyburczy
Mark Schildhauer
Matt Jones
Jim Reichman

National Center for Ecological Analysis and Synthesis
735 State Street, Suite 300
Santa Barbara, CA 93101

Prepared For:
The California Ocean Science Trust

June 12, 2007

Table of Contents

"The great thing about standards is there are so many of them to choose from."
– Andrew Tanenbaum

1. INTRODUCTION

While the utility of metadata has been recognized within the environmental research community for some time (Michener et al. 1997; Gross and Pake 1995), widespread consideration of best practices for scientific metadata has been the focus of major institutions for only a few years (Fegraus et al. 2005; Hamre et al. 2004, Isenor 2005, IOOS DMAC 2005). This increase in emphasis has arisen in part due to increasing concerns that data are not adequately preserved by researchers after the publication of their studies in scientific journals (Green et al. 2005). Many valuable data sets have been rendered useless or reduced in value because contextual information crucial to understanding the data has been lost or forgotten (Michener et al. 1997). Metadata provide a solution to this problem, by storing the information needed to interpret data alongside the actual measurement values. Standardization of metadata ensures that data sets are documented in a uniform fashion, and that relevant metadata are not omitted.

The California Ocean Science Trust (CalOST) has commissioned this report to clarify the benefits and costs associated with metadata approaches for data preservation and usage; to identify and compare several potentially relevant metadata standards for Marine Life Protection Act (MLPA) needs; and to make recommendations as to which standards should be adopted for MLPA baseline monitoring data. This effort assumes that metadata constitute an effective way to meet the key anticipated data discovery and integration challenges so that synthetic and integrative analyses can accurately inform adaptive management strategies for the state's Marine Protected Areas (MPAs).

There are many technological approaches to collecting and storing scientific data within a common framework for access and analyses, and the merits of these various approaches, particularly how to associate multiple data sets that have a broad range of themes and structures, are still subject to much debate in both the general database and scientific database literature (Jones et al. 2006). These discussions can get very technical, and cannot be discussed in any detail within this short report. Similarly, metadata standards can become very complicated, in terms of the intended purpose of the standard, the range of concepts covered, the technical details of how the standard is constructed, and the availability of tools to support the creation and accessing of documents or data sets within that standard. Many of the standards we discuss have a voluminous amount of literature associated with them that clarify these issues, and we can only attempt to summarize them here on behalf of MLPA researcher needs.

The table included as Appendix C below (in US legal page format) attempts to summarize the relevant aspects of a number of metadata approaches that might be useful for MLPA needs. Further information about many of these standards can be found by visiting the associated websites, which are provided when available.

Data supporting MLPA efforts will be collected under the auspices of a number of programs and agencies, including the California Ocean Science Trust's Marine Protected Area Monitoring Enterprise, the State Coastal Conservancy, the Ocean Protection Council, the California Department of Fish and Game, and the California Sea Grant Program. Data collection will occur via multiple awards to a large number of researchers, many working relatively autonomously and largely outside the formal organizational structure of the supporting programs. These studies will investigate a range of complex topics relative to marine community structure and composition, over a diverse set of habitats. The studies will focus on identifying and modeling a number of critical biotic and abiotic factors impacting local marine ecosystems, with special attention to identifying suitable areas for MPAs, monitoring their status once established, and intelligently informing adaptive management decisions (http://www.dfg.ca.gov/mrd/mlpa). Moreover, socioeconomic data will also be collected, some of which might be derived from external sources; and these must also be integrated into the analysis framework for the MLPA.

Broad-ranging, loosely coordinated investigations as envisioned under the MLPA monitoring program, raise major challenges in terms of developing a comprehensive database solution. Rather, this is exactly the type of project that can perhaps best benefit from a rich metadata documentation approach, in order to loosely couple highly diverse sets of data together into a common framework for discovery, access, and interpretation (Jones et al. 2006). It is this perspective which motivates the focus of this document--identifying and contrasting various metadata solutions for integrating diverse environmental and ecological data sets.

1.1 Metadata Defined. Put simply, metadata is data about (describing) data. It is the set of information, outside of actual measurement values, that is needed to discover, identify, access, interpret, and analyze a set of data (Jones et al. 2006). Metadata is not a term that is rigorously defined, but emerges in terms of "context"—the additional information that someone needs, to discover, access or interpret the data for some given purpose. However, here we do differentiate metadata that are formally structured from those that are not. Informal metadata would include, for example, a natural language document that might describe in a set of paragraphs what a data set contains or is useful for. We typically reserve the use of "metadata" to mean "structured metadata" where the information is divided into smaller pieces, or fields, each with some specific and well-defined content, such as "Data set Owner" or "Data set Size". These metadata elements can collectively constitute a complete metadata description of a data set, within a given metadata definition. Metadata fields can cover a broad array of topics, including the units and precision of measurement, the data set owner and contact, the taxonomic, spatial and temporal coverage of the data, the format and layout of the data, and much more. It is also worth noting that certain elements of metadata are critical for computers to be able to import, parse, or operate upon the data as more than an opaque object. For example, in the case of a table of data, basic information about the number of columns and their contents might be essential for the computer to be able to import and represent that data object as a table, with the appropriate number of rows, and like values aligned in columns. The ability of metadata to enable "computations" upon the data is a major benefit.

The various perspectives on metadata and their use, and the tools to facilitate metadata creation and querying, are highly variable, often complex, and rapidly evolving. This brief report represents a snapshot of the issues and standards that we consider to be most directly relevant to MLPA metadata considerations at this point in time. The brevity of this report has constrained us from more in-depth examination of existing standards, and led to reduced coverage of some emerging standards that might prove useful in the future.

1.2 Types of Metadata. Metadata can be classified according to a variety of criteria. Within the ecological discipline, one popular classification was created by Michener et al. (1997), in which metadata fields were organized into five classes according to five common questions a data user might want to ask. An alternative approach is to group metadata according to the benefits and costs to the data owner, rather than focusing on the data user. In this case, metadata are typically grouped and listed in order of increasing information content and complexity. These are provided only as a general guideline to frame discussion; they are not strict classifications, and any given piece of metadata might reasonably be assigned to more than one category.

Identification and Citation. Metadata fields in this category are typically the absolute minimum required of data within most any metadata standard. Fields in this category provide the information necessary to identify a data set for discussion and potential use, and to adequately contact or cite the data owner when used, with an appropriate title, unique identifier, data owner and data contact.

General Description. Information in this category, such as abstracts, keywords, and spatial, temporal, and taxonomic coverage, provide a basic understanding of the type of data contained in a data set. They are generally useful for data discovery and search, and often contain free-form text descriptions of the data. Methodology, project descriptions, sampling descriptions, and other such information falls into this category. These metadata are beneficial for selecting data appropriate for a specific use.

Logical Structure. These metadata provide detailed information about the logical model used to represent the data. Some data are represented in relational tables with a particular structure, others are represented as images or geospatial features. Metadata describing the logical structure document these decisions, including the number and data types for the measurements represented in the data, integrity constraints among relations, and geospatial data organization. These metadata are typically useful for selecting data appropriate for a specific use, and also are critical in assisting with the choice of appropriate applications with which to analyze the data.

Physical Structure. Physical structure metadata describes the format of the data, storage location, methods for accessing the data, and so forth. These metadata, in combination with descriptions of the logical structure, have great potential for use by computers for automating data integration and analysis. Because of this, a high degree of structure for these fields is a desirable characteristic of a metadata standard.

Semantic Interpretation. Documenting the semantic meaning of the various columns or attributes of a set of data is a non-trivial task. Metadata standards for this type of content are relatively new, and many are still under intensive development. This information has the potential to allow further automation of data synthesis and analysis activities by exposing information about the meaning of data measurements, contextual relationships among measurements and sampling designs, and the relationship of data measurements to higher-order scientific concepts. For example, semantic annotations could conceptually differentiate wet-weight and dry-weight measurements, even though both have mass as their units of measurement. They could also help researchers determine that measurements labeled “mass”, “biomass”, and “wet-weight” across different data sets were all measuring the same parameter. This semantic information can be used by semiautomated reasoning systems to enable new capabilities, such as the ability to discover data in ways that are scientifically relevant, integrate data sets with one another, transform data so that it can be used in analyses and models, and understand and interpret data collected by others.

Other Information. Other specialized types of metadata also exist. Several standards allow users to specify access control rules for their data. Provenance metadata, which indicate where data sets came from originally and how they have been processed since that time, is another emerging topic for comprehensively describing data. Since MLPA baseline data is to be made public, and will generally involve newly collected data, these types of metadata will not be discussed here, though they could be useful for future MLPA activities.

1.3 Metadata Standards. Before metadata standards were widely used, many researchers still kept metadata for their data holdings, often simply in prose form, typically in a laboratory or field notebook. However, the metadata fields that were recorded varied greatly among researchers and data sets (Fegraus et al. 2005). Even when similar metadata were recorded by different researchers, the format of the metadata varied greatly between data sets, making it difficult to search for data sets based on particular qualities. This problem is evident today when one does a Web search for data—it is hard to locate these based on specific criteria, such as data about some given topic(s), collected from some given area, by some known person, during some prescribed time frame. Metadata standards and specifications can significantly alleviate this problem by specifying a uniform set of content, and sometimes structure, for metadata (Fegraus et al. 2005). Metadata content standards specify only which metadata fields should be included, while metadata implementation standards specify both metadata content and the structure and layout of that content. For the purposes of this paper, metadata specifications that have been officially approved by standards organizations such as the ISO, and other major metadata specifications and agreements, as well as those specifications which are simply broadly adopted, shall both be termed “metadata standards” for simplicity. Individual organizations might need to consider whether the adoption of a metadata “standard” implies formal sanctioning or required use by a governing body (compliance), or is motivated more by the utility derived from adhering to a “specification” that is in broad usage within the community, and enables a powerful set of features for, e.g., scientific research.

Uses of Metadata. Structured metadata can provide many benefits to data users and data owners. The most common are listed here.

Preservation. The proximate reason for most institutions to adopt metadata standards is simply to prevent information loss. Ultimately the goal is to allow the data to be reused in the future, but the extent to which a data set will be used in the future is often not known. By fully documenting the context and meaning of a data set, data owners prevent their research from disappearing or becoming unusable (Michener et al. 1997). However, for this to be effective, metadata contributors must fully embrace the responsibility for providing complete and accurate metadata. It should also be noted that even if stored with adequate metadata, data can be lost if stored in a format that becomes obsolete.

Discovery. Basic description metadata, if shared, allows data users to search for data of potential interest (Isenor 2005), creating opportunities for collaboration and synthesis. Advanced semantic metadata allows very focused discovery of particular types of data from an even larger collection of data sets.

Access. Metadata concerning the appropriate use and access restrictions can document that data should only be used and distributed according to the directives of the data owner. These metadata can be essential in protecting potentially sensitive data, or guarding human subject privacy rights. This can also include instructions for data users about how to properly cite the data set when it is used.

Interpretation and Analysis. Metadata on the logical structure of data can aid researchers in successfully interpreting a data set so that the data are used properly in analyses. Metadata on the physical structure of data can potentially be used in software and statistical programs to automate certain aspects of analysis.

Dynamic Data Loading. Metadata concerning logical and physical structure, and semantic interpretation, can be used to automate data integration, allowing data from different data sets to be loaded dynamically into a common database system, or other data frameworks (IOOS DMAC 2005).

Benefits of Standardization. While metadata alone can provide many of the benefits listed above, some of them, especially those that allow for automation of certain tasks, require that the metadata be stored in a common format. Unless metadata conforms to a known, uniform structure, it is exceedingly difficult to write programs or other tools to facilitate common data activities like search, selection, integration, and analysis. Even if metadata is not stored in a uniform structure (i.e. no metadata implementation standard), enforcing a standard content for metadata makes certain that important information regarding data sets is available for future use. A metadata standard also provides the important benefit of promoting the use of well-defined fields that can lead to the consistent capture of critical information, with a high degree of confidence in their interpretation, and adequate levels of detail.

1.4 Costs of Metadata and Standards. Though the benefits of metadata and standards are numerous, they do come with a price. Metadata standards can increase the workload of data owners, and can impose restrictions on the types and formats of data that can be accommodated.

Overhead and Maintenance. The adoption of a metadata standard can impose some overhead on the data owner. While it is generically useful to document essential aspects of a data set that might not be explicit within the data itself, creating and maintaining standards-compliant metadata can require a significant time investment (IOOS DMAC 2005), particularly if the number of data sets is large or the level of required metadata content is high. The investment of resources may not be justified if a particular set of data is unlikely to be reused in the future. Also, if a metadata standard becomes obsolete, existing entries will have to be converted to a new format. Data owners may balk at participating if the costs of creating metadata are not outweighed by the perceived benefits to the individuals involved. However, recent synthesis efforts have demonstrated that even seemingly unlikely data sources may be critical to future scientific understanding, especially from a management or policy perspective where researchers are doing retrospective comparisons (v. Jackson et al. 2001).

Adoption of a metadata standard also typically entails work for the stewards of the metadata, particularly if the metadata desired significantly exceeds the standard's minimum required metadata. Metadata stewards must ensure that researchers contribute metadata sufficient to meet community guidelines, and not merely the bare minimum to be compliant with a standard. It can be complicated to monitor metadata contributions when using more flexible standards, which typically allow more user discretion as to what metadata is included. User discretion also introduces the prospects for entering erroneous metadata, so programmatic checks and metadata steward overview might be necessary to correct or update metadata as more information becomes available about a data set.

Conformance. Many standards place restrictions on the data they describe. Some standards are limited in the types of data they can describe (e.g. only geographic data) or in the required format of the data (e.g. only data in netCDF format). Researchers whose data are in an unsupported format would have to convert their data, which can be time-consuming and disruptive to data owners. Owners who collect data of a type outside the scope of a particular metadata standard will be unable to comply with the standard at all. For example, researchers performing laboratory analyses would be unable to describe their data in Geography Markup Language, which is confined to data with explicit geographic content, and might not have the relevant details about the data that could benefit a laboratory researcher. The costs of conforming to any particular metadata standard should be outweighed by the anticipated benefits in terms of future use of the data.

1.5 Controlled Vocabularies and Ontologies

Controlled Vocabularies. While metadata standards provide content guidelines for what to include in descriptions of data, and structure for where different pieces of information should be kept, they rarely provide a consistent terminology for naming common scientific or technical elements and processes (e.g. ecosystems, rocky intertidal zone, taxonomic groups, community interactions). For example, one researcher may discuss "predator-prey relationships" while another talks about "trophic interactions", even though both are describing the same types of scientific phenomena. This lack of consistency in terminology can impede data discovery, as researchers are forced to remember and search using an array of synonyms for each concept they want to find data on. Controlled vocabularies provide a finite set of terms with which to describe properties of interest, and can greatly enhance the usability of a set of metadata when consistently employed to augment free-form descriptions. Existing controlled vocabularies vary significantly in level of detail, form the nineteen theme keywords used in ISO 19115, to the hierarchical and far more expansive GCMD Science Keywords (http://gcmd.nasa.gov/Resources/valids//gcmd_parameters.html). Institutions can also develop their own controlled vocabularies for describing concepts and processes more specific to their fields of interest.

Ontologies. Ontologies represent a further elaboration of controlled vocabularies, where the individual words are also related to one another via a rich set of relationships. While metadata standards can be very useful for facilitating data discovery and access, activities such as data integration and analysis often require more detailed, structured information than that provided by metadata standards alone, in order to properly interpret and process the deeper, more semantically complicated aspects of a data set. Ontologies provide this capability by enabling researchers to annotate their data using a controlled set of interrelated and scientifically relevant terms, to explicate aspects of the data set of interest that might not be evident in the raw data or even the associated metadata. For example, a column in a data set might be called "kg" – but this does not clarify for us that the data in that column are actually **wet biomass** measurements of a **quadrat harvest** of **macroalgal cover**, and that the **biomass** should be measured in some **mass unit** such as **kilograms** (emphases here indicate the "annotation terms" or "key words" that would be drawn from the ontology). Unlike simple controlled vocabularies of keywords, ontologies provide the capability to embed these terms in a network of related terms, so that it is possible to use them for advanced reasoning—such as determining whether descriptions of the data are consistent (e.g. data are sampled from either **pelagic** or **benthic marine environments**, and **not both**); or related among columns (e.g. data in column 1 are **counts** of **recruits,** of a taxonomic identity specified in column 2 as **barnacles,** on a **settlement plate** of an area listed in column 3 as .25 $\mathbf{m^2}$). The standardized terminology and well-defined semantics within an ontology, allow processing applications to better utilize the information to perform functions such as data search and integration. Ontologies, by serving as independent reference standards, provide a basis for consistently expressing highly structured and rigorous relationships using scientifically meaningful terms attached to data and metadata. The combination of using ontologies with structured metadata paves the way for a number of advanced semantic applications to operate upon the data.

While less commonly used than metadata standards or controlled vocabularies, several ontologies are currently under active development. The Observation Ontology (Extensible)—or OBOE for short, is a formal ontology being developed by the NSF-funded SEEK project (Science Environment for Ecological Knowledge; http://seek.ecoinformatics.org). OBOE provides a framework by which data sets can be annotated with controlled terms, by selecting scientifically relevant concepts as defined in an ontology (e.g. terms like *pelagic, rockfish,* or *recruitment*), and "attaching" these to a data set or column within a data table. OBOE is written in a language called "OWL-DL", that is an accepted World-wide Web Consortium (W3C; http://w3c.org) recommendation for adding much greater detailed information about the "meaning" of any document, including data sets, as envisioned for the Semantic Web (Berners-Lee et al. 2001). OBOE is unique in having as its focus the general explication of scientific observations—that is, clarifying the context, types of measurements, and inter-relationships of any given observation with other observations. The OBOE ontology forms the basis of this annotation framework, with an initial software application that enables linking OBOE terms with EML-described metadata.

The Semantic Web for Earth and Environmental Terminology (SWEET) was initially developed by adapting the GCMD Keywords into an ontological framework. It has expanded in a variety of areas, notably by incorporating the keywords used in the Earth System Modeling Framework and the grid concepts of the Earth Science Grid. SWEET is composed of a set of ontologies, including ones for units, temporal entities, spatial entities, physical properties, and human activities, among others. It is written in another version of OWL, dubbed "OWL Full" by the W3C. A mapping from GMCD

Keywords to the SWEET ontology has been developed by project personnel to facilitate the transition from the controlled vocabulary to an ontological framework.

The use of ontologies for annotating data and metadata is quite promising, and represents an active area of research in the field of data modeling and information management. A detailed review of this field is beyond the scope of this document, so elsewhere we simply point out where we believe ontologies might provide particular added-value to the metadata issues which constitute the main focus of this document.

2. CRITERIA FOR EVALUATION

2.1 Content Area and Structure. These criteria are the most fundamental to the selection of metadata standard(s) appropriate for MLPA baseline and monitoring data. The standard(s) must contain metadata sufficient to allow data users to successfully interpret baseline and monitoring data for future analyses and resource management activities. The different types of metadata that should be included are discussed in the subheadings below.

To facilitate the effective use of data, the standard(s) must not only encompass the appropriate metadata elements, but also provide a robust structure within which those elements are stored. That is, while metadata intended primarily for human use (e.g. abstracts, project descriptions, methods) can be free-form text or loosely formatted, metadata that computers are likely to need for data analysis and integration purposes (e.g. units of measurement, data file format) must be formatted in a way that is machine-readable. This generally involves breaking down the metadata into discrete fields, each of which contains a specific type of information, and which is to be entered in a prescribed manner. Standards in which metadata are broken down to a greater extent are said to have finer granularity to their metadata.

Storage format. Metadata standards are typically expressed with attention to formal rigor, such that it is possible to develop or use convenient software tools for creating and validating metadata documents within that standard. The format of a metadata document depends on choosing some specific, and well-defined “storage syntax”. There are a number of these, and each allows for somewhat different levels of expressiveness, with different levels of existing software tool support. Commonly used formats over the last five years include, roughly in ascending order of complexity, the Extensible Markup Language (XML), a standardized language in which information is stored in a hierarchical fashion; the Resource Description Framework (RDF), a model where information is represented in subject-predicate-object statements; the Web Ontology Language (OWL), a language that can be used to described formal semantic relationships between concepts and objects; and the Unified Modeling Language (UML), a standardized graphical notation used to specify relationships in a model. These various storage languages provide some advantages when trying to edit or visualize the metadata, since standard software tools exist which can be used to work with XML, OWL, etc. While a metadata document might be stored in any one of these formats, the choice of storage format can have major implications in terms of how those metadata can then be effectively used, e.g., for fast and efficient searching.

Identification and Citation. This information is essential for data users to obtain, use, and discuss data of potential interest. It also enables data providers to be cited appropriately when their data is used. This latter point can be particularly important in getting data owners to comply with MLPA metadata guidelines.

General Description. Abstracts, keywords, and other general descriptive information allow researchers to discover data of potential interest within data management systems that exploit this information. Information such as the geographic, temporal, and taxonomic coverage of a particular data set are particularly useful in this regard.

Physical Structure. Describing the physical structure of the data can greatly facilitate the integration of multiple data sets, or even the analysis of a single set of data. Granularity of this type of metadata is particularly important, as software programs and computer scripts could potentially make use of this information to partially automate data analyses.

Logical Structure. Data users need to know the logical framework under which MLPA data sets were collected. It will enable them to select baseline data appropriate for their monitoring activities, and to ensure that they perform proper analyses on baseline and monitoring data. Data lacking this type of metadata could unknowingly be used in analyses for which they violate critical assumptions, or be otherwise used improperly.

Semantic Structure. Semantic metadata provide machine-readable information about the underlying meaning of a data set, which can be used by computer software to facilitate a variety of activities, including data search, analysis, and integration (IOOS DMAC 2005).

2.2 Flexibility. The MLPA data collection activities will be performed by a variety of institutions, agencies, and individual researchers. These data are likely to include a wide variety of measurements, including species abundances and distributions, oceanographic and biophysical measurements, and socioeconomic data such as resource economics. Given the distributed nature of data collection and the heterogeneity of the data, the chosen standard(s) will have to be extremely flexible in the types and formats of data they are able to describe. Flexibility (the types of data that can be described with metadata) and Content Areas (the type of metadata stored; see above) are the two most important criteria when determining the appropriateness of a metadata standard.

Support for Different File Formats. Researchers often vary in their preferred method of file format for data storage. This is due to many factors, including software and hardware platforms used, institutional regulations, research discipline default formats, and individual preference. Since MLPA data will span many researchers and institutions, viable metadata standard(s) will absolutely need to be able to accommodate a wide variety of data formats.

Support for Different Data Types. MLPA baseline data is proposed to include biological, environmental, and socioeconomic information, and these typically require a broad range of data types, including standard tables, relational data, vectors and matrices, large character fields, geospatial vector images, photo-images, etc. Thus, metadata standards that restrict themselves to describing only a narrow range of data types will be unsuitable choices for MLPA data.

Support for Relational Databases. Some researchers or institutions may prefer to store their data in a relational database system, such as MySQL, PostgreSQL, Microsoft Access, or FileMaker Pro, etc., rather than as a series of individual ASCII-text files. Furthermore, several of the data systems that the MLPA wishes to be able to inter-operate with, including the Biogeographical Information and Observation (BIOS) System and the recreational fisheries databases, house their data in this fashion. Therefore, the chosen metadata standard(s) ideally would be able to describe data held in relational database structures, Otherwise, the potential complexity of the relational structure will confound straightforward production of useful metadata descriptions. This might be alleviated by producing metadata descriptions for a set of scientifically useful “views” of the relational data, if such a set of pre-defined views can be developed.

Support for Spatial Data. Since one of the goals of the MLPA is to assess the impacts of MPAs on marine ecosystems, much of the data collected is likely to contain spatially explicit information. Selected metadata standard(s) should thus provide metadata fields appropriate for geo-referenced and other spatial data.

2.3 Accessibility/Usability. In order to get data owners to comply with MLPA metadata guidelines, the there must be tools available for selected standard(s) to simplify the process of metadata creation and modification, and to facilitate research activities such as data search, storage, and interpretation. Metadata standards that are both human- and computer-accessible will provide the greatest benefit to MLPA activities.

Human Access/Usability. Creation of metadata can impose a significant investment of time on data owners. Aids, such as detailed documentation, writing guides, or software tools, can greatly decrease the amount of time needed for data owners to efficiently create metadata entries. The availability of these tools can encourage MLPA data owners to provide metadata for their data. Other useful tools, such as search portals and workflow applications, which use captured metadata to provide added utility to data users, can also encourage participation in metadata activities. The MLPA should therefore consider the availability of these kinds of tools when selecting a metadata standard.

Computer Access. Although humans can easily understand text descriptions and other loosely formatted metadata, computers require a high degree of structure to be able to easily find and process information. Standards that break down the metadata into small pieces of information (i.e. standards that are finer-grained), store them in a structured manner, and use controlled content (such as vocabularies) for them, will enable computers to provide more precisely defined access to metadata. This allows current and future software programs to better automate data access, integration, and analysis activities.

2.4 Stability. To be an effective standard, a metadata system must be relatively stable through time, and must enjoy widespread use. Standards that have been widely adopted or endorsed by major institutions are more likely to persist over time, minimizing the cost of having to convert from obsolete “standards”.

2.5 Interoperability. The MLPA wishes to interface explicitly with a variety of data and metadata systems. Because of this, it is important to choose metadata standard(s) that are interoperable with other metadata standards, and that can be used to merge MLPA data with data in other systems.

Other Standards. The MLPA is specifically interested in compatibility with the standards used by the Biogeographic Information and Observation System (BIOS), Knowledge Network for Biocomplexity (KNB), CalFish, Integrated Ocean Observing

System (IOOS), Central and Northern California Ocean Observing System (CeNCOOS), Southern California Ocean Observing System (SCCOOS), Pacific Coast Ocean Observing System (PaCOOS), and California Department of Fish and Game (CDFG). Other data and metadata systems that might be of interest include the Long-Term Ecological Research Network (LTER), Ocean Research Initiative Observatory Networks (ORION/OOI), Partnership for Interdisciplinary Study of Coastal Oceans (PISCO), Ocean Biogeographic Information System (OBIS), Global Change Master Directory (GCMD), and Marine Environmental Data Index (MEDI). Since these systems use a variety of metadata standards, the MLPA should focus on selecting standard(s) that can operate easily with those used by these institutions, while being sure of compatibility with standards used by the most important potential partners of the MLPA.

Crosswalks, or conversions between metadata standards, have been developed for some of the standards discussed, and will be noted, as they can aid greatly in achieving interoperability between standards. However, crosswalks can vary significantly in their ease of use, as the different content and structure of the various standards can present formidable hurdles. Sometimes it is possible to convert from one standard to a second, but not to perform the reverse operation. As a trivial example, suppose standard “A” has a field for a researcher’s name, while standard “B” divides the name into separate fields for first and last name. Converting from “B” to “A” will simply involve combining the first and last names together, and can be easily automated, whereas converting from “A” to “B” would require figuring out which parts of the entry for the researcher’s name corresponded to the first name and last name-- a much more difficult task.

Data Systems. The MLPA wishes to use data from and operate with several institutions with data systems that either do not use metadata, or have decided not to adopt a metadata standard, such as the Multi-Agency Rocky Intertidal Network (MARINe), RecFin, PacFin, and others. Selection of a metadata standard should thus consider how that standard could be used to aid in using data from these systems with MLPA data. For systems lacking metadata, this includes determining the ease with which data from the outside system could be described with the MLPA standard(s).

2.6 Existing Tools and Frameworks. Although technically outside the scope of metadata standards per se, it is important to recognize that some metadata standards are already supported by a suite of technology tools that can be used to create, store, and query metadata. Some standards will also provide access to tools that help automate analyses and integration of data. Effective use of metadata standards lacking such tools will almost certainly require concerted development effort to construct them. Availability of these tools should be considered when selecting which standard(s) to adopt, especially since the MLPA is placing an emphasis on making their data accessible to a broad scientific community.

3 METADATA STANDARDS

3.1 The Knowledge Network for Biocomplexity (KNB) Ecological Metadata Standard (EML). Ecological Metadata Language (EML) is a metadata specification developed by the ecology discipline and for the ecology discipline, but is also capable of generically describing scientific data sets (Fegraus et al. 2005). EML contains metadata fields for a wide range of content, presented in a structure designed to be machine- and human-readable. It includes fields for the identification, general description, and physical and logical structure of data. This degree of comprehensiveness is balanced by a high degree of flexibility to the standard, allowing the metadata provider to ignore fields that are not applicable to a particular set of data or application. While originally developed for the discipline of ecology, the standard was constructed to be broadly applicable to scientific data, including quantitative information from the social sciences and economics.

EML has fields for describing many important aspects of scientific data, including details of dataset spatiotemporal coverage, methods, and sampling design. It provides a detailed structure for describing the logical model of a data set, including entity and attribute definitions, support for formally defining the interpretation of data values, and support for defining measurement units (e.g. kg/m^2) and coded values used in data sets. An EML dataset description can describe multiple data entities and the relational linkages among them. EML is implemented as a series of XML document types that can be used in a modular and extensible manner. Each EML module is designed to describe one aspect of the total metadata that should be included with any ecological dataset. One area of metadata not explicitly covered by EML that is available in some other metadata standards, such as SWE and the remote sensing extension of CSDGM (see descriptions below), is detailed information for remote sensing systems, and detailed information for the interpretation of information from ground-based sensors.

The development of EML was funded in part by the National Science Foundation, and is the structural underpinning of the Knowledge Network for Biocomplexity (KNB) program. Since it's inception, EML has been adopted by the National Center for Ecological Analysis and Synthesis (NCEAS), the Long-Term Ecological Research (LTER) Network, the Partnership for Interdisciplinary Study of Coastal Oceans (PISCO), and the University of California of Natural Reserve System (UCNRS), among others. It is also used in South Africa, Taiwan, and other countries. The Ecological Society of America has recently launched a system that utilizes EML for the voluntary registration of metadata for the data supporting analyses referenced in manuscript submissions for its journals. The intention is to require such metadata registrations sometime in the near future.

EML was developed with the FGDC CSDGM, BDP, ISO 19115, GML, DCMI (see descriptions below), and other standards in mind, relative to compatibility. A crosswalk between EML and the BDP has been developed, allowing metadata to be easily converted between the two standards. EML generally delivers a greater amount of metadata content and finer granularity than these other standards, making it more suitable for use by computer programs and scripts. Several cross-platform, freely available computer applications involving EML are already under active development. These include Morpho, a full-fledged EML metadata creation and search tool; Metacat, a server for metadata and data storage, data discovery, and creation of basic metadata; and Kepler, a scientific workflow application, designed to facilitate data integration and analysis in part by reading EML metadata to present a visual representation of data sets.

3.2 International Standards Organization (ISO) ISO 19115:2003 Geographic Information – Metadata. "ISO 19115:2003 defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data" (ISO 2003). This standard defines the metadata fields that should be included in any geographic data set, including metadata concerning the identification, description, logical and physical structure of the geospatial data. It should be noted that the standard does not define an implementation for this metadata – that is, it does not define how this metadata should be presented or formatted. Currently, the Open Geospatial Consortium (OGC) is working with the ISO to develop GML (see below) into an XML format for ISO 19115, which will be adopted as ISO 19136. The North American Profile to ISO 19115, or ISO 19139, is another widely-used XML format for storing the content specified in ISO 19115. As an ISO standard, ISO 19115 has gained broad acceptance across disciplines dealing with geographical information. Version 3 of the FGDC CSDGM (see below) is slated to be compliant with this standard, as is GML, which speaks to the level of support for the standard. EML was designed with attention to compatibility with ISO 19115, but a comprehensive crosswalk between the two does not currently exist. Several common GIS applications can partially automate the process of metadata creation for ISO 19115, making it relatively simple to produce this type of metadata when using one of those GIS applications for data management .

For the purposes of the MLPA, a potential drawback of ISO 19115 is that it is primarily focused on geographical metadata, although mechanisms exist for adding metadata profiles or extensions to complement the ISO 19115 elements. Data sets not including geographic information, such as certain types of socioeconomic data or data from laboratory analyses, would almost certainly require developing these custom extensions to ISO 19115, in order to be highly useful.

3.3 Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). The CSDGM was designed, similar to the ISO 19115, to house metadata for geographic data sets. Indeed, the next version of the standard will be compliant with ISO 19115. The standard is quite comprehensive in its scope, covering identification, description, and logical and physical structure of geospatial data sets. The description of the data logical model is focused on providing attributes that are referenced to geospatial features, and so lacks some of the flexibility needed to describe non-geospatial data, especially more complex relational models, such as those used in many ecological databases. It also does not use controlled vocabularies for many critical features of the logical model, such as units and data types, making it more difficult to do automated machine processing based solely on CSDGM metadata. The CSDGM requires geographic information for all data sets to which it is applied. This makes the standard unsuitable for data sets lacking a strong geographic context. This has been mitigated to some extent by the development of FGDC-endorsed modifications and extensions of the CSDGM, called CSDGM profiles. Some of these profiles, such as the NBII BDP (see below), allow the omission of geographic metadata.

Federal agencies that produce geographic data, such as the Environmental Protection Agency, have been required to use the CSDGM since 1995, which has led many institutions that collaborate with these agencies to adopt it as well. In its initial report on metadata, the IOOS Data Management and Communications (DMAC) Subsystem (2005) recommended that the IOOS adopt the CSDGM as the de facto standard until a more thorough review of metadata standards was conducted. Due in part to the federal metadata requirements, CSDGM has become the de facto standard for many institutions. Several IOOS regions, including CenCOOS and SCCOOS, use CSDGM metadata. BIOS uses CSDGM metadata as well, though it is separately developing a data model for a unified database; the database appears to have a proprietary metadata structure for describing it. There are a variety metadata creation tools for CSDGM on the Windows OS platform, as well as digital workbooks to guide data owners through the process of metadata creation. ESRI's ArcCatalog and other industry-standard GIS applications also provide tools or plug-ins that automate or aid in the creation of CSDGM metadata as well.

3.3.1 National Biological Information Infrastructure (NBII) Biological Data Profile (BDP) to the CSDGM. The BDP is a modification to the CSDGM designed to provide a better metadata framework for biological data sets. It includes some additional sections, such as that for taxonomic coverage, while removing the requirement for geographic information. The Biological Data Profile was developed simultaneously with the first versions of EML, and several structures from EML were directly incorporated into the BDP, including taxonomic coverage, geologic time, detailed physical structures for ASCII tables, and others. These modifications increase both the comprehensiveness and the flexibility of the BDP, allowing most biological data sets to be described in BDP with a good level of detail. One major limitation is that BDP metadata (and CSDGM by extension) can only reference one data file per metadata record, which prohibits accurate description of complex relational models that are common in environmental sciences. This may not be of concern when documenting high-level metadata, but could be a severe limitation for describing the logical and semantic structure of data.

The BDP is the official format of metadata in the NBII Clearinghouse and has been widely adopted. As previously stated, a conversion to and from EML has been developed, allowing LTER data providers to post over 5000 metadata descriptions in the NBII Clearinghouse using EML metadata that is automatically converted to BDP format.

3.3.2 Shoreline Profile to and Remote Sensing Extension of the CSDGM. The Shoreline Profile was developed by the NOAA Coastal Services Center to accommodate metadata specific to shoreline data, such as information on tides. It retains the requirement for geographic information, and may not be suitable for subtidal data due to requirements for tidal information. The Remote Sensing Extension to the CSDGM provides additional metadata fields for remote sensing activities, such as the geometry of the measurements taken and the properties of the remote sensors used. Remote sensing activities may thus be better described in CSDGM through the use of the metadata fields provided in this extension.

3.4 Open Geospatial Consortium (OGC) Standards

3.4.1 Geography Markup Language (GML). GML was developed by the OGC for the description of GIS data. It provides metadata elements for describing data and geographic features. GML is one of a suite of standards developed by the OGC, including the Web Map Service (WMS) and the Web Feature Service (WFS), which provide methods for transferring and processing geospatial images, and geospatial feature data, respectively. The inherent interoperability of GML with these other standards provides a significant advantage for leveraging the metadata for analysis and management activities. This is evident in the growing number of Web-based tools that enable one to flexibly manipulate and visualize cartographic information. Many of these services could be relying on GML-marked up data. Ironically, however, the highly popular "Google Earth" (http://earth.google.com) does not use GML, but rather its own geospatial file format standard, KML.

For the purposes of the MLPA, GML is limited by its primary emphasis on geographic information, but it is very appropriate for describing data from GIS applications. GML is fully ISO 19115 compliant, and is currently being developed as an implementation of ISO 19115 in XML, as ISO 19136 (Hamre et al. 2004). Once adopted, its use by researchers and institutions is likely to increase. GML has a degree of similarity to both EML and CSDGM, though neither is fully compatible with GML.

3.4.2 Sensor Web Enablement (SWE). SWE is a suite of standards under development by the OGC to facilitate the discovery and use of data from sensor systems and devices. The standards include the Observation & Measurements (O&M) Schema for recording data from sensors, Sensor Modeling Language (SensorML) for describing the properties of the sensors themselves, and Transducer Markup Language for supporting streaming of data to and from sensor devices. SWE should enable sensors and sensor data to be used by a variety of OGC services for receiving notifications and data from, and sending instructions to, sensor systems.

The SWE framework will enable advanced functionality when deploying ground and sky-based sensor systems, and provide for more detailed metadata about these types of devices than other extant metadata standards. These formats may not be appropriate for documenting data that does not come from sensor systems, however, such as data collected directly from observation and measurements by field researchers (e.g. fisheries observers on fishing vessels, visual transects of wildlife diversity or abundance by divers, etc.). The SWE framework is currently under development and not yet an OGC standard. It could become highly useful for achieving close compatibility and data transparency among sensor systems deployed by MLPA, once the framework is finalized and approved, and appropriate technology implementations are available.

3.5 Directory Interchange Format (DIF). DIF was originally developed as a format for exchanging scientific information for what has now grown to become the NASA Global Change Master Directory (GMCD), a clearinghouse for data spanning a variety of disciplines. "[DIF] is simply the 'container' for the metadata elements that are maintained in the IDN database, where validation for mandatory fields, keywords, personnel, etc. takes place. The DIF is used to create directory entries that describe a group of data. A DIF consists of a collection of fields which detail specific information about the data" (DIF 2007). DIF is a standard that minimally contains just eight metadata fields that generically describe a data set, though there are a number of additional optional pieces of information that can be included, and is intended to assist an investigator with determining whether a particular data set is of interest. MEDI uses DIF for its metadata, and OBIS has adopted the standard on a limited basis. These organizations host internet "data portals" where users can perform keyword searches to discover data.

The DIF standard is flexible enough to describe disparate data, although it makes little attempt to describe the format and structure of the data in detail. The approach greatly simplifies metadata entry, which has led to the broad use of the standard by data owners. However, it leaves the data user without a clear understanding of how to interpret the actual data file and perform analyses—a shortcoming that is also apparent in the native profiles (that is, without special extensions) for several of the other geospatially-related metadata standards, such as the CSDGM or ISO 19115. The minimal metadata also makes it difficult to convert metadata from this standard to others, and will lead to information loss over time because detailed information about the data set is not captured by the standard.

DIF was a highly innovative approach to data interchange when it was conceived and approved back in the 1980's, and remains a useful metadata format to this day, due in part to the large number of data sets already documented with DIF. However, the grain of many of DIF's metadata fields are relatively coarse, and permit verbose text explications of features like "access constraints", and lack of structured fields for detailed descriptions of data columns, the units of measurement contained within them, etc. This is likely due to the very different technological landscape when DIF was developed, which was prior to the Web, and in the early days of relational database systems. Modern metadata approaches are typically more ambitious, attempting to enable computational operations on the data, such as automated data integration or subsetting, based on metadata descriptions.

3.6 Network Common Data Format (netCDF) Markup Language (ncML). NetCDF is more of a self-describing file format than standalone metadata standard -- one that is optimized for the description and exchange of array- or grid-oriented scientific data. NcML is an XML syntax for describing the metadata (header) information for data in netCDF format-- so it is also relatively inflexible with regards to data format. The standard is commonly used in the disciplines of climatology and oceanography, however, and has been adopted by SEACOOS. Identification information, basic descriptions, and the logical and physical format of the data are covered by the standard, making it relatively comprehensive for those specialized data types. Metadata is captured in a highly structured format, in part enabled by the data format shared by all data described by this standard. This allows the metadata to be leveraged very effectively by analysis and integration software for netCDF files.

Because netCDF is a binary format, one must employ a netCDF application in order to create or use the data. This is likely to decrease the accessibility of the data to scientists and managers unfamiliar with the data format, or those using applications that do not readily import or export netCDF. However, netCDF is a widespread standard with excellent software library support, so conversion of netCDF files and ncML metadata to other standards and data systems should be straightforward.

3.7 Dublin Core Metadata Initiative (DCMI). DCMI is a basic set of only 15 metadata terms designed to be applicable to all digital resources. This includes metadata for the identification and basic description of data sets. Like ISO 19115, DCMI does not specify a metadata format; it simply indicates the kinds (defined terms) of metadata that should be provided. As a metadata content standard that confines itself to very basic metadata, DCMI is quite flexible, and interoperable with many other standards. In fact, the DCMI has been the foundation for a variety of metadata standards, including EML, ISO 19115, CSDGM, and GML. The accessibility of the standard will vary somewhat according to the actual metadata implementation used. However, the minimal nature of the content tends to make the standard easily accessed by humans, but lacking in power for any advanced machine processing. Recent initiatives to represent DCMI metadata using RDF have allowed for a greater degree of interoperability among metadata standards that are referencing elements from the DCMI.

3.8 Darwin Core and the Ocean Biogeographic Information System (OBIS) Schema. The Darwin Core is a federated model for the description of specimen data within natural history museum collections, which represent primary information sources about the world's biodiversity. It differs from a metadata standard in that it explicitly specifies data content and structure. Put another way, the Darwin Core Standard is for expressing the raw digital data describing the structure and basic descriptive metadata (e.g. taxonomic information) for a particular set of specimens, which are physical objects or artifacts within museum archives. It has proven extremely useful for confederating the various catalogues of information about specimen collections from museums and other institutions around the world, and was endorsed by the GBIF organization as a provisional standard for interchange of this sort of information. This status is currently under review.

OBIS created a data schema that extended Darwin Core to include information more specific to the needs of marine researchers interested in describing patterns of biodiversity in the world's oceans, including for example, repeated observations of a tagged individual, or description of a specimen's "DepthRange". However, recent efforts by OBIS, in conjunction with the GCMD, have used the DIF format for data and metadata, rather than the OBIS data schema.

Many natural history museums continue to use Darwin Core variants as the basis for exchanging information about specimens. Specify, a database program to track natural history collections, is used by hundreds of museums worldwide and can export collections data in Darwin Core format using the DiGIR protocol. Darwin Core is not a comprehensive solution for MLPA use though, as it can only represent a small subset of the types of data to be collected by MLPA researchers.

3.9 Open-source Project for a Network Data Access Protocol (OPeNDAP) metadata. OPeNDAP is a data transfer protocol, widely used by researchers in oceanography and climatology, located at facilities like the National Oceanic and Atmospheric Administration (NOAA), or nodes of the Distributed Oceanographic Data System (DODS). OPeNDAP is not strictly speaking a "metadata standard", but rather a well-defined and rich protocol for data exchange that allows for very effective data querying, integration and subsetting of information across many different data storage formats, as well as diverse and distributed computing platforms and applications. For example, OPeNDAP provides the capability for advanced subsetting and integration of *distributed data* from within popular scientific analysis frameworks like MATLAB, or C programs. This ability is especially useful for the transfer of small subsets of extremely large data sets, and makes OPeNDAP the only standard discussed in this report that routinely enables operations upon data at the level of the individual record instead of the file. OPeNDAP has its own internal data model that describes many scientifically relevant data structures, including arrays and relational tables, among others. It has good support for spatially explicit data matrices through its 'grid' data type. The OPeNDAP protocol also supports other data formats that require some amount of metadata about the physical structure of a data set, including netCDF and HDF file formats.

Since OPeNDAP is more of a data access protocol than a metadata standard in the true sense, it relies upon both standardized data formats and metadata descriptions of the logical structure of data in order to operate. It does not explicitly include the more extensive descriptive, logical, or identifying metadata such as found in EML or the BDP standards. Metadata files for data in the OPeNDAP system could be constructed, however, according to one of the other metadata standards listed here.

The documentation for retrieving data from an OPeNDAP server is generally straightforward, but the ability to install an OPeNDAP server, which is required for sharing data, requires a level of technical expertise probably not possessed by the average researcher. OPeNDAP could be useful to the MLPA as a means of distributing MLPA data from a central server, both within the MLPA and to outside colleagues. For the purposes of facilitating powerful data discovery and interpretation, however, data in an OPeNDAP server would need to be augmented by information stored according to another metadata standard. There is currently an effort underway to clarify how OPeNDAP and EML might complement one another, in terms of a framework that provides greater access to data at the record level, coupled with finer-grain metadata descriptions for interpreting those data (http://reap.ecoinformatics.org)

3.10 Single Institution "Standards". The MLPA requested reviews of the metadata documentation and standardization approaches of many organizations associated with their efforts. These included MARINe, PacFin, RecFin, CalFISH, and others. In reviewing these sites online, however, one discovers that many of these data systems have made little effort with regards to exposing their metadata, or adhering to metadata standards.

Instead, most of these organizations have gone to the effort of constructing detailed relational databases for their data, based on the particular needs of their users, and well known properties of their data. Thus, these databases are not, nor do they likely contain, metadata standards or systems. Such approaches can be very useful if all collaborators can initially agree on an overarching design which successfully services their specific analytical needs and interpretations of the data. But essentially, they are cocooned within their own framework—the data model only pertains to what they have defined beforehand, and such systems can be difficult to modify and extend as these needs change. Moreover, effective integration with external data sources also becomes difficult, in terms of reconciling semantics, scaling and unit issues, etc.

Often what might be considered as "metadata" is captured in a relational database, or even a simple table, as a repeating value of some (probably) idiosyncratically named variable. In this case, a "raw data" value might not be evident as metadata, simply because it is stored amongst the raw data with no indication of its global relevance for the data set. For example, "geospatial information" might be included in a column that is called "SITE", and the value of SITE might be an abbreviation (e.g. "PBSMR") that needs to be joined with information contained in some other table in order to access more detailed descriptions of the site, such as its full name (e.g. “Piedras Blancas State Marine Reserve”), latitude and longitude, etc. This demonstrates the problem with transparency of those data to users who are not already intimately familiar with the particular database. These “metadata” can be most effectively “exposed” to outside researchers by mapping the values within the database to well-established metadata fields. For example, the concept of who is the collector of some set of data might be captured in each of a number of individual database solutions as “raw data” (or it might not, due to oversight). But in any case, understanding how that information was captured (by referencing only surname? initials? full contact information?), and in which fields of the database (the column called “owner”? “tech”?) would require careful perusal of the database structure and contents. Here it becomes clear that standardized metadata terms can serve as the common set of labels to facilitate discovery and interpretation *across* data resources being collected by various groups, whether these are stored as single tables, or complex relational databases, etc. Standards can also help assure that important metadata information is indeed captured, as well as consistently interpreted.

As it is, for many single institutions or single project efforts, the types and formats of the “metadata” are extracted from the database in text and diagrams on the websites of the organizations, or collected in some wordprocessed document which might typically be called the data catalogue. These descriptions and documents, if they exist, represent the “metadata” for the data holdings of these organizations. These are not going to be very useful or efficient for computer-assisted data discovery or access, due to their idiosyncratic structures and semantics; and even less useful for any advanced computation such as data subsetting or integration, for the same reasons. Moreover, for those not already quite familiar with the data, finding these ancillary documents, and then properly interpreting them, will be quite challenging.

4 RECOMMENDATIONS

The final recommendation of a metadata standard for MLPA data collections depends in part on (1) the level of metadata desired by the MLPA for their baseline data, (2) the amount of time investment the MLPA wishes to impose on the owners of the baseline data, and (3) the amount of leverage the MLPA can exert on data owners to comply with the selected standards. Below we consider two cases: one which requires minimal investment by data owners to create basic data descriptions, and one which requires a high level of metadata content to fully describe the data structure.

4.1 Recommendations for Basic Metadata Descriptions. This scenario allows for the creation of a data registry, in which basic metadata descriptions are supplied for data, without documenting the logical or physical structure in detail. This would enable data users to search data holdings on keywords and other basic criteria, and potentially be very effective for discovering interesting or complementary data to inform a given analysis. Without further effort, this information would be of limited value for guiding data analysis and other such efforts. The advantage of this system, however, is that with minimal time investment by data owners, the awareness of what types of data people are collecting, and in which regions, etc. would become very evident. This capability might facilitate increasing acceptance of the system by researchers, while providing a global catalog of all MLPA-relevant data.

Adopt the EML Standard. EML provides all the necessary content and structure for creating basic metadata entries. Though EML is capable of handling very detailed metadata, most fields are optional and thus it can also be used to enter only very basic metadata. Other extensive standards, such as CSDGM, are not capable of accepting a minimal level of detail in this manner without violating the standard (i.e. many more fields are required in CSDGM). Other standards for entering basic metadata, such as DCMI and DIF, do not have the capacity for housing detailed metadata descriptions alongside more basic entries. Using EML has the benefit of capturing basic metadata records for all MLPA data holdings with minimum cost, while allowing those who are interested to create more detailed entries. In addition, EML imposes no restrictions on the formats or types of data described. EML also has basic identification metadata that closely mirror the DCMI elements. For more specialized data, like specimen collections, adoption of the Darwin Core metadata might also be highly useful.

An advantage of EML over DIF and DCMI is the existing suite of free, production-ready software tools that are customized to operate on EML. Metacat, a server program for storing EML metadata and entries, comes packaged with web registry forms that facilitate the process of metadata creation and maintenance. It also provides web-based search functionality based on a variety of criteria. Morpho, a full-featured EML editor, could be used by researchers desiring more metadata content than is provided by the web forms. The availability of these tools makes EML a very strong candidate for adoption by the MLPA in this scenario. Close compatibility and tools for bi-directional conversion between EML and the NBII's BDP, will also be advantageous to many organizations (see below).

Support for DIF. Since several data clearinghouses relevant to the MLPA use DIF, it is important that DIF be supported by the system. Although a crosswalk from EML to DIF does not currently exist, given the relative simplicity of the DIF format, creating one should not be difficult. Once this was completed, metadata from the MLPA system could be automatically registered with these other systems, including the GCMD and MEDI. Correspondences between the research and data needs of MLPA and the data being collected by the OBIS community, which uses both Darwin Core and GCMD/DIF, might also be well worth closer examination.

Develop Metadata Guidelines. Given the flexibility of EML, the MLPA should strongly consider developing a set of guidelines for the metadata content to be included in entries on the system. These guidelines should include the use of controlled vocabularies, such as GCMD keywords or emerging ontologies, where appropriate, as well as clear explanations of which metadata fields should be used, and examples of how those fields should be populated. Since many of the systems the MLPA wishes to interface with require CSDGM metadata, such as BIOS, IOOS, as well as all data from federal agencies, the guidelines could also specify which additional metadata fields would be necessary to allow metadata entries to be converted from EML to CSDGM, and vice versa.

Accommodating Single Institution "standards". The recommendations above presuppose that a number of different types of data will be collected under the auspices of MLPA, both within MLPA and externally, and we recommend a standardized metadata approach for documenting these. There are several options, however, for the MLPA to achieve greater compatibility with external organizations having their own individualized approaches for metadata, and these might be particularly useful if MLPA researchers are already aware of valuable, external data sets which they want to tightly confederate into their framework. There will also probably be opportunities to reduce the structural and logical heterogeneity among data sets within MLPA-sponsored efforts *per se*, through careful planning for how the data should be collected and stored.

While standardized metadata approaches are invaluable for facilitating interpretation and discovery across diverse data sets, the adoption of standardized data models can reduce this heterogeneity at the outset. Developing and promoting well-defined data models for the collection of data motivated by the same research questions, but potentially collected by various parties at different times and places, will immediately provide the benefit of close integration of those data at the individual record level. The adoption of standardized data collection protocols and models (as described in choice 1 below) should be encouraged for highly similar data collection efforts, whenever this does not impede MLPA progress or compromise research goals. Three approaches, which are not mutually exclusive, are:

1) Adopt specific, detailed data models used by other organizations where appropriate, or develop these within the MLPA. This option involves carefully prescribing protocols for data collection and reporting, and thus falls outside the scope of any specific metadata standard, although several of the metadata standards described in this report could capably provide documentation about these protocols. With this approach, data collected according to a certain data model could all be easily integrated, but combining data collected using different underlying data models will still be difficult. Development of standard data models and data collection protocols is recommended whenever there is a need for a highly coordinated effort, intended to address identical, focused research questions, even if these are spatially or temporally distributed.

2) *Adopt a more generalized data model, and then use it to map to the data models of these outside organizations*. This will afford the same discovery and integration benefits to the data housed by outside organizations as to those within the MLPA, but these more general models often do not permit the type of tight integration that researchers need to do synthetic analyses—details are often lost. This approach also often involves an arduous mapping for every new data source that one wants to integrate with the system.

3) *Adopt a metadata standard and expect any potential collaborators to also describe their data according to that standard, or have MLPA researchers create compliant metadata for the outside data systems*. This last possibility will provide the greatest flexibility in accommodating at least effective data discovery and access across highly diverse, and potentially autonomous projects.

Monitor Emerging Standards for Ontologies. Ontologies have the potential to greatly simplify data search, analysis, and integration in the future. Developing standards such as OBOE, SWEET, and the ontologies in the Marine Metadata Initiative Ontologies Repository should be periodically reviewed to assess their maturity and applicability to MLPA data. In the case of the OBOE ontology, EML metadata will be a primary target for enrichment via annotation, and software tools to facilitate the annotation process are currently under development.

4.2 Recommendations for Structural Level Metadata. This scenario considers the development of a metadata system in which the physical and logical structure of data is described for MLPA baseline data. While this places a greater burden on data owners to provide this information, it ensures better preservation of the data, and can provide better computational access to the data via analytical tools and software. Given the importance of MLPA baseline data and its potential for reuse in long-term monitoring studies and other research, this option should be strongly considered for adoption by the MLPA.

All the categorical recommendations in the above section also obtain within this more advanced scenario. While both EML and BDP provide the ability to capture metadata about the logical and physical structure of data, EML has a finer-grained structure for describing these sorts of metadata. These finer-grained metadata fields provide for potentially powerful ways of accessing the data using computer software, as detailed descriptions of a data set's physical and logical structure are usually critical for developing automated applications that can appropriately operate upon the data. The finer-grained structure in EML also means that EML can be converted to BDP in an automated fashion; and tools to accomplish this exist. The reverse conversion from BDP to EML can not be completely automated because BDP lacks some detailed information that EML contains, so BDP's coarser-grained structure inhibits seamless conversion. By documenting datasets in EML, researchers will be able to obtain both EML and BDP compliant versions of their metadata, effectively eliminating the need to choose between the standards. This will allow data sets to be used in applications, data clearinghouses, and other situations that require either of these two formats. If, on the other hand, data sets were described initially using BDP, they would be unable to be used in situations that require EML. While BDP and CSDGM arguably have greater acceptance among relevant oceanographic and environmental institutions at this time, EML will facilitate integration with these institutions' metadata standards, as well as with other institutions that use EML itself, such and LTER, PISCO, and the KNB. EML is also actively being investigated as to its use with advanced knowledge processing tools, such as annotation by ontologies.

Other standards, such as ncML, GML and SWE, that require specific data formats, types, or collection methods, may prove excessively restrictive to data owners who are not familiar with those formats, or are engaged in efforts that do not have a primary geospatial component in their analyses. However, to the extent that the MLPA anticipates significant usage of GIS technologies, it will be worth examining the utility of adopting GML-compliant applications, and potentially requiring GML, ISO 19115, or one of the other emerging geospatial metadata standards, for any GIS coverage information. The MLPA might also consider the use of the CSDGM or BDP standards for data with a significant GIS component, as many of the commonly used GIS software platforms automatically generate CSDGM metadata. Similarly, as the SWE framework is finalized, the MLPA should investigate its utility for documenting data from various sensor network systems.

The MLPA wishes to conduct a variety of analyses with its baseline data. Kepler, a scientific workflow program currently in a "beta-stage" of development, could potentially provide an ideal environment for performing analyses of data described with EML (http://www.kepler-project.org). A Metacat server could be set up specifically for

MLPA metadata, or MLPA metadata could be loaded into existing Metacat servers on the KNB. Morpho can facilitate individual researchers documenting their data with EML. The existence of these tools further supports the recommendation for the adoption of EML as the metadata standard for MLPA baseline data. Workflow solutions, which are capable of using metadata standards for powerful data discovery and integration, represent an exciting area of investigation in scientific computing at this time. Other free and commercial workflow engines are currently under development, which might also be viable solutions for MLPA. This is an area of interest that could benefit from continued evaluation, and further research on MLPA's behalf.

Develop Metadata Guidelines. Given that physical structure and other metadata in EML is optional, the MLPA will need to develop guidelines for the metadata that should be included for baseline data. The EML fields already available in the Morpho application could provide a good starting point for discussion. The metadata recommendations of the IOOS Data Management and Communications (DMAC) Meatadata and Data Discovery Expert Team (IOOS DMAC 2005) can also provide guidance when developing these guidelines. Because it is difficult for researchers to recall or obtain important metadata to be added to an entry at a later date, it is suggested that the guidelines err on the side of including more metadata rather than less, during initial population of the data catalog.

Support BDP and CSDGM. Many institutions and systems of specific interest to the MLPA, including IOOS, BIOS, and the NBII Clearinghouse, use the BDP and CSDGM standards for their metadata. The MLPA will therefore want to make the crosswalk from EML to BDP readily available to researchers. Alternatively, the MLPA could use the EML to BDP crosswalk to make available BDP versions of all MLPA metadata, possibly publishing these metadata in clearinghouses such as NBII. This will allow metadata in the MLPA system to be used in systems requiring BDP metadata with minimal effort. If the MLPA also wishes to interface with systems such as GCMD, OBIS, and MEDI, it should consider developing a crosswalk from its metadata standard to DIF as well.

Monitor Emerging Standards for Ontologies. Even moreso than in the prior section, "intelligent" and automated discovery and processing of data, and the construction and documentation of executable analyses, will be based on advanced semantic approaches that enable computers to more capably interpret and operate on data. Ontologies, and formalized frameworks for expressing rich conceptual models of the underlying data, are currently active areas of investigation within the field of computer science and informatics. These represent the cutting edge of how database technologies are merging with Web applications and analyses. While immediate impacts of these advances are unlikely to be experienced by MLPA researchers, the solutions adopted in the present can influence future possibilities of how and whether the MLPA informatics solutions will be able to take advantage of these technologies as they mature. The MLPA should be alert to emerging trends in technology, and establish ongoing associations with technologists who are informed in these areas, but also understand the needs and goals of the MLPA.

5 REFERENCES

(For references to the websites of individual metadata standards, please see Appendix C.)

Berners-Lee, T., Hendler, J. and Lassila, O. 2001. The semantic web. Sci. Am. 284(5):34-43.

Fegraus, E. H., Andelman, S., Jones, M. B., and Schildhauer, M. 2005. Maximizing the value of ecological data with structured metadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation. Bull. Ecol. Soc. Am. 86(3): 158-168.

DIF. DIF writer's guide. http://gmcd.nasa.gov/User/difguide/difman.html . (March 2007)

Green, J. L., Hastings, A., Arzberger, P., Ayala, F. J., Cottingham, K. L., Cuddington, K., Davis, F., et al. 2005. Complexity in ecology and conservation: Mathematical, statistical, and computational challenges. Bioscience. 55(6): 501-510.

Gross K. L. and Pake C.E., eds. 1995. Final Report of the Ecological Society of America Committee on the Future of Long-term Ecological Data (FLED) Vol. I: Text of the Report. 123pp. http://intranet.lternet.edu/archives/documents/other/fledvol1.pdf.

Hamre, T, Keiran, M. and Matthews, B. 2004. Analysis of marine and XML data standards. EU MarineXML Project. MarineXML_D6_v5 (17/08/2004). http://www.iode.org/marinexml/files/MarineXML_D6_v5.pdf . (March 2007).

IOOS DMAC. 2005. Data management and communications plan for research and operational integrated ocean observing systems: interoperable data discovery, access, and archive. National Office for Integrated and Sustained Ocean Observations. Pub. no. 6.

ISO. 2003. Geographic information -- metadata. ISO 19115:2003. Available http://www.iso.org . (March 2007).

Isenor, A. W. 2005. Final report of the ICES/IOC study group on the development of marine data exchange systems using XML. http://www.iode.org/marinexml/files/SGXML_final.pdf . (March 2007).

Jackson, J. B. C., Kirby, M. X., Berger, W. H., Bjorndal, K. A., Botsford, L. W., Bourque, B. J., Bradbury, R. H. et al. 2001. Historical overfishing and the recent collapse of coastal ecosystems. , Science. 293: 629-638.

Jones, M. B., Schildhauer, M. P., Reichman, O. J., and Bowers, S. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Ann. Rev. Eco. Evo. & Sys. 37: 519-544.

Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D. and Villa, F. In Review. An ontology for describing and synthesizing ecological observation data. Ecol. Info.

Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B. and Stafford, S. G. 1997 Nongeospatial metadata for the ecological sciences. Ecol. Appl. 7:330-342.

Open Geospatial Consortium, Inc. 2006. OGC Sensor Web Enablement: Overview and High Level Architecture. Available: http://www.opengeospatial.org/pt/06-046r2 . version 2.0

Appendix A: Glossary of Acronyms

BDP	Biological Data Profile (http://metadata.nbii.gov)
BIOS	Biogeographic Information and Observation System (http://bios.dfg.ca.gov)
CalOST	California Ocean Science Trust (http://calost.org)
CDFG	California Department of Fish and Game (http://www.cdfg.ca.gov)
CeNCOOS	Central and Northern California Ocean Observing System (http://www.cencoos.org)
CSDGM	Content Standard for Digital Geospatial Metadata (http://www.fgdc.gov/metadata/csdgm)
DCMI	Dublin Core Metadata Initiative (http://dublincore.org)
DMAC	Data Management and Communications (http://dmac.ocean.us)
DIF	Directory Interchange Format (http://gcmd.nasa.gov/User/difguide/difman.html)
DODS	Distributed Oceanographic Data System
EML	Ecological Metadata Language (http://knb.ecoinformatics.org/software/eml/)
FGDC	Federal Geographic Data Committee (http://www.fgdc.gov)
GCMD	Global Change Master Directory (http://gmcd.nasa.gov)
GIS	Geographic Information System (http://www.gis.com)
GML	Geography Markup Language (http://www.opengeospatial.org/standards/gml)
IOOS	Integrated Ocean Observing System (http://www.ocean.us)
ISO	International Standards Organization (http://www.iso.org)
KNB	Knowledge Network for Biocomplexity (http://knb.ecoinformatics.org)
LTER	Long-Term Ecological Research Network (http://www.lternet.edu)
MARINe	Multi-Agency Rocky Intertidal Network (http://www.marine.gov)
MEDI	Marine Environmental Data Index (http://ioc.unesco.org/medi/)
MLPA	Marine Life Protection Act (http://www.dfg.ca.gov/mrd/mlpa)
MPA	Marine Protected Areas
NBII	National Biological Information Infrastructure (http://www.nbii.gov)
ncML	netCDF Markup Language (http://www.unidata.ucar.edu/software/netcdf/ncml/)
netCDF	Network Common Data Format (http://www.unidata.ucar.edu/software/netcdf/)
OBIS	Ocean Biogeographic Information System (http://www.obis.org)
OBOE	Extensible Observation Ontology (http://seek.informatics.org)
OGC	Open Geospatial Consortium (http://www.opengeospatial.org)
OPC	Ocean Protection Council (http://resources.ca.gov/copc)
OPeNDAP	Open-source Project for a Network Data Access Protocol (http://resources.ca.gov/copc)
ORION/OOI	Ocean Research Interactive Observatory, Ocean Observatories Intiative (http://www.orionprogram.org)
OWL	Web Ontology Language (http://www.w3.org/2004/OWL)
PacFIN	Pacific Fisheries Information Network (http://www.psmfc.org/pacfin)

PaCOOS	Pacific Coast Ocean Observing System (http://www.pacoos.org)
PISCO	Partnership for Interdisciplinary Study of Coastal Oceans (http://www.piscoweb.org)
RDF	Resource Description Framework (http://www.w3.org/RDF)
SCCOOS	Southern California Ocean Observing System (http://www.sccoos.org)
SWE	Sensor Web Enablement (http://www.opengeospatial.org/projects/groups/sensorweb)
SWEET	Semantic Web for Earth and Environmental Terminology (http://sweet.jpl.nasa.gov/)
UCNRS	University of California of Natural Reserve System (http://nrs.ucop.edu)
UML	Unified Modeling Language (http://www.uml.org)
XML	Extensible Markup Language (http://www.w3.org/XML)

Appendix B: Metadata Requirements

Criteria	Discovery	Access	Analysis	Data Loading
2.1 – Content and Structure				
2.1.1 - Identification	Required	Required	Required	Required
2.1.2 - Description	Required	Required	Required	Required
2.1.3 - Physical Structure	--	Helpful	Required	Required
2.1.4 - Logical Structure	Helpful	--	Required	Required
2.1.5 - Semantic Structure	Helpful	--	Helpful	Required
2.2 - Flexibility	Optimal	Optimal	Required	Required
2.2.1 - File Formats	Optimal	Optimal	Required	Required
2.2.2 - Data Types	Optimal	Optimal	Required	Required
2.2.3 - Databases	Helpful	Helpful	Optimal	Required
2.2.4 - Spatial Data	Helpful	Helpful	Optimal	Required
2.3 - Accessibility				
2.3.1 - Human Access	Required	--	Required	Required
2.3.2 Machine Access	Helpful	Required	Optimal	Required
2.4 - Stability	Helpful	Optimal	Optimal	Required
2.5 - Interoperability	Helpful	--	Helpful	Optimal
2.5.1 Metadata Systems	Helpful	--	Helpful	Optimal
2.5.2 - Data Systems	--	--	Helpful	Optimal
2.6 - Existing Tools	Helpful	Optimal	Helpful	Required

Key

Required - Criterion is necessary to perform the desired function.
Optimal - Criterion is not an absolute requirement, but significantly improves the desired functionality.
Helpful - Criterion provides auxiliary information that in some cases aids in performing the desired function.

Appendix B:

mpa MONITORING ENTERPRISE

CALIFORNIA

MPA Monitoring Information Management System

USER NEEDS ASSESSMENT

OCTOBER 2010

Version 1: March 2016

About this Document

This report was prepared by Exa Data and Mapping Services, Inc., a consultant team selected to conduct a user needs assessment to inform development of an online marine protected areas monitoring information management system. California's Marine Life Protection Act requires establishment of a statewide marine protected areas network, and monitoring to ensure the network is meeting its goals and to inform future management decisions. The information management system will steward and share monitoring data, results, and associated information. This report, and a summary, *MPA Monitoring Information Management System User Needs Assessment: In Brief*, are available from the Monitoring Enterprise website.

Acknowledgements

We are grateful to Peggy Myre and the rest of the Exa team for their hard work and creativity in designing and conducting the user needs assessment. We thank Elizabeth Chornesky for her advice and assistance throughout the project and in finalizing this report. In addition, we thank the many individuals who participated in the user needs assessment for their time and thoughtful comments. The assessment was made possible by the support of the California Ocean Protection Council, the California Ocean Science Trust, the David and Lucile Packard Foundation, and the Resources Legacy Fund Foundation.

Citation: *MPA Monitoring Information Management System User Needs Assessment*. MPA Monitoring Enterprise, California Ocean Science Trust, Oakland, California, USA. October 2010.

About the MPA Monitoring Enterprise

The MPA Monitoring Enterprise was created in 2007 to lead the design and implementation of science-based, impartial and cost-effective monitoring of and reporting on the network of marine protected areas established in California under the Marine Life Protection Act. We develop monitoring that assesses and tracks the condition of ocean ecosystems and evaluates the effects of marine protected area design and management, in order to evaluate the performance of marine protected areas in meeting policy goals and inform future management decisions. We work closely with the California Department of Fish and Game and the California Ocean Protection Council and engage scientists and stakeholders to ensure monitoring is based on the best available science, reflects public interests, and meets management needs.

CALIFORNIA OST Ocean Science Trust
Connecting Science to Ocean Management

The MPA Monitoring Enterprise is housed within the California Ocean Science Trust, a non-profit organization established pursuant to the Coastal Ocean Resources Stewardship Act of 2000 to provide scientific guidance to the state on ocean policy issues. More information about the MPA Monitoring Enterprise can be found at monitoringenterprise.org.

Version 1: March 2016

mpa
MONITORING
ENTERPRISE

MPA Monitoring Information Management System User Needs Assessment

Developed to support implementation of California's Marine Life Protection Act

October 2010

Version 1: March 2016

SUMMARY

The user needs assessment (UNA) described in the enclosed report was conducted to inform the design of a future on-line monitoring information management system (IMS) for California marine protected areas. The State of California is currently engaged in establishing a statewide MPA network. Monitoring of the MPA network is required under state law, and will generate a large quantity and wide variety of data, analyses, reports and associated information that should be made available to decision-makers and the public. The IMS will house and provide access to that data and information, but must be designed according to the needs of its users to be effective.

Through the UNA, the views of individuals reflecting the diverse and geographically dispersed potential user population for the IMS were solicited. The UNA explored the perspectives of potential system users related to the types of information that are of interest, how that information should be synthesized and presented, which IMS features and functions are of highest priority, and institutional considerations. The results of the UNA have clear implications for the design of the future IMS, as is described briefly below, and in detail in the enclosed report.

Background

The 1999 California Marine Life Protection Act (MLPA, Chapter 10.5 of the California Fish & Game Code, §2850-2963) directs the state to complete a statewide network of marine protected areas (MPAs). The MLPA also requires monitoring of MPAs to facilitate adaptive management of MPAs and ensure that the MPA network meets the goals of the Act.

The statewide MPA network is being implemented through a regional approach. Five regions have been defined: North Coast, North Central Coast, Central Coast, South Coast, and San Francisco Bay. As of October, 2010, MPAs have been implemented in the Central and North Central Coast regions, are undergoing regulatory review in the South Coast region, and are in the planning stage in the North Coast region. Work in the San Francisco Bay region has not yet begun.

The MLPA Initiative, a public-private partnership, works with stakeholders and scientists to develop MPA network proposals in each region, and submits recommendations to the California Fish and Game Commission to consider for legal designation. In each region, the MLPA Initiative establishes a Science Advisory Team (SAT) to provide technical input on MPA proposals, and a Regional Stakeholder Group (RSG) to provide broad stakeholder representation. The MLPA Initiative has also established a Statewide Interests Group (SIG), composed of members from key interest groups around the state, to provide input on MLPA Initiative implementation. (For more information on the MLPA Initiative, see www.dfg.ca.gov/mlpa.)

The MLPA Initiative's regional SATs and RSGs, and the SIG, as key participants in implementing the MLPA, are also likely to be interested in MPA monitoring. Similarly, state agencies directly or indirectly involved in MPAs are also key audiences for monitoring information. These include the California Department of Fish and Game (DFG), the agency with statutory authority for implementing the MLPA, the California Fish and Game Commission, the decision-maker under the MLPA, and others. The MPA Monitoring Enterprise works with members of these and other groups and organizations to lead development of efficient, cost-effective MPA monitoring for each region after MPA planning is completed.

The MPA Monitoring Information Management System

Monitoring to assess the effectiveness of the emerging statewide MPA network in meeting MLPA goals will entail the collection of many different types of data, including ecological data about species and habitats, for example, and socioeconomic information about human activities in and around the MPAs. The MPA Monitoring Enterprise will lead analyses of monitoring data to generate summary reports useful for assessing MPA network effectiveness and informing future management decisions.

The future MPA monitoring IMS must make the data, analyses, reports, and other relevant information readily available to MPA decision-makers, managers, stakeholders, and scientists, as well as the broader interested public. To determine how best to present this broad array of types of information to diverse audiences, it is critical to begin with a thorough assessment of the needs and priorities of the future IMS users. This UNA is thus the first step in ensuring the IMS will be useful and effective.

User Needs Assessment Approach

The UNA was designed to seek opinions from individuals expected to have a high level of interest in the state's marine protected areas (MPAs), including members of the regional RSGs, SATs, and the SIG, and from all five of the MLPA regions. Questions posed in the assessment were designed to solicit opinions relevant to designing the future IMS. These questions on preferred information types (e.g., raw data, maps, summaries), information topics (e.g., ecosystem types, human activities in and around MPAs, general MPA information), system tools (e.g., data download tools, data analysis tools, mapping capabilities), and other possible contents or functions of the system.

Internet surveys and telephone interviews were employed to solicit opinions from a broad array of potential users of the future IMS from across California. Invitations to participate in the internet survey were sent to 412 people and an additional 49 were invited to participate in telephone interviews. In addition, an open invitation to participate in the internet survey was extended to the 2,376 subscribers to the MLPA Initiative listserv. In total, 519 people completed the internet survey, and 35 people were interviewed by telephone.

Key Findings

While respondents' opinions varied, it was nonetheless possible to identify clearly recurrent themes. For example, many respondents requested that the IMS provide access to raw MPA monitoring data, not only to allow independent analyses, but also ensure transparency and credibility. Another clear recommendation was to provide monitoring information at a variety of spatial scales, ranging from individual MPAs to the state as a whole. The report describes many other such findings, and the similarities and differences in the views of respondents with different MPA interests.

In addition, four patterns of user preferences were identified based on desired levels of information synthesis (ranging from raw data to highly summarized reports) and system interactivity (ranging from a simple site with simple search functions to a highly interactive site with complex analytic tools). These four patterns, termed user personas, have important implications for the design of the IMS that are discussed in the final section of the report.

Next Steps

This report provides a wealth of information to consider in designing the future MPA monitoring IMS, and illustrates some likely trade-offs. The next step is to begin to flesh out technical specifications and system requirements that would reflect the user needs and priorities identified through the UNA, and further explore key potential decision points in system design.

We intend to adopt an 'agile' approach to designing and building the IMS, identifying a sequence of steps that will incrementally add value and function to the system. This will allow the IMS to be built, tested, and refined, in stages, and adjustments to be made along the way to ensure usefulness and efficiency. Of course, the ultimate judges of the usefulness of the IMS will be its users – the decision-makers, resource managers, stakeholders, and scientists the system is meant to serve. We will use a variety of tools and mechanisms to evaluate this along the way.

We thank the many participants in the UNA and hope we will continue to hear from people interested in California's MPAs and their future management about how to make the IMS effective.

California Marine Protected Areas Monitoring Information Management System – User Needs Assessment

Final Report

April 2010

Prepared for:

Marine Protected Areas Monitoring Enterprise

through the
California Ocean Science Trust

Prepared by:

Data & Mapping Services

Exa Data & Mapping Services, Inc. Team:

Sound GIS
CoastalVision, LLC
T.C. Hoffman and Associates, LLC
Adita Agrawal

Table of Contents

LIST OF FIGURES

LIST OF FIGURES, CONT.

LIST OF TABLES

LIST OF TABLES, CONT.

APPENDICES

Acknowledgements

Thank you to Elizabeth Chornesky for your detailed review and edit, and for providing useful and interesting insight into the results; Cheri Recchia for additional review and comprehensive oversight of the project; Drew Carey and Allison Bailey for providing data analysis, as well as thought-provoking discussion and perspective; Tegan Hoffman for guidance in the project design and planning; Aditya Agrawal for needs analysis experience and GIS support; Ezra Beaver for launching and managing the internet survey results; and Kimberly Galimanis for graphics support.

1. Introduction

Exa Data and Mapping Services, Inc. (Exa) conducted a User Needs Assessment (UNA) by surveying people potentially interested in scientific monitoring information collected to assess California's marine protected areas. The project was conducted for the Marine Protected Areas (MPA) Monitoring Enterprise, under the auspices of the California Ocean Science Trust. The Monitoring Enterprise was created to lead the effort to monitor MPAs as required by the Marine Life Protection Act (MLPA). Information derived from the UNA will assist in developing an information management solution for storing and providing access to the MPA monitoring data and associated information. The project's purpose is to ensure that the needs of the people with vested interest in the data, and those most likely to want access to the information, are understood and considered in making system design choices.

A UNA provides a process for discovering and evaluating user needs and desires for information in order to develop an information management system (IMS) to meet those identified needs. Thus, user needs drive the specifications for the IMS, rather than the more typical 'build it and they will come' approach. This *user-centered design* shifts the focus from the designer of the system to the user of that system, thereby improving the usability of the data and the utility of the applications.

Information collected in this project will be used by the MPA Monitoring Enterprise for a subsequent system requirements analysis, and ultimately will inform decisions about what existing or new systems should be integrated into a final solution to deliver monitoring data and information in an effective and efficient manner. Because the UNA is the first component of a full requirements analysis – and thus will lay the foundation for subsequent assessments of the system's desired functional capabilities and related technical specifications – this document does not focus on specific technical solutions or recommendations. Instead, it reports key findings and their implications within the more general context of information management and system design.

1.1 Goals and Challenges of the User Needs Assessment

The initial step of the UNA was for Exa to work with the MPA Monitoring Enterprise to define the anticipated scope of the IMS. The IMS must be able to give people useful information that will help them reach their own goals. As a corollary, the system must deliver information synthesized from the monitoring data, helping to bridge the gap between science and policy. This synthetic information should be institutionalized for monitoring-related activities, but also generalized for other users to diversify and deepen the longer-term usefulness. The IMS will be internet-based, although no conclusions have been made as to the optimum infrastructure or platform, or as to whether this is a new system or will be integrated with an existing infrastructure.

The UNA sought to survey all types of potential users because of the wide variety of interests involved in establishing and adaptively managing California's MPA system. This goal presented a challenge in designing the survey because of the diverse experiences and backgrounds of the potential survey population. The audience for the IMS includes not only decision-makers, resource managers, stakeholders, and scientists, but also members of the general public. Thus, the survey group consists essentially of anyone with an interest in the state MPAs. The user-centered design philosophy adopted in this UNA and by the MPA Monitoring Enterprise required carefully characterizing, prioritizing, and assessing user roles in order to specify the audience that would be invited to participate in the UNA.

The solution was to identify the full range of ways in which people are associated with MPAs. These were then assembled into *archetypes*. Archetypes categorized potential users by their role, association, or familiarity with MPAs, such as resource manager, stakeholder, etc. The classification was based only on role, interest, job function, or other demographic information; it was independent of any assumed preferences for, or interaction with, the monitoring information itself. Once the archetype list was created (Section 1.2), it provided a framework for ensuring that the UNA included representatives of each archetype (Section 2.1). The archetype categories also provided a means of analyzing the UNA data to assess whether different potential user groups had significantly different needs (Section 3).

The UNA also sought to identify the implications of user needs for the IMS design, costs, and tradeoffs among design options. Certain survey questions addressed these issues – such as those related to desired types of analytical tools and interaction modes – thereby enabling clear links between the survey topics and specific elements of the IMS framework that can usefully be informed by the survey findings. This approach should help ensure that the UNA supports design choices about the IMS in addition to informing criteria for assessing its success. Section 1.3 of this report (Designing the Assessment) describes the IMS framework, the relationship of the framework elements to the UNA survey topics, survey instruments chosen for the project, and an overview of how the UNA data were processed.

1.2 Defining User Roles: Archetypes

Defining the archetypes proved to be one of the more challenging aspects of the UNA. The archetype list evolved in parallel with the development of lists of people who would be invited to take the survey or be interviewed, and the analysis of the survey and interview data (Figure 1-1). This process was initiated at the first kickoff meeting of the MPA Monitoring Enterprise and Exa and was finalized only after all of the data were collected.

Initially, we defined archetypes in terms of their assumed need for and interaction with monitoring information. We subsequently decided that this reasoning was circular, because the purpose of the UNA was to determine the user's need for information without pre-conceived bias. Therefore, we re-defined archetypes independent of their assumed or expressed need for information, and based entirely on the user's role, association, or familiarity with MPAs. However, the limitations of this archetype categorization became apparent with the completion of the online survey, as many respondents could easily fit into more than one archetype. Therefore, during the data analysis phase, a set of objective rules was developed to establish the primary role of each respondent in order to best assign them to one of the archetype categories.

The careful attention to archetypes and geography shown in Figure 1-1 reflected the MPA Monitoring Enterprise and Exa's commitment to ensuring the full range of potentially interested parties was integrated into the UNA and to avoid bias in the results or their interpretation. The MLPA requires implementation of a statewide MPA network. The MPAs are being planned on a region-by-region basis through the MLPA Initiative, which engages a wide diversity of MPA interests and perspectives. MPA monitoring must be similarly responsive to this diversity. Archetypes are simply a tool for accomplishing this purpose and do not reflect any underlying assumptions about the particular perspective of any given survey respondent. The results section (starting p. 12) includes a summary of the original distribution of archetypes invited to take the UNA survey and the actual distribution of archetypes that participated.

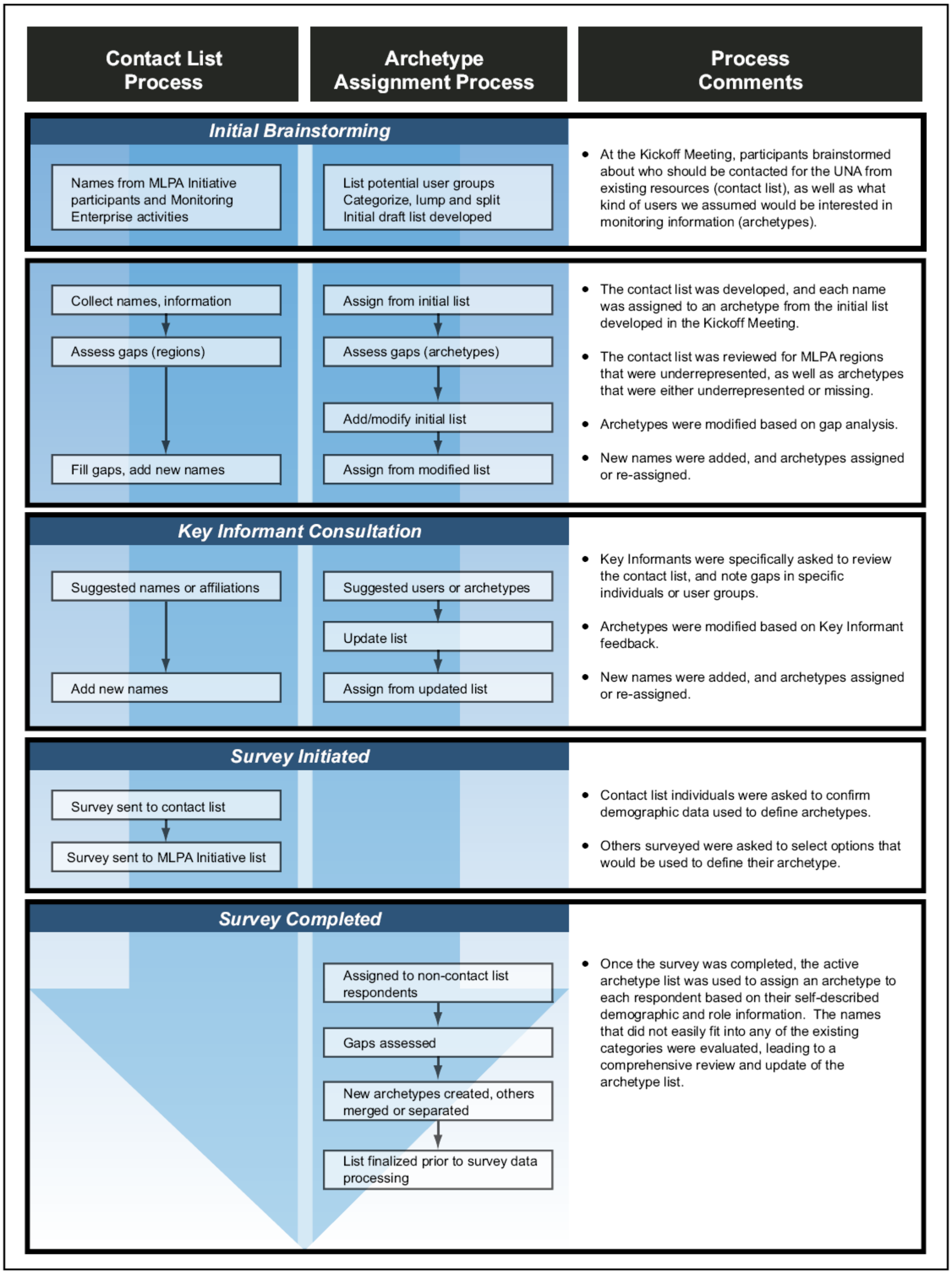


Figure 1-1. Flow chart documenting the parallel processes of developing the contact list of those who would be invited to participate in the survey, and of creating the archetype list. Activities are ordered in time from top to bottom.

The final categories used in the archetype assignment process included the following, described more fully below:

- Decision Makers
- Resource Managers
- Policy Informers or Influencers
- Scientists
- Information Managers
- Educators
- Stakeholders
    - Commercial Fishing
    - Recreational Consumptive
    - Recreational Non-Consumptive
    - Environmental Non-Governmental Organization (ENGO)
    - Local Coastal Manager
    - Military
    - Tribal
- Students
- Citizens
- Miscellaneous

Decision-Makers – People with decision-making authority related to MPAs. Examples included individuals that make decisions at both state and federal levels, including members of the California Fish and Game Commission and legislators who make high-level management or policy decisions about MPAs or related issues and staff who advise those decision-makers.

Resource Managers – Local, state, or federal agency staff responsible for management and/or regulatory decisions about MPAs or issues directly affected by MPAs. Organizational affiliations of this archetype included the California Department of Fish and Game (CDFG), the National Marine Sanctuary Program, and sanitation districts. The primary focus of the MPA Monitoring Enterprise, and consequently of the planned IMS, is on MPAs implemented under the Marine Life Protection Act. People who involved in the management of other types of MPAs were included because of their potential interest in monitoring results for California's MPAs.

Policy Informers or Influencers – Policy analysts, as well as those who influence policy. People in this archetype do not directly make decisions about MPAs but are interested in achieving the overarching goals of the MLPA Initiative. Examples of organizational affiliations include the California State Coastal Conservancy, the MLPA Initiative, and the Ocean Protection Council.

Scientists – Scientists who are typically affiliated with academic or government institutions, but also some independent consulting scientists. Individuals classified as Scientists conduct monitoring and research related to MPAs (basic or applied). Many also have an educational role with undergraduate and/or graduate students. This archetype excludes scientists affiliated with stakeholder organizations; although these individuals may have similar information needs, they are likely to use this information in furthering their organization's mission relative to MPAs and are thus classified on the basis of their affiliated organizations. The scientist archetype included research and applied scientists, and individuals who manage scientists or scientific organizations. Represented disciplines included biological and physical oceanography, social science, mathematics, modeling, and economics.

Information Managers – Technologists or managers of technologists, commonly affiliated with government or academic institutions. People in this group provide tools, data, analyses or maps to support MPA or marine management and policy. This archetype excludes Information Managers associated with stakeholder groups for the same reason as scientists affiliated with stakeholder organizations were excluded from the scientist archetype; such individuals were classified on the basis of their affiliated organizations. Individuals included in this category included website developers and/or webmasters; GIS analysts; and data analysts.

Educators – People whose primary role is education of children and/or adults about MPAs or the marine environment. This archetype excludes college and university educators (most were classified as scientists), as well as educators associated with stakeholder groups. Individuals in this archetype included teachers (Grades K-12) and education coordinators in museums and aquariums. Examples of organizational affiliations of individuals within this archetype are COSEE (Centers for Ocean Sciences Education Excellence) and several aquariums in California.

The following archetypes were created to more finely differentiate various stakeholder groups to ensure each was sampled in the UNA.

Stakeholder: Commercial Fishing – People who fish commercially as well as those involved in seafood marketing, commercial fishermen's associations, species-specific commissions, and commercial organizations that support fishing (e.g., bait shops, boat charters). Internet survey respondents that selected commercial fishing as their primary role associated with MPAs were classified with this archetype unless they provided an alternative organizational affiliation.

Stakeholder: Recreational Consumptive – People involved in recreational fishing and shellfishing, as well as members of recreational fishing clubs and associations; diving clubs with a focus on spearfishing; and commercial businesses that support recreational fishing (including media, charter boats). For example, internet survey respondents that selected recreational fishing as their primary role associated with MPAs were classified with this archetype unless they provided an alternative organizational affiliation.

Stakeholder: Recreational Non-Consumptive - Recreational boaters (yachts to kayaks) and supporting businesses (marinas); divers and snorkelers (non-consumptive); and people involved in other marine and beach-related activities. Internet survey respondents that selected recreational activities as their primary role associated with MPAs (but not fishing-related) were classified with this archetype unless they provided an alternative organizational affiliation.

Stakeholder: Environmental Non-Governmental Organization (ENGO) –People who work for or support organizations that advocate for environmental protection or sustainability. This category includes all employees of ENGOs regardless of their individual role within the organization. There were a wide variety of organizational affiliations in this category.

Stakeholder: Local Coastal Managers – City, town, and port managers who are responsible for facilities that affect, or are directly affected by, MPAs, and who are concerned with information about MPAs that will help them to manage their facility, permit, or coastal resource. Organizational affiliations of this archetype included cities, counties, ports, and harbor districts. This archetype excludes facilities with a specific commercial or recreational purpose (e.g., supporting recreational activities) that are covered in the commercial or recreational stakeholder groups described above.

Stakeholder: Military – People associated with federal military facilities who have concerns about enforcement, security, and environmental impacts and regulations. Individuals in this archetype have attributes similar to those of Local Coastal Managers. Organizational affiliations of this archetype included coastal Navy and Air Force facilities.

Stakeholder: Tribal – All respondents associated with a tribal-related organization were classified as Tribal regardless of the individual's role within the organization or tribe.

The final archetypes described below were created after the internet survey was completed to enable classification of respondents that did not easily fit into one of the existing archetypes.

Students – Respondents who specifically identified themselves as students. All members of this archetype were university graduate students.

Citizens – Respondents who specifically identified themselves as citizens, without any other available organizational, recreational, or commercial association with MPAs.

Miscellaneous – The miscellaneous classification was applied to the remaining internet respondents that did not fit into other categories, and included consultants, artists and photographers.

1.3 Designing and Evaluating the Assessment

This section summarizes how the assessment was designed and implemented to ensure that the results could readily translate into the next stages of IMS development by the MPA Monitoring Enterprise. The team initially considered three survey approaches: one-on-one interviews (telephone and in-person); focus group interviews; and an internet-based survey. We chose to conduct an internet survey because it allowed us to obtain information from a broad survey population that would potentially match the wide range of potential users. Internet surveys are also cost-effective and allow collection of quantitative and readily processed user needs data. The other survey method that we chose was one-on-one interviews conducted via telephone. Telephone, as opposed to in-person, interviews allowed us to survey a relatively large number of people in a detailed manner. We used the telephone interviews to develop qualitative and more nuanced information than was possible through the online survey. Focus groups were not implemented primarily due to the high cost relative to the amount of information gathered.

Two types of telephone interviews occurred during the course of the project: an initial set of key informant interviews to elicit information that would assist in designing the assessment and a subsequent set of scripted interviews that focused on the same topics as the online survey. A *key informant* was defined as someone familiar with the MLPA Initiative process who would need or could benefit from information generated through monitoring of the MPAs. The group of individuals selected to serve as key informants also included people who were likely to have creative ideas or useful insights that would aid in accomplishing the Monitoring Enterprise's goals for information management, synthesis and visualization. The key informants represented several key categories of users, including ones that could speak to the information needs of the scientific, stakeholder, and policy-related communities. The interviews with key informants were conducted a) to ensure the contact list of potential user and user groups (archetypes) was comprehensive; and b) to seek input on survey topic areas that would be relevant to potential users of MPA information.

The list of survey topic areas developed for the User Needs Assessment was iteratively refined up to the time of actual survey implementation. Based on initial interviews with key informants, we identified seven survey topic areas that encompassed the various interests and experiences of the potential IMS users. These topics provided an outline for developing specific questions (Table 1-1) that were used in the internet survey (Appendix A) and for guiding the telephone interviews.

Table 1-1. User Needs Assessment Survey Topics

UNA SURVEY TOPIC	DESCRIPTION
Content	Preferences for types of monitoring information; user needs for other related information.
Synthesis	Level of detail of the information to be offered by the IMS; related to preferred presentation of the information.
Interactivity	User preferences for utilities, tools, and functions available for browsing, viewing, downloading, or otherwise interacting with the IMS.
Human-Computer Interface	The user experience and preferences, how they interact with the internet, ease of use of a site, including physical or cultural barriers.
Institutional	Organizational issues that will impact how the website will be run and maintained, including partnerships, institutional agreements for data sharing.
Audience	Attributes of the potential user group and how these attributes affect design issues.

Upon final completion of the internet surveys and scripted telephone interviews, the data were first analyzed separately for each survey method and then for each topic area (see Sections 3 and 4 of the report). The data from the two survey methods were then combined for two topic areas (synthesis and interactivity) to derive the overarching conclusions found in Section 5 of the report.

<u>Audience as System Driver</u> – The user-centered design of this analysis required explicitly defining the potential audience for the IMS (Section 5.1). We first analyzed the UNA data according to archetype. However, because archetype indicates an individual's affiliation rather than their likely preferences and uses of an IMS, we then developed a second, more independent, designation – the "User Persona" –to categorize people who participated in the UNA according to their information use attributes. Distinctions between "archetype" and "user persona" can be drawn from the definitions that follow:

> ***Archetype*** – A category of a potential IMS user based on associated role with, or interest in, marine protected areas. The archetype serves as a model that represents the interests of a group of like users, but does not infer any assumed interaction with, or preference for, MPA information. The purpose of the archetypes is to ensure comprehensive coverage of the diverse interests and perspectives of those interested or engaged in MLPA implementation.
>
> ***User Persona*** – Classification of a potential IMS user according to their information use habits and preferences. This classification was derived from the surveys and interviews and is independent of the individual's role with, or interest in, marine protected areas. User personas (see p. 63) are idealized categories of users that help the IMS design match the end-users' actual needs and preferences. The user persona construct creates empathy and facilitates discussion. The purpose of the user personas is to identify and characterize major patterns in what IMS users want from, and how they wish to interact with, the IMS in order to inform IMS design.

User Need Themes – Several themes arose consistently throughout the UNA that will be indirectly relevant to decisions about the IMS design. These themes often reflect the users' varying philosophies or perceptions and are summarized in Section 5.2 as potential guiding principles for the next stages of the IMS development.

Implications of the UNA Results for the IMS Technical Framework – To ensure that the UNA data could inform technical issues that will arise in the subsequent IMS design and development process, we developed an IMS Framework that links the survey topics (Table 1-1) to elements of the IMS design (Table 1-2). Section 5.3 of the report synthesizes the results of each survey topic in terms of its implications for each element of the IMS framework shown in Table 1-2. Section 5.3 also includes a general assessment of the relative effort and costs of various implementation options.

Table 1-2. IMS Framework Element Description and Examples

Element	Description	Example Issues
Architecture[1]	Model of the way the IMS is set up; where the database, interface, and applications are stored, and communication pathways between these elements.	File server vs. client server Centralized vs. distributed and distribution components (residence of database/apps) Role of the Monitoring Enterprise (interface, database, and/or applications) Proprietary vs. open source
Components	Software; hardware; networking.	Web server and middleware options External data products (e.g., CDs)
Modules	System functionality grouping; applications; modular design issues.	Modular phasing Use of commercial applications, partnering Program code and browser standards
Interface	Website design and logic, graphical user interface.	User tiers (levels of entry) Language and/or disability compliance
Database	Organization, standards, and protocols that are used to store and serve data.	Relational vs. GIS/Geodatabase Data and metadata standards and codes Templates Archiving and long-term maintenance

[1]Note that the system will be internet-based.

2. Data Collection and Analysis Methods

2.1 Survey Population

The group targeted for participation in the UNA survey would, under ideal circumstances, include proportionate representation of the various actual users of MPA monitoring information. However, because this user group does not yet exist, Exa worked with the MPA Monitoring Enterprise to develop a contact list that included all of the archetypes of interest and geographic representation from all MLPA regions in California. The final contact list of potential users who were invited to participate in the UNA included 466 people distributed across archetypes as shown in Table 2-1. Most (412) were invited to participate in the online survey, but a subset (49 people or 10.5% of the contact list) was reserved for participation in the telephone survey (Section 2.3). Participants in both the internet and telephone surveys were told their responses would remain anonymous in the analysis and reporting of results.

Table 2-1. Distribution of Archetypes in the Final Contact List

Archetype	Count	Percent
Decision-Makers	10	2.1%
Resource Managers	37	7.9%
Policy Informers or Influencers	24	5.2%
Scientists	151	32.4%
Information Managers	9	1.9%
Educators	28	6.0%
Stakeholders		
Commercial Fishing	48	10.3%
Recreational Consumptive	39	8.4%
Recreational Non-Consumptive	14	3.0%
Environmental Non-Governmental Organization	64	13.7%
Local Coastal Manager	15	3.2%
Military	3	0.6%
Tribal	6	1.3%
Total Stakeholders	189	40.6%
Miscellaneous	18	3.9%
TOTAL	**466**	**100.0%**

An open invitation was extended to members of the MLPA Initiative listserv subscriber list (2,376 people) to participate in the online survey. This open invitation subsequently “went viral” as it was passed on to other potential participants through emails and website postings. The survey design allowed us to separately analyze responses of participants who responded to this “open survey” from those who were on the contact list. Contact list participants were asked to verify their institutional affiliation, title, and contact data during the survey.

Invitees from the contact list were given two follow-up reminder emails to take the survey; individuals on the MLPA Initiative subscriber list were given one follow-up reminder email. Both the invitee and the open surveys were closed after five weeks.

2.2 Internet Survey Methods

We selected Survey Gizmo (www.surveygizmo.com) as the survey platform based on such features as its robust software, spam control, contact list management feature, data management options, and reporting options.

The diversity of the survey population made designing the survey challenging. Developing a different survey for each archetype was impractical. Moreover, it was important to avoid assumptions about user expertise based exclusively on archetype. Therefore, we selected a nested approach, in which answers to certain questions led to more detailed follow-up questions depending on the self-reported experience or interest of the respondent (see full survey instrument in Appendix A). For example, if the user specified that he or she was a data producer as well as a consumer, he or she was prompted to answer specific questions about what kind of data he or she anticipated contributing to the system.

The first question in each set of nested questions is referred to as a 'gateway' question in this report. Because each gateway question narrowed the pool of respondents that was shown subsequent nested questions, the term "eligible" respondents describes the response rate of that subset of respondents. The following hypothetical example demonstrates how the nested approach worked, and the use of the term eligible respondents:

Illustrative Example:

100 people took the survey
80 people selected a particular option in a gateway question and were shown a follow-up question
50 people answered the follow-up question
10 answered "Essential" to that question

Response rates were calculated as follows:
Response rate to the gateway question = 80/100 total respondents, or 80%
Response rate to the follow-up question = 50/80 eligible respondents, or 62.5%
Percent of eligible respondents that answered "Essential" = 10/50, or 20% of the eligible respondents that answered the follow-up question

The raw internet survey data were processed and cleaned prior to data analysis. These steps included:

- Remove duplicate (redundant) surveys;
- Merge surveys completed by the same individual;
- Remove surveys with no name and no demographic information (other than email);
- Remove surveys with no response (someone clicked on the link and then closed the survey).

The data were then downloaded into an Access database and responses assigned standardized codes for each survey topic and selected rating or choice. Each respondent was identified with a unique numeric ResponseID and then assigned an archetype (Section 1.2).

2.3 Telephone Survey Methods

The telephone interviews were semi-structured. Survey topic areas (Table 1-1) provided a template for each interview, with multiple questions for each topic. A script of related questions guided the conversations, but not all questions were asked of all interviewees. Instead, the goal of the interviews was to allow each interviewee to expand on his or her particular areas of interest, concern, and/or expertise. Many of the interviews departed from the standard list of questions. In total, 35 telephone interviews were conducted.

Three interviewers conducted the interviews. The common interview format and extensive discussions among interviewees enhanced consistency among the interviews. The anticipated duration of each telephone interview was up to one hour. In practice, some were as short as 15 minutes to a half hour because of some interviewees' limited availability. Notes taken during the interview captured each conversation's content; then each interviewer summarized the key points of the interview with an interpreted synthesis.

Completed telephone interview transcripts were reviewed to identify recurring themes. Some, but not all, of these themes came directly from the nine interview topics. Snippets of text were copied from the raw interview transcripts into spreadsheet worksheets for each of the themes. This provided a systematic method for evaluating how many times a particular theme or category within a theme was mentioned and whether the interviewee provided a positive or negative comment (e.g., that a particular function or content either was or was not necessary or desirable). Section 4 provides an overall analysis of the telephone interview results. Section 5.2 synthesizes what interviewees identified as the three top goals, needs, or issues for the IMS; this provides their perspectives on the essential functions of the system, *versus* those that would be nice to have.

3. Internet Survey Results

The survey was opened on April 23, 2009 and closed on June 1, 2009. Results of the internet survey presented below are organized by survey topic (Table 1-1). Appendix A provides a copy of the full internet survey. Within the text that follows, italicized quotes identify direct quotes from open comments portions of the survey. At the end of each survey topic, a summary of findings is presented, along with preliminary comments on the implications of those findings for each element of the IMS framework (Table 1-2). A summary of the responses to the final two open comment questions (list three "out-of-the-box" ideas and three top crucial aspects of the IMS) are provided in Section 5.2. Section 5.3 synthesizes implications of the online and telephone survey results for the IMS framework and makes related recommendations for the IMS design.

3.1 Response Rate

Of the 417 people on the contact list who were invited to take the internet survey, a total of 128 (30.7%) responded. An additional 391 people responded via the open survey. Table 3-1 compares the archetype distribution from the contact list to the actual number of respondents from both the invitee and open surveys. Stakeholders and scientists made up a large proportion of the contacts list (40.6% and 32.4%, respectively). These two groups also comprised a large proportion of the survey respondents (stakeholders 59.3%, scientists 15.4%). The relatively high response rate by stakeholders was due to the high number of responses by recreational stakeholders (156 consumptive and 59 non-consumptive).

Table 3-1. Distribution of Archetypes in the Contact List and Actual Respondents

Archetype	Contact List[1]		Internet Survey Respondents			
	Count	Percent	Contact List	Open Survey	Total Count	Total Percent
Decision Makers	8	2.1%	3	3	6	1.2%
Resource Managers	30	7.9%	10	8	18	3.5%
Policy Influencer/Informers	12	5.2%	5	7	12	2.3%
Scientists	140	32.4%	44	36	80	15.4%
Information Managers	4	1.9%	3	13	16	3.1%
Educators	28	6.0%	5	16	21	4.0%
Stakeholders						
Commercial Fishing	46	10.3%	13	17	30	5.8%
Recreational Consumptive	39	8.4%	15	141	156	30.1%
Recreational Non-Consumptive	14	3.0%	2	57	59	11.4%
Environmental Non-Governmental Organization	59	13.7%	19	27	46	8.9%
Local Coastal Manager	15	3.2%	6	5	11	2.1%
Military	3	0.6%	1	2	3	0.6%
Tribal	5	1.3%	1	2	3	0.6%
Total Stakeholders	181	43.4%	57	251	308	59.3%
Citizens				42	42	8.1%
Students				6	6	1.2%
Miscellaneous	14	3.4%	1	9	10	1.9%
TOTAL	**417**		**128**	**391**	**519**	

[1]Includes people invited to take the internet survey unless interviewed by telephone.

We considered the 30.7% response rate by invitees on the contact list to be good, particularly given the time required to take the internet survey (the median survey time was 30 minutes). At least one invitee from each archetype responded, and the response rate of survey invitees per archetype ranged from 28-42%, except for recreational non-consumptive stakeholders (14.3%), educators (17.9%), and tribal stakeholders (20%). The response rate of information managers, 75% was the highest among the various archetypes. The response to the open survey helped to fill some gaps in the distribution of respondents from the contact list, especially recreational stakeholders and educators.

For each topic discussed below, we include a response rate that provides the number of people who selected at least one option in each question. For multi-option questions, the variation in response rate among options was typically $\leq$5%, and never exceeded 10%, suggesting that if someone chose to answer a question about one option, they provided an opinion about every option.

3.2 Attributes of the Respondents

The initial "About You" sections of the survey provided information on the respondent attributes (Appendix A, Sections 4-5). The data assisted in classifying respondents by archetype (Section 2.1) and provided information useful to characterize the respondent population. Almost all of the respondents (494 out of 519, or 95%) answered at least one of the questions in this section.

Respondent Affiliations with MPAs

The first question in the "About You" section was intended to help distinguish users whose professions were related to management, monitoring, or policy of MPAs from those who had a commercial or recreational interest in MPAs. Of the 519 people who took the survey, 221 people, or 42.6%, indicated that at least part of their job was related to either "establishment, management, or monitoring of MPAs" or "marine science, education, marine resource use, policy or management" (Figure 3-1). Of those, one third associated themselves with education or research, and almost as many with government or public service (Figure 3-1). A smaller number were self-described as working in a non-profit or non-governmental organization (20.8%), marine industry (10.4%), or "other" (5.9%).

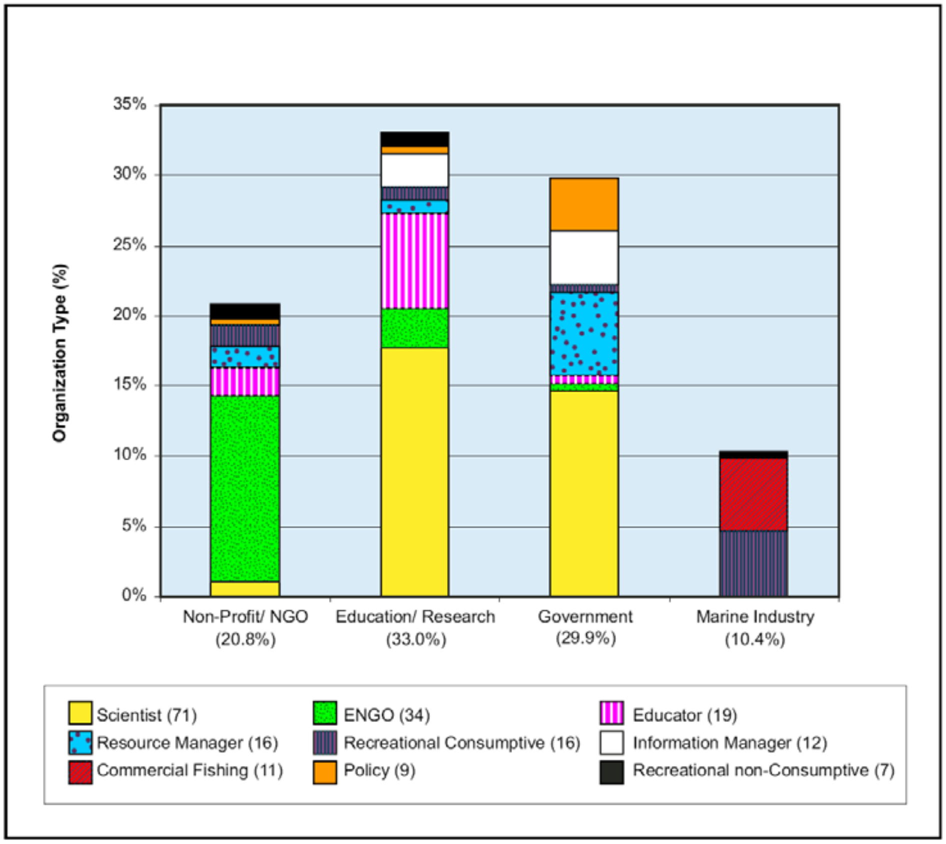


Figure 3-1. Distribution of organization types for users self-described as having a direct or indirect professional association with MPAs (221/519 or 42.6% of respondents). The archetype distribution is shown relative to the total percent within each category; numbers in parentheses represent the number of individual respondents per archetype.

The distribution of archetypes within each affiliation category was generally consistent with expectations based on the organization type. For example, environmental non-governmental organization stakeholders (ENGOs) dominated the non-profit/NGO category, and more than half of those who described their organization as education- or research-related were scientists. The government or public service affiliation was similarly dominated by scientists, and also included resource managers, information managers, and policy-related individuals. The marine industry category was dominated by the commercial fishing and recreational consumptive stakeholders.

Follow-up questions about direct or indirect professional associations with MPAs elicited further detail on the organizational affiliation of each respondent as shown in Table 3-2 (see Appendix A for questions). Of the respondents in the education research field, more than two thirds reported they were active at the college level (Table 3-2). Those in government were largely split between state and federal agencies, with fewer responses coming from those in local government. Respondents affiliated with non-profit and non-governmental organizations (NGOs) were mostly from the environmental or conservation community, although a few were affiliated with various other NGO types.

Table 3-2. Organizational Affiliations of Respondents with Professional Association with MPAs

Organization Category	Discipline or Area	Number	Percent
Education/Research	College	49	67.1%
	K-12	6	8.2%
	Non-Profit Research Institution	9	12.3%
	Public Outreach	5	6.8%
	Other (informal, all)	4	5.5%
Government/Public	Federal	26	39.4%
	State	28	42.4%
	County, Regional, City/Town	9	13.6%
	Tribal	1	1.5%
	Other (district)	2	3.0%
Charity/Non-Profit/ Non-Governmental Organization	Environment/Conservation	30	65.2%
	Marine education	4	8.7%
	Other[1]	12	26.1%
Marine Industry	Commercial Fishing	12	52.2%
	Recreational Fishing	3	13.0%
	Other[2]	8	34.8%
Other		13	
Total		**221**	**42.6%[3]**

[1]Marine management policy, Marine or coastal tourism, Marine trade association, Philanthropy, Recreational fishing, Research, Scuba diving, Snorkeling or free-diving, Sustainable development
[2]Aquaculture, Fishing/Boating Supplies/Services, Marina, Port/Commercial shipping, Scuba/Snorkeling, Watersports
[3]Total of 221 out of 519 respondents

More than half of the survey respondents (282 of 519) indicated they had an interest in marine recreational activities (Table 3-3). This information enriched our understanding of the respondent population. For example, although only about 10% of the survey respondents identified a professional affiliation with marine industry (commercial or recreational fishing) (Figure 3-1), this should be considered in light of the number of respondents who indicated that they conduct or participate in commercial (10%) or recreational (36%) fishing activities (Table 3-3). In the absence of other kinds of identifying information, this kind of information was used to help assign archetypes.

Overall, the respondent population had good representation both of people who have a professional association related to MPA management/monitoring (42.6%) as well as of people who have a commercial or recreational interest in MPAs (54.3%). Many respondents had multiple, and sometimes overlapping, roles that may affect their interest in MPA monitoring data. Based on the above analysis, the survey appears to have reached a wide variety of interests and affiliations, although not with equal representation. Possible gaps in the survey population include people from local government, from education/research institutions outside of the college environment, and from non-environmental NGOs.

Table 3-3. Commercial and Recreational Interests Indicated by Respondents

Commercial or Recreational Activity	Number[1]
Fishing – recreational, boat-based	189
Beachgoing	170
Conservation	166
Diving, SCUBA, snorkeling, or freediving – non-consumptive	153
Wildlife viewing	151
Fishing – recreational, shore-based	136
Diving, snorkeling, or freediving – consumptive (e.g., spear fishing)	135
Watersports (swimming, surfing, windsurfing, kitesurfing)	125
Ocean kayaking or canoeing	118
Tide pooling	115
Power boating	93
Fishing – commercial	53
Sailing	50
Personal Water Craft (PWC)	24

[1]Total number of people selecting at least one activity was 282/519 (54.3%)

Regional Interests

The next series of "About You" questions queried the respondents as to their interests in various regions of California (based on MLPA regions), and spatial scales of information (MPA specific, region, statewide) (Appendix A, see question 5). These questions also provided a way to explore potential geographical biases in the survey data.

The MLPA is being implemented in California sequentially, region-by-region. Thus potential users from different parts of the state may have experience with different stages of the MLPA implementation process. For example, at the time of the survey, MPA planning was just starting in the North Coast Region, whereas state MPA designation was completed in 2007 in the Central Coast Region. Regional differences also exist in which issues are likely to be of greatest interest to potential IMS users and therefore might influence their responses.

First, respondents were asked what region(s) interested them: the Central Coast, North Central Coast, North Coast, San Francisco Bay, South Coast, or All of California. All respondents were shown this question, and 496 (95.6%) selected at least one region of interest. More than half of the respondents (52.2%) indicated that they were interested in the entire state of California. The South Coast was selected by 42.9% of respondents, the Central Coast by 23.2%, the North Central Coast by 22.2%, the North Coast by 19.8%, and San Francisco Bay by 12.7%. This regional pattern is consistent with the physical locations of survey respondents, as indicated by information they provided on city and postal code and the IP address collected by the survey software (see Appendix B for map of locations for survey respondents). For example, more than half of survey respondents were located in southern California, and the South Coast region was the most frequently selected region of interest. The relatively large population and ongoing MLPA planning process of southern California at the time of the survey may also have contributed to over-representation of that part of the state in the survey population and results. This potential bias should be taken into account when weighing options in the IMS design phases; for example, there may be issues of particular interest to the South Coast region, such as water quality, that may be over-represented in the survey results.

The next question about geographical scale asked respondents to indicate whether they were interested in accessing information about one or more specific MPA(s), the MPAs within a study region, and/or the full statewide MPA network. Most of 519 respondents (93.6%) selected at least one of these options (Figure 3-2). More than half (63.2%) were interested in one or more individual MPAs. The comments offered by 39 respondents, however, showed that many of these individuals were actually interested in the MPAs found in a specific area or region (e.g., *"South Coast Area, specifically Oceanside to Mex. Border*," and "*Sonoma and Mendocino coasts, as well as Monterey/Carmel area.*") rather than individual MPAs (Table 3-4). Approximately 20% indicated interest in accessing information from the statewide MPA network, in contrast to the more than half that indicated an interest in all of the MLPA regions throughout the state of California (above).

When the results were broken out by archetype, most archetypes tracked the pattern of the combined respondents (Figure 3-2). One exception was resource managers, who selected all three options with relatively equal frequency, perhaps reflecting their responsibilities at all three spatial scales. No information managers selected the statewide option, for no obvious reason.

Based on these results, it appears that most respondents are interested in accessing information through a local portal, even if, as their comments suggest, their perspective ranges over a larger area. This finding is consistent with the comments (discussed below) indicating that many users would like the ability to click on a map and 'drill down' into more detailed information. The implications of this multi-scale capability of the IMS are discussed further in the Interactivity section of the report (see p. 33).

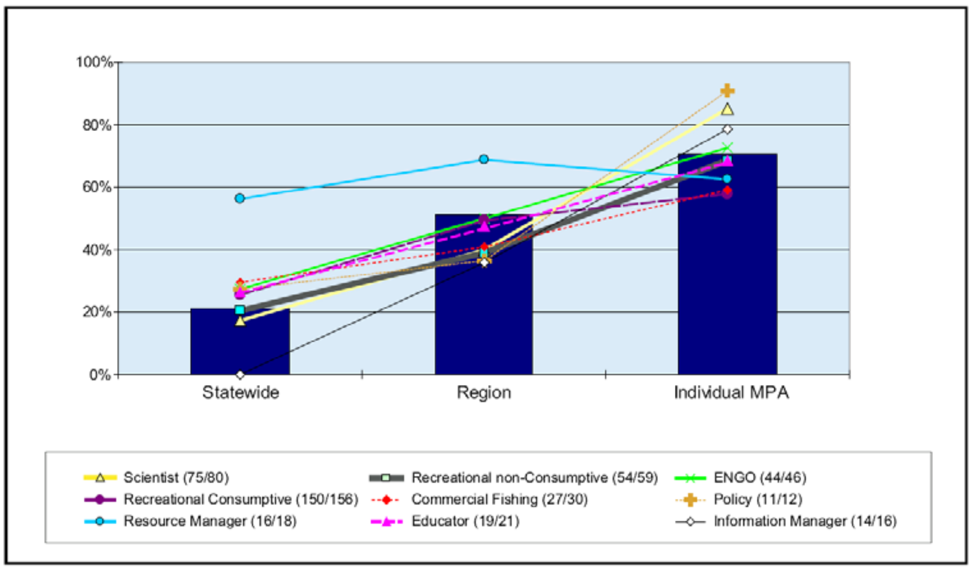


Figure 3-2. Percent of respondents who indicated they were interested in accessing MPA monitoring information at three different spatial scales. Columns show the percent of respondents who selected that option averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey for each archetype. The number of respondents selecting at least one option in this category was 486/519 (93.6%).

Table 3-4. Specific Marine Protected Areas Mentioned by Respondents

Marine Protected Area (Current or Proposed)	Number of Mentions[1]
Asilomar	1
Bodega/Bodega Head	2
Bolsa Chica	2
Carmel Bay/Pinnacles	3
Catalina	4
Channel Island	6
Duxbury	1
Ed Rickets	2
Elkhorn Slough	1
Farallons	3
Goleta Slough	1
Humboldt Bay	1
James V. Fitzgerald	3
La Jolla	3
Laguna Beach	5
Lovers Point	1
Malibu	1
Mia J. Tegner	2
Montare	1
Monterey Bay	2
Moro Cojo Slough	1
Morro Bay	2
Newport Beach	3
Pacific Grove	2
Palos Verdes	6
Point Loma	2
Point Reyes	1
Russian River	1
San Clemente Island	1
San Diego Area[2]	1
Saunders Reef	4
Stewarts Point	3
Vandenberg	1

[1]A total of 39 individuals cited at least one specific MPA

[2]San Diego River, Mission Bay, Famosa Slough, San Diego Bay, South San Diego Bay, Tijuana Estuary

Software Use

A final aspect of the survey population that was assessed in the "About You" section was the users' levels of experience with the various software systems listed in Table 3-5. Previous software experience is important because it could influence how users access and use the IMS. Most of the respondents (96.1%) selected at least one option, and the three applications most frequently used by the survey population were email, web-browsing, and word processing (Table 3-5). Standard data processing tools (database, spreadsheet, graphics) as well as image processing were the next most frequently used applications, while specialized data analysis tools were used least frequently by the survey population.

Table 3-5. Use of Types of Software Applications

Application	Never	Rarely	Sometimes	Frequently	Total
Email	0.2%	0.0%	2.2%	97.6%	454
Web-browser	1.1%	0.5%	5.9%	92.6%	443
Word processing	1.1%	4.1%	13.7%	81.1%	438
Spreadsheet	5.6%	12.8%	23.5%	58.1%	430
Database	8.6%	23.0%	32.3%	36.1%	421
Graphics (illustration, production)	9.9%	24.3%	32.0%	33.9%	416
Image processing	11.2%	26.0%	31.3%	31.5%	419
GIS	33.5%	29.6%	23.4%	13.5%	406
Statistical Software	38.6%	32.2%	17.9%	11.3%	407
Modeling tools	45.8%	27.1%	17.9%	9.2%	402
Programming Languages	61.2%	20.8%	9.0%	9.0%	399

The total number of eligible respondents who selected at least one option was 499 of 519 (96.1%)

The respondents were also asked if there were any other types of software that they considered to be important. Eleven respondents specified social networking, video or music editing software, or web-based collaborative software (e.g., Sharepoint; "*Collaborative Adaptive Management Tools*"), and three respondents specified navigation or charting software. Three responded directly to the GIS option by mentioning their preference for Google Earth or "*non-GIS mapping tools*." One person suggested "*support for both Windows and Mac operating systems*" and one requested "*Mobile Phone Based Browsers*."

The results showed the survey respondents have widely varying types of experience with software. Not unexpectedly, most have experience with more common software tools (email, web browsing), and fewer have experience with high-end software such as modeling and programming. This widely variable experience occurred within each individual archetype as well. For example, of the 39 respondents that that selected "Frequently" for programming, scientists and recreational non-consumptive respondents each made up 23.1% of this group, followed by recreational consumptive stakeholders (20.5%) and then information mangers (15.4%). The implications of these results are that the potential IMS users will vary widely in their levels and types of experience with different software tools and that such variation occurs across all archetype groups.

3.3 Information Content

The Information Content section of the survey ("MPA Monitoring Information" – Sections 7-9 of Appendix A) examined the users' interest in having various categories of information available from the IMS that relate to whether the MPAs have met the goals and objectives of the MLPA. Although the focus and content of monitoring data will be formulated through a separate monitoring planning process, the results of this survey are potentially useful in prioritizing what monitoring information will be made available through the IMS. Different types of data require different methods of database management, and potentially will influence data format and software decisions when designing the IMS.

MPA Monitoring Information

Most respondents (90.5%) answered at least one of the MPA monitoring information content questions. In evaluating the relative importance of information on various ecosystem types, kelp beds were most frequently identified as essential (70.6% of respondents who answered the question, or 332) and deep soft bottom the least frequently (32.1%) (Figure 3-3). Each ecosystem type was rated essential by at least a third of the online survey respondents (Figure 3-3). The not needed category was <6.4% for all categories.

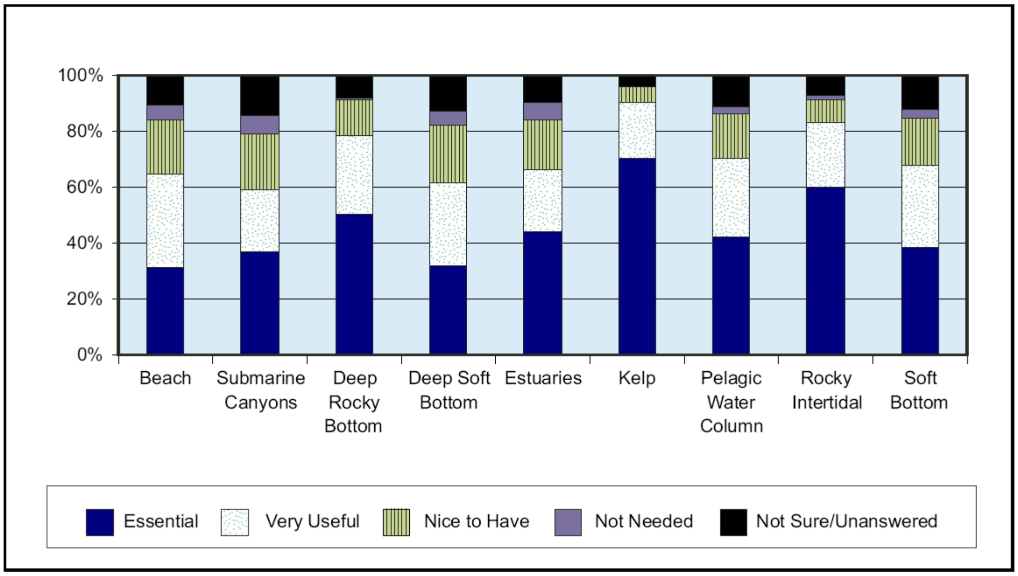


Figure 3-3. Distribution of all responses that indicated interest in monitoring information related to specific types of ecosystems. The number of respondents selecting at least one option in this category was 470/519 (90.5%).

Several interesting patterns emerged when comparisons were made among archetypes (Figure 3-4). Most importantly, the archetypes generally tracked one another in the frequency with which they identified each ecosystem type as essential, with educators being the most common outlier. Stakeholders associated with ENGOs rated three ecosystem types as essential more frequently than did the other archetypes: sandy beaches, estuaries, and rocky intertidal. Scientists, in contrast, most frequently rated kelp beds as essential, followed by deep rocky bottom and rocky intertidal – a pattern that is consistent with the emphasis of existing scientific monitoring activities. In the open comments for this category, several respondents specifically cited sea grass (eel or surf; 5 people), water quality, and/or fresh water/sea water interaction (7 people), and two mentioned artificial reefs.

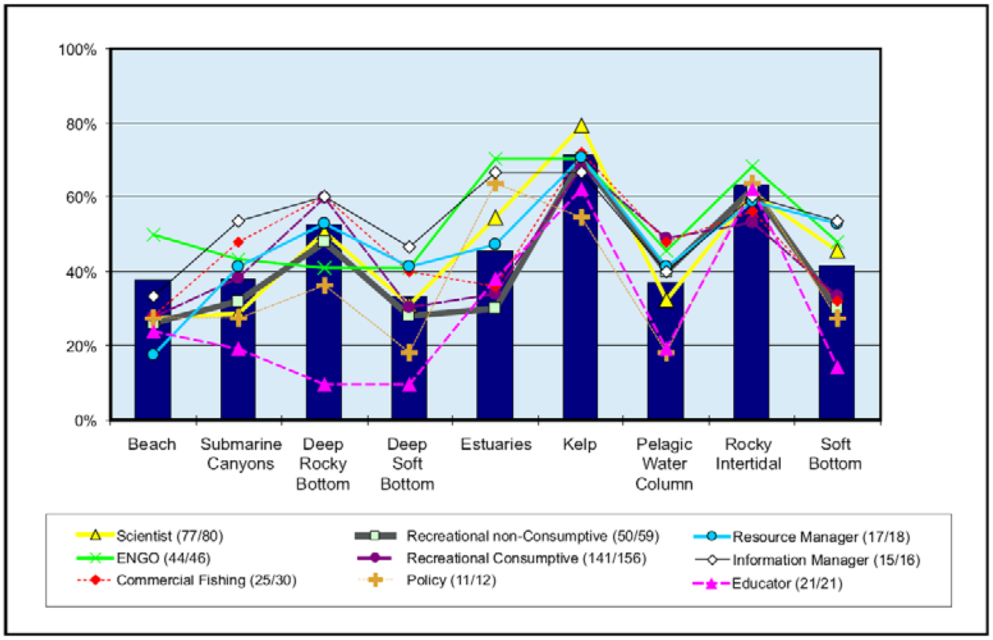


Figure 3-4. Percent of respondents who indicated each ecosystem type as essential content of the IMS. Columns indicate the percent of respondents that answered essential averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey for each archetype. Figure is based on response of 383 people who selected essential for at least one ecosystem as shown in Figure 3-3.

Most respondents (90%) selected at least one option in response to the question asking their relative level of interest in information about different biological taxa – birds, fish, etc. (Figure 3-5). Fish were most frequently rated as essential by respondents (76.7%), and no respondent indicated that information about fish was not needed (Figure 3-5). More than half of the respondents also thought that information on invertebrates and plants/algae was essential. Four comments specifically highlighted abalone as the invertebrate of interest.

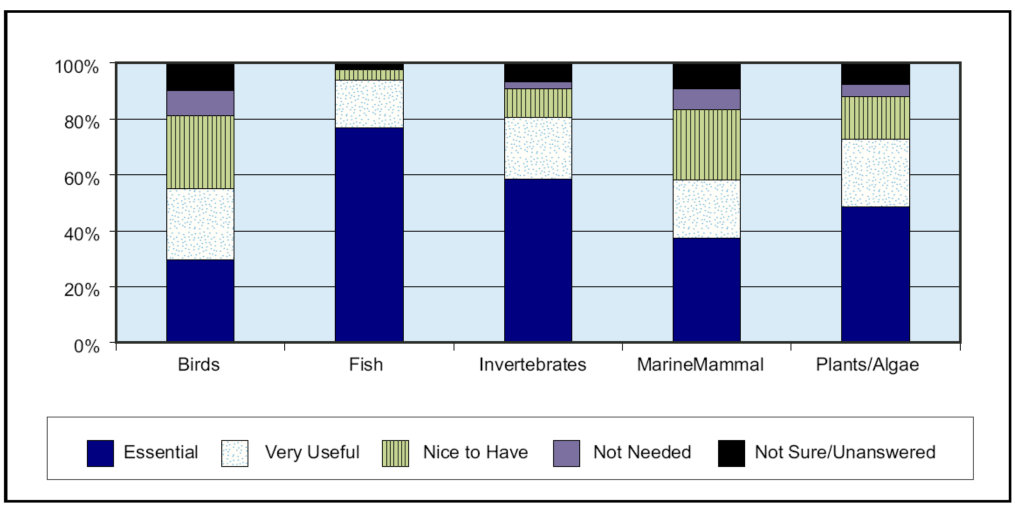


Figure 3-5. Distribution of responses that indicated interest in taxon-related monitoring information. The number of respondents selecting at least one option in this category was 467/519 (90%).

Although birds were least frequently identified as essential (25% of respondents), more than half identified information on birds as either essential or very useful. Overall, results of the taxa question indicated that potential IMS users were interested in all of the taxa categories identified in the survey (Figure 3-5).

Most archetypes showed the same relative levels of interest in the various taxa (Figure 3-6). The major outlier was the more than 50% of ENGO-related stakeholders (29 of 44 ENGO respondents) who identified monitoring information on birds as essential (Figure 3-6).

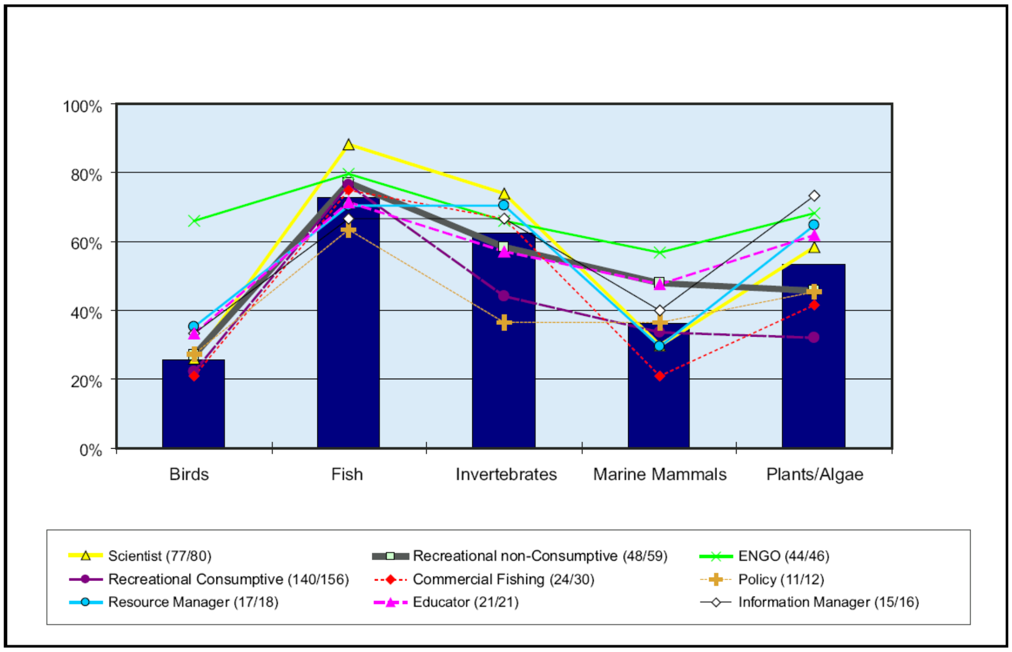


Figure 3-6. Percent of respondents who indicated specific taxa as essential content of the IMS. Columns indicate the percent of respondents who answered essential averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey per archetype. Figure is based on response of 385 people who selected essential for at least one taxon as shown in Figure 3-5.

Respondents were also asked to assign relative importance to various kinds of information on human uses. Most respondents (90%) selected at least one human use; recreational and commercial fishing were rated essential by the highest number of respondents (Figure 3-7). Most of the categories of human uses were rated as essential or very useful by more than half of respondents. The exception was Marine Transportation, which was rated nice to have, not needed, or not sure/unanswered by more than half of respondents.

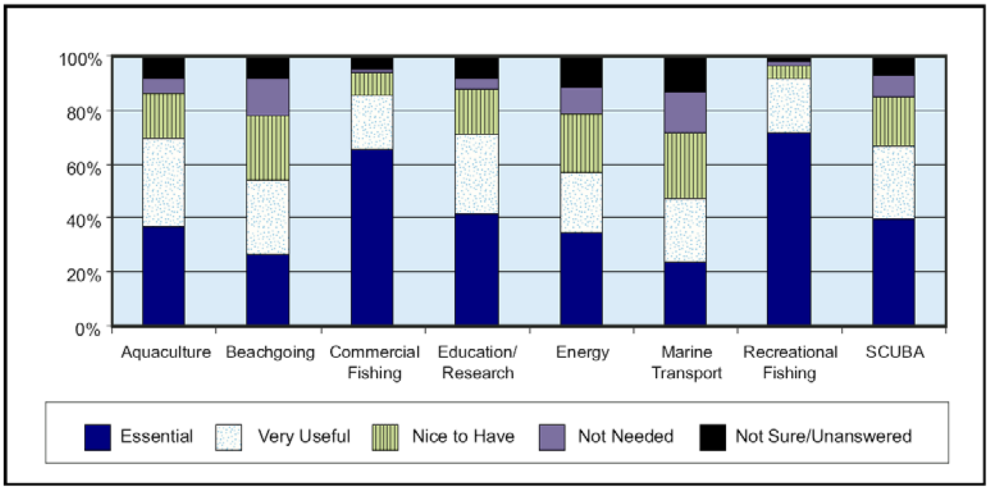


Figure 3-7. Distribution of responses that indicated some level of interest in human-use information. The number of respondents selecting at least one option in this category was 467/519 (90%).

Comparison across archetypes showed their interests were consistent with expectations based on stakeholder concerns (Figure 3-8). For example, 80% of commercial fishing stakeholders indicated information on commercial fishing use of MPAs was essential, while 81.7% of recreational consumptive stakeholders indicated information on recreational fishing use was essential. Equally interesting, respondents also desired information about uses beyond those suggested by their respective archetypes, although typically to a lesser extent.

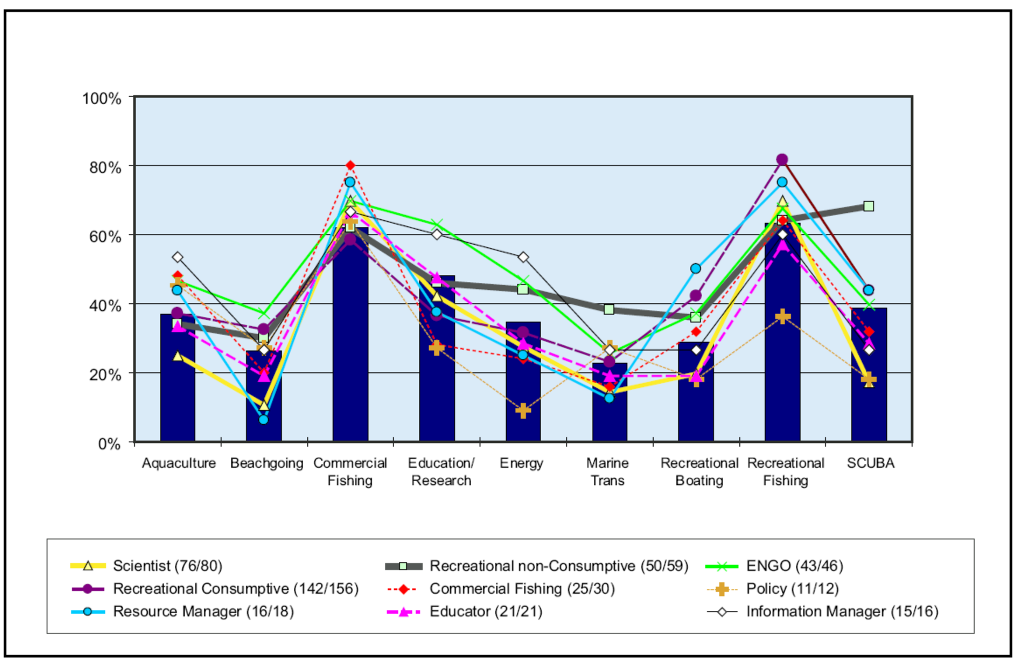


Figure 3-8. Percent of respondents who indicated specific human uses are essential content of the IMS. Columns indicate the percent of respondents that answered essential averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey for each archetype. Figure is based on response of 397 people who selected essential for at least one human use as shown in Figure 3-7.

From the perspective of the IMS development, the actual content of the monitoring data will have a primary impact on database issues, that is, what data are stored, how the data are categorized, organized, and documented, and other issues. The findings on users' level of interest in different types of information content provide helpful insights for the IMS designer in terms of user perceptions of what information content is essential and/or not needed. Regardless of what monitoring information is ultimately collected and served by the IMS, the information will need to be provided with associated explanatory information about why certain monitoring data were collected or excluded.

Other MPA-Related Information

In addition to questions about access to various types of monitoring data, the survey asked respondents to rate the need for the IMS to provide access to other types of data related to California's MPA network or to MPAs more generally (Section 8 of the survey; Appendix A).

In considering the various categories of MPA descriptive information shown in Figure 3-9 (boundaries, location, etc.), more than two thirds of respondents rated each category as essential, and none indicated that these data were not needed (Figure 3-9). The category most frequently rated as essential was allowed uses of the MPAs (79.9%). Although other sources currently manage and serve this information, these results have implications relevant to database issues and also to institutional considerations related to the IMS development (architecture; Table 1-2).

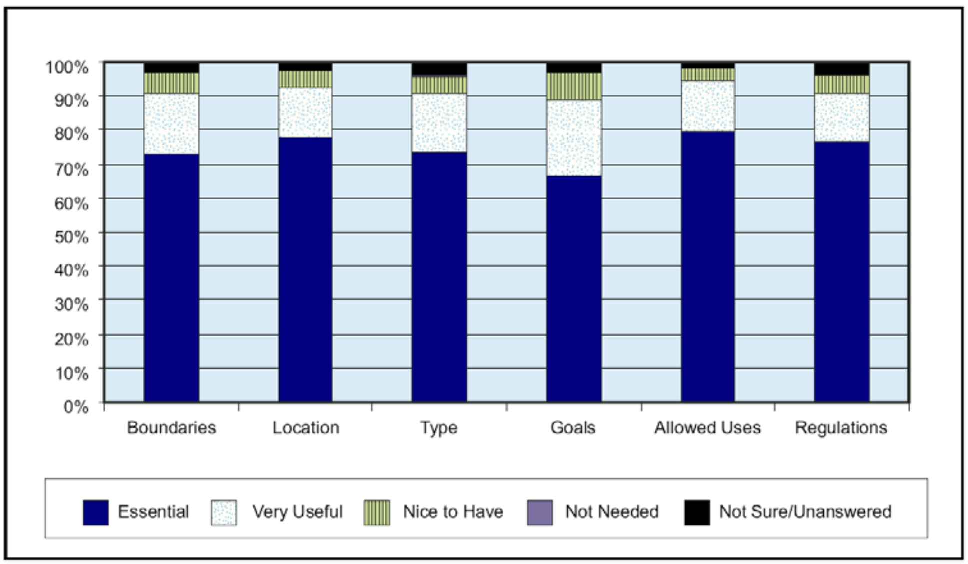


Figure 3-9. Distribution of all responses that rated other MPA descriptive information. The number of respondents selecting at least one option in this category was 463/519 (89.2%).

The pattern shown in Figure 3-9 was consistent for most archetypes. However, fewer resource managers rated the MPA-related information as essential, with frequencies ranging from 52.9% (boundaries) to 64.7% (regulations; 17 of 18 resource managers answered this question). The archetypes that most frequently rated this type of MPA information as essential were decision makers, local coastal managers, and military stakeholders. All five of the decision makers who answered this question indicated that MPA boundary, location, type, and use were essential, while four thought goals and regulations were essential. All of the eight local coastal managers who answered this question rated location and regulations as essential, while the two military stakeholders who responded rated all five categories as essential.

When asked to rate the various categories of MPA and monitoring related publications or educational materials shown in Table 3-6, far fewer respondents rated these information materials as essential content for the IMS to host (less than 20% for most categories). The exception was "California-based monitoring objectives and protocols" which was considered essential by more than half of the respondents who answered this question. This pattern held for all archetypes except for resource managers and information managers, of whom only 37.5% and 33.3 %, respectively, rated "California-based monitoring objectives and protocols" as essential.

Table 3-6. Response Summary for Monitoring Publication and Educational Resource Needs

Category	Essential	Very Useful	Nice to Have	Not Needed	Not Sure
Publications About Monitoring					
California-based monitoring objectives and protocols	266 (58.3%)	121 (26.5%)	52 (11.4%)	8 (1.8%)	3 (0.7%)
Monitoring protocols from other states or countries	53 (11.6%)	132 (28.9%)	182 (39.9%)	65 (14.3%)	7 (1.5%)
General aquatic and marine monitoring concepts and theories	86 (18.9%)	148 (32.5%)	160 (35.1%)	42 (9.2%)	4 (0.9%)
Documents about MPAs in other states or countries	59 (12.9%)	134 (29.4%)	174 (38.2%)	66 (14.5%)	5 (1.1%)
Educational Resources					
MPA science bibliography for students (K-12)	81 (18.1%)	111 (24.8%)	139 (31.1%)	97 (21.7%)	13 (2.9%)
MPA laws/policies for students (K-12)	76 (17.0%)	109 (24.4%)	133 (29.8%)	105 (23.5%)	14 (3.1%)
Instructional DVDs	75 (16.8%)	119 (26.6%)	140 (31.3%)	90 (20.1%)	14 (3.1%)
Posters/brochures	73 (16.3%)	131 (29.3%)	129 (28.9%)	88 (19.7%)	12 (2.7%)
Teacher guides	84 (18.8%)	119 (26.6%)	124 (27.7%)	95 (21.3%)	14 (3.1%)

Response rate for Publications about Monitoring was 87.9% (456 out of 519 selected at least one option); response rate for Educational Resources was 86.1% (447 selected at least one option).

Respondents rated most other categories of publications and educational resources shown in Table 3-6 more frequently as not needed than as essential. "General aquatic and marine monitoring concepts and theories" is the only category that reversed this pattern. Overall, except for California-based monitoring objectives and protocols, respondents selected the categories of very useful or nice to have most often for all categories of publications and educational resources.

A few patterns emerged when the data on individual archetypes were considered. Certain archetypes assigned an essential rating to information about "general aquatic and marine monitoring concepts and theories" more frequently than others. All were stakeholders, including recreational non-consumptive (33%); commercial fishing (26.1%); local coastal managers (25%); and recreational consumptive (21.2%). Scientists consistently rated educational resources of least importance in comparison to the other archetypes; only 5.3% rated DVDs and 13.3% rated teacher guides as essential. Educators, in contrast, most frequently rated educational resources as essential, ranging from 28.6% for posters/brochures to 47.6% for teacher guides.

The respondents appeared to hold widely divergent views on educational materials, as was illustrated by this selection of their written comments:

- A decision maker – "*this seems like it could be efficiently accomplished through one of numerous partner entities*."

- An educator, in contrast – "*Education is the key toward the MPA success. The public needs to be aware of MPAs and supportive of ocean conservation in order for MPAs to be effective. K-12 education is especially important*."
- Request from one recreational consumptive stakeholder – "*bibliography of publications that link ecosystem based management with MPA and real life adaptive fisheries management…*"
- One scientist on potential problems with educational material – "…*these have GOT to be unbiased*."
- A second scientist suggested alternative educational resources –"*There are exciting activities emerging using online platforms like Google Earth and social networking sites (like e-Pals) that are far more innovative and engaging than some of these static options*,"
- While a third scientist suggested a more traditional, hands-on approach – "*The most effective outreach is direct exposure of K-12 students to the organisms and habitats by someone capable of identifying organisms and explaining phenomena. Open house days at a field station are in the experience of many of us*."

These results suggest that – with the exception of California-based monitoring objectives and protocols – most users thought that contextual information (documents that can provide background information to help put the monitoring results in context) would be nice to have, but is not necessary. Most also consider educational materials a low priority, and/or should be included as links to existing resources that do not violate any perception of bias. Section 5.2 further discusses issues and options related to providing user access to ancillary, contextual and educational information.

3.4 Information Syntheses

Users' preferred levels of information synthesis, ranging from raw data to final synthesis findings, have important implications for the IMS framework and how data will be processed, transformed, and/or presented into a useful form. Information about preferred presentation format for synthesized information (maps, graphs, etc.) is also important. Insights derived from survey questions related to preferred synthesis level and format (Section 6, Appendix A) will help inform the design of an information model for the IMS that describes the flow of data from collection (raw data), through processing, summary, reporting, and final presentation.

Section 6 of the survey (Appendix A) presented queries about synthesis and presentation types in a nested fashion, so that only a subset of users was asked particular questions, depending on their responses to the initial gateway question (Figure 3-10). Consequently not all respondents were shown every question, and the results below are presented in terms of the number of *eligible* respondents shown each question.

Synthesis Level: Overview

Among the 481 respondents who answered the gateway question asking what level of information synthesis they would like to access (left side of Figure 3-10), almost 80% selected each of the three most highly synthesized information products – information summaries, scientific reports, and summary tables (Figure 3-11). Far fewer expressed an interest in detailed data tables and summary statistics (48.7%) or raw data (27.4%) (Figure 3-11).

Initial Question

I would like access to the following types of MPA monitoring information (pick all that apply):

- ☐ Information summaries, reports for non-specialists, or key findings as they relate to the goals and objectives of specific or multiple MPAs.
- ☐ Scientific reports that describe monitoring results and conclusions drawn from the collected data.
- ☐ Summary tables, presentation, graphs, images, videos, or maps.
- ☐ Detailed tables of data, statistical summaries, compiled databases.
- ☐ Raw field/laboratory data with associated calibration or validation information.

Follow-Up Options

What options for information summaries that are automatically generated would you be interested in accessing?

- ☐ Data Tables
- ☐ Graphs
- ☐ Photographs
- ☐ Videos
- ☐ Maps

What options for processed and compiled data would you be interested in accessing?

- ☐ Databases
- ☐ Statistical Summaries

Assuming you want to download raw data, select the option that best describes your expectations.

- ☐ I would prefer extracting data in one file, and would expect that the data have been standardized across all fields.
- ☐ I would prefer downloading data files from the original sources, even though there may be differences, as long as the data standards are well documented.
- ☐ I would prefer being able to select a data standard that meets my own specifications.
- ☐ I'm not sure.

Please select the most common data formats you have downloaded to use in the past.

- ☐ Text files
- ☐ Spreadsheets
- ☐ Databases
- ☐ GIS file formats
- ☐ Don't generally download data

What level of metadata (methods, calibration, documentation, and validation information) should be included in the Monitoring Enterprise database?

- ☐ Standardized metadata, including data source, methodology, validation, exclusions, etc. for every dataset.
- ☐ Summary metadata/methods for every dataset.
- ☐ Methods described in monitoring reports or publications.
- ☐ Programmatic methods and standards documents.
- ☐ Other

Figure 3-10. Structure of survey questions on desired synthesis level; respondent selection of the bottom three options in response to the initial gateway question on the left resulted in the respondent being shown the follow-up options framed on the right. Respondents were asked to rank each option on the right as to its level of usefulness (e.g., from "essential" to "not needed"). See Appendix A.

This pattern was generally consistent across archetypes with the following provisions. Scientists selected all options more frequently than did other archetypes except information summaries, which they selected less frequently (Figure 3-11). Conversely, policy-related individuals selected information summaries more frequently and raw data less frequently than did any other archetype. The frequency with which commercial fishing stakeholders selected raw data was equal to that of scientists and higher than other archetypes (Figure 3-11). Almost all of the resource managers expressed an interest in information summaries and summary tables.

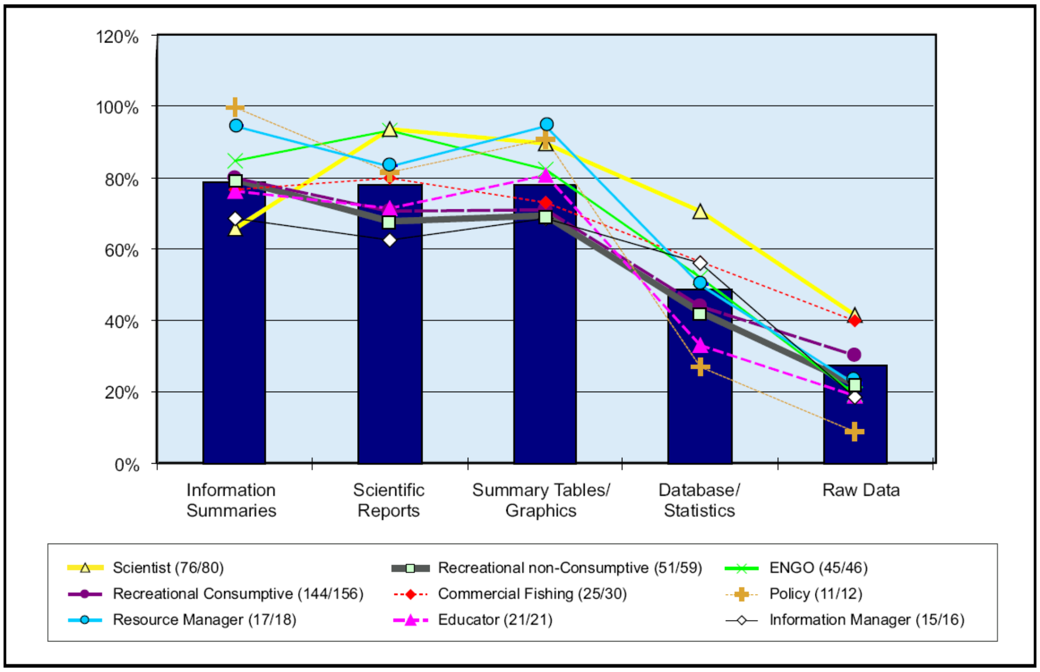


Figure 3-11. Percent of respondents who indicated an interest in the specified levels of information synthesis. Columns indicate the percent of respondents averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey for each archetype. The number of respondents selecting at least one option in this category was 481/519 (92.7%).

Specific comments provided by respondents about information synthesis lend perspective to the perceived differences among archetype and are further discussed in Section 5.2. One commercial fishing stakeholder commented "*I would like access to all information taken from them [the MPAs]*," while recreational stakeholders called for "*Independent science monitoring*" and "*Non-selective scientific input.*" These comments suggest at least these stakeholders want transparency in the development and interpretation of monitoring data. "*Scientists in my organization would want the raw data and more detailed formats*," commented an information manager, underscoring the potential interest in the scientific community in using raw monitoring data as a research resource.

Synthesis Level Follow-Up: Summary Tables and Graphs

The survey asked those respondents who indicated an interest in summary tables and graphs to rate the different kinds of summary formats (Table 3-7). This set of respondents presumably desired more detailed summary information than would be available in highly synthesized results, suggesting that although they need detail, they do not necessarily want to do any information processing themselves. Thus, to service this need, either the IMS or someone external to the system would need to do the data processing and presentation. Among the 405 eligible respondents, more than 70% considered maps to be essential (Table 3-7). Close to half of the respondents also considered graphs (46.9%) and tables (45.4%) to be essential data summary formats. Overall, greater than 50% of respondents judged all data presentation formats except videos as essential or very useful. Fewer than half (43.1%) rated videos as essential or very useful, and approximately 10% said they were not needed.

Table 3-7. Percent of Eligible Respondents Selecting Specific Types of Summary Data Presentation Formats

Summary Data Presentation Format	Essential	Very Useful	Nice to Have	Not Needed	Not Sure
Maps	70.0%	23.4%	3.9%	0.2%	0.7%
Graphs	47.1%	37.7%	11.4%	0.7%	1.0%
Tables	45.7%	32.8%	15.4%	1.5%	1.7%
Photos	23.8%	40.3%	29.0%	3.0%	1.0%
Videos	14.1%	29.0%	38.2%	10.2%	2.0%

The total number of eligible respondents selecting at least one option was 405/519 (78%).

Comparison across archetypes showed all assigned an essential rank to maps more frequently than to the other summary presentation formats (Figure 3-12). Scientists ranked maps, graphs, and data tables as essential with approximately equal frequency. Although the majority of information managers, like scientists, ranked both maps and data tables as essential, far fewer (~20%) assigned this rank to graphs. In keeping with their role of communicating scientific information to more varied audiences, the two presentation formats most frequently rated essential by educators were maps and photographs.

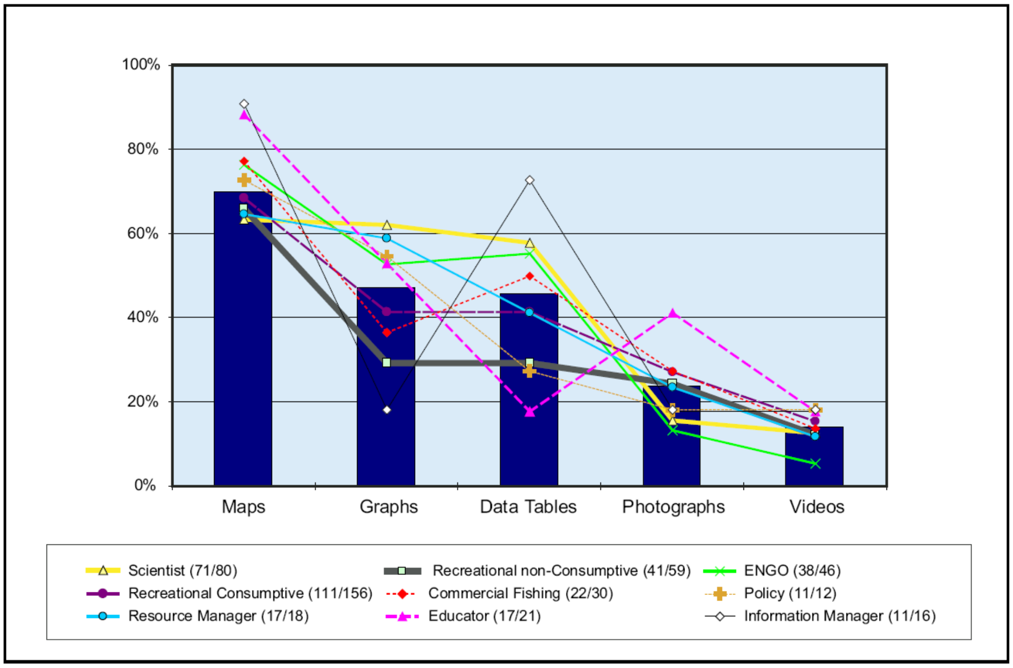


Figure 3-12. Percent of respondents who rated each summary format as essential for the IMS. Columns indicate the percent of respondents who answered "essential" averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey for each archetype. Figure is based on response of 335 people who selected "essential" for at least one summary format.

These results have important implications for the IMS framework. The clear preference for maps suggests that, at a minimum, there should be a mapping display capability. A key design issue will be whether the summary data products are created outside of the IMS (as separate products by a scientist or data processor), or if there is a process (application) that is used to generate standard summary data products. The latter option will necessitate the software, coding, and maintenance (of the database and code) in order for this capability to be maintained. Finally, the results suggest that photographs and videos are of least interest, unless, of course, the primary monitoring data are collected in these formats.

Synthesis Level Follow-Up: Detailed Data

As shown in Figure 3-10 (page 27), the survey asked for more specific information from those respondents who indicated they were interested in detailed data tables, statistical summaries, or compiled databases (253 people, 48.7% of all respondents). Eligible respondents were asked to rate the relative importance of databases or statistical summaries (Table 3-8). Almost all respondents to this question indicated that statistical summaries (91.7%) and databases (82.2%) would be essential or very useful.

Table 3-8. Eligible Respondents' Ranking of Specific Detailed Data Formats

Information Type	Essential	Very Useful	Nice to Have	Not Needed	Not Sure
Statistical Summaries	47.4%	44.3%	5.1%	0.0%	2.0%
Databases	43.1%	39.1%	13.8%	0.0%	2.0%

The total number of eligible respondents selecting at least one option was 253/519 (48.7%).

Most of the archetypes rated statistical summaries and databases as essential with approximately equal frequency. Information managers and policy people rated databases as essential more frequently than statistical summaries. Resource managers and educators showed the reverse pattern.

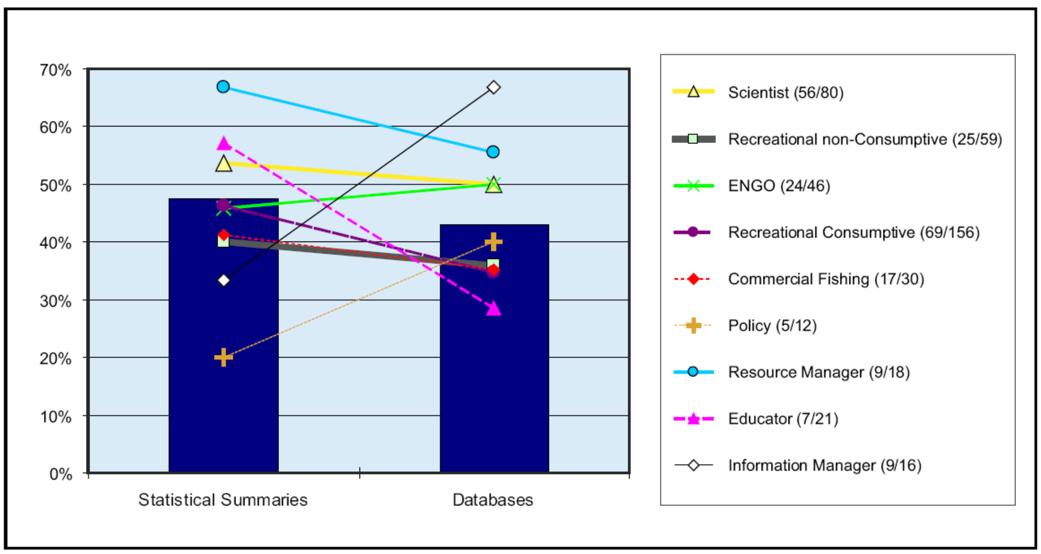


Figure 3-13. Percent of respondents who indicated it was "essential" for the IMS to provide specific types of detailed information. Columns indicate the percent of respondents who answered "essential" averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey for each archetype. Figure is based on response of 157 people who selected "essential" for at least one type of detailed format.

These results show that there is an audience for more detailed data; about half (49%) of the survey respondents distributed across all of the archetypes identified it as an overall need for the IMS (Figure 3-12), and most of these individuals indicated that both statistical summaries and databases would be essential or very useful (Table 3-8). The decision to store all the data in consistent, normalized database to facilitate querying and statistical calculations has implications related to the database and infrastructure aspects of the IMS framework. For example, centralized data management can be relatively labor-intensive and costly, so alternative approaches for facilitating access might be worth exploring. This may not satisfy the desires of all users, such as the respondent who suggested that "*The data should be made available in a format that allows it to be re-purposed by other web-based applications*."

Synthesis Level Follow-Up: Raw Data

As shown in Figure 3-10, people who selected raw data in the initial gateway question were then shown three more detailed follow-up questions about preferred download method, file format, and metadata style. The purpose was to identify those respondents who would presumably be doing their own data processing, and would need both the raw data as well as documentation about how those data were collected and organized. The eligible respondents who selected raw data comprised only about a third (142 people, 27.4%) of the respondent pool, but comprised a more diverse audience than anticipated. The follow-up questions were somewhat technical about downloading preferences, metadata, etc., so the number of individuals who selected "not sure" for these questions was relatively high, as discussed below.

The first question about download method was intended to gather more information for database management and standardization issues of the IMS framework. The question was geared to assess the expectations of those who would want to download raw data for their own data processing purposes. Most (98.6%) of the eligible respondents expressed an opinion by selecting one of the following options as shown in Table 3-9:

- I would prefer extracting data in one file, and would expect that the data have been standardized across all fields (units, nomenclature, etc.).
- I would prefer being able to select a data standard that meets my own specifications.
- I would prefer downloading data files from the original sources, even though there may be a differences (in units, etc.), as long as the data standards are well-documented.
- I'm not sure.

Each of these options has implications for how the monitoring data would be managed: from a centralized, standardized database (first choice), to a more distributed approach (third choice). The second option (customizing a data standard) would require a programmed feature to convert data to the user specifications, and would also be the most flexible (and most complex) approach. These results suggest that data standardization is important, but development of tools for user-defined data standards would not be necessary.

Table 3-9. Percent of Eligible Respondents' Expectations for Specific Download Method

Archetype (number answered)	Standardized	Custom Standard	Original Source	Not Sure
Citizen (7)	42.9%	42.9%	14.3%	
Decision Maker (1)		100.0%		
Educator (4)	25.0%	25.0%	50.0%	
ENGO[1] (9)	44.4%	22.2%	22.2%	11.1%
Fisher-Commercial (12)	50.0%	8.3%	33.3%	8.3%
Information Manager (3)	66.7%		33.3%	
Local Coastal Manager (1)				100.0%
Policy (1)				100.0%
Recreation-Consumptive (47)	31.9%	19.1%	23.4%	25.5%
Recreation-non-Consumptive (12)	58.3%	25.0%		16.7%
Resource Manager (4)	75.0%		25.0%	
Scientist (34)	44.1%	29.4%	20.6%	5.9%
Student (4)		75.0%		25.0%
Tribal (1)	100.0%			
Average (140)	40.7%	23.6%	20.7%	15.0%

[1]Environmental Non-Governmental Organization

The total number of eligible respondents selecting at least one option was 140/142 (98.6%).

The second question posed to the eligible respondents asked about their experience downloading data in different formats (text files, spreadsheets, databases, GIS format, don't download data) (Figure 3-10). Respondents were allowed to select more than one option. Most of the eligible respondents (96.5%) provided information on their experience with various downloaded file formats (Table 3-10). Between 15 and 25% of commercial fishermen and recreational users noted that they do not download data (Table 3-8). This result suggests that some users are interested in making sure that raw data are accessible for transparency, but may not download the data themselves.

Table 3-10. Percent of Eligible Respondents' Experience with Specific Download File Formats

Archetype (number answered)	Text Files	Spread-sheets	Databases	GIS Format	Don't Download
Citizen (7)	75.0%	87.5%	75.0%	37.5%	
Decision Maker (1)		100.0%	100.0%		
Educator (4)	75.0%	75.0%	25.0%	25.0%	
ENGO[1] (9)	77.8%	100.0%	77.8%	77.8%	
Fisher-Commercial (12)	66.7%	58.3%	33.3%	41.7%	25.0%
Information Manager (3)	100.0%	100.0%	100.0%	100.0%	
Local Coastal Manager (1)	0.0%	100.0%	0.0%	100.0%	
Policy (1)	100.0%	100.0%	100.0%	100.0%	
Recreation-Consumptive (44)	57.4%	70.2%	55.3%	27.7%	19.1%
Recreation-non-Consumptive (12)	76.9%	76.9%	53.8%	15.4%	15.4%
Resource Manager (4)	100.0%	100.0%	100.0%	50.0%	
Scientist (34)	90.9%	90.9%	57.6%	51.5%	3.0%
Student (4)	80.0%	80.0%	60.0%	60.0%	
Tribal (1)	100.0%	100.0%		100.0%	
All Respondents (137)	75.9%	83.2%	59.9%	43.1%	10.9%

[1]Environmental Non-Governmental Organization

The total number of eligible respondents selecting at least one option was 137/142 (96.5%).

The majority of respondents had experience with downloading data in spreadsheets and text files (Table 3-10). Fewer noted experience using databases and in GIS format. Because of the small number of respondents in many archetype categories, it is difficult to draw comparisons across archetypes.

The final question sought perspectives on metadata from those who indicated an interest in raw data. Respondents were allowed to select one or more of the following categories:

- Standardized ("Standardized metadata, including data source, methodology, validation, exclusions, etc. for every dataset.");
- Summary ("Methods described in monitoring reports or publications.");
- Methods in reports ("Summary metadata/method for every dataset.");
- Programmatic methods ("Programmatic methods and standards document.");
- Other (a blank space was included for text entry with this option).

Most (93%) of the eligible respondents expressed an opinion about what metadata were needed (Table 3-11). Standardized metadata was the category recommended by the highest proportion, followed by detailed methods descriptions in reports ("Reports" in Table 3-11). Less than half indicated that programmatic methods were needed. Many people selected more than one or all of the methods; this is consistent with advice from two stakeholders who said that "*all available*" metadata should be included.

Table 3-11. Percent of Eligible Respondents' Perspective on Metadata

Archetype (number answered)	Standardized	Summary	Reports	Programmatic Methods
Citizen (7)	71.4%	71.4%	85.7%	57.1%
Decision Maker (1)	100.0%	0.0%	0.0%	0.0%
Educator (4)	100.0%	100.0%	100.0%	25.0%
ENGO[1] (9)	77.8%	66.7%	66.7%	44.4%
Fisher-Commercial (12)	91.7%	25.0%	66.7%	33.3%
Information Manager (3)	66.7%	100.0%	66.7%	33.3%
Local Coastal Manager (1)	0.0%	0.0%	100.0%	100.0%
Policy (1)	100.0%	100.0%	100.0%	100.0%
Recreation-Consumptive (41)	73.2%	63.4%	61.0%	39.0%
Recreation-non-Consumptive (10)	100.0%	40.0%	50.0%	30.0%
Resource Manager (4)	100.0%	75.0%	50.0%	50.0%
Scientist (34)	91.2%	55.9%	58.8%	38.2%
Student (4)	75.0%	100.0%	75.0%	75.0%
Tribal (1)	100.0%	100.0%	100.0%	100.0%
Total (132)	83.3%	59.8%	63.6%	40.9%

[1]Environmental Non-Governmental Organization
The total number of eligible respondents selecting at least one option was 132/142 (93%).

In summary, although only about a third of the survey respondents indicated an interest in raw data (Figure 3-11 above), the follow-up questions suggested an additional reason for making raw data available. Specifically, certain respondents who indicated limited experience downloading data and metadata still felt that raw data should be available, presumably to ensure transparency in the evaluation of MPA condition, performance, and impacts. This issue of data transparency arises in several sections of the report and is discussed in Section 5.2. One solution that would fulfill this need may be to ensure access to raw data, but not necessarily through a managed internet interface (i.e., via CD, or other external data product). The need expressed by the most technical users for access to standardized data files (with normalized units, etc.) and metadata could be fulfilled either by including a managed back-end database component as part of the IMS, or adopting data standards, then enforcing them on the data providers. The advantages and disadvantages of these two options are discussed further in Section 5.3.

3.5 Interactivity

Interactivity: Overview

Two parts of the survey – Section 14 ("Data Analyzers") and Section 12 ("Communicate and Collaborate") – explored issues relevant to the IMS' interactivity (See Appendix A). The purpose of these sections was to determine what tools and utilities (applications) users would like to have on the site.

Interactivity: Data Analysis Tool Preferences

Section 14 of the survey ("Data Analyzers") used a gateway question to divide respondents into those who want to interact with the data online, those who want to access, download, and work with the data using their own software, and those who consider online tools to be unnecessary (Figure 3-14). Respondents who selected the second or third option shown in Figure 3-14 indicated an interest in some level of interactivity or analysis, and they were shown follow-up questions asking them to rate specific analysis tools.

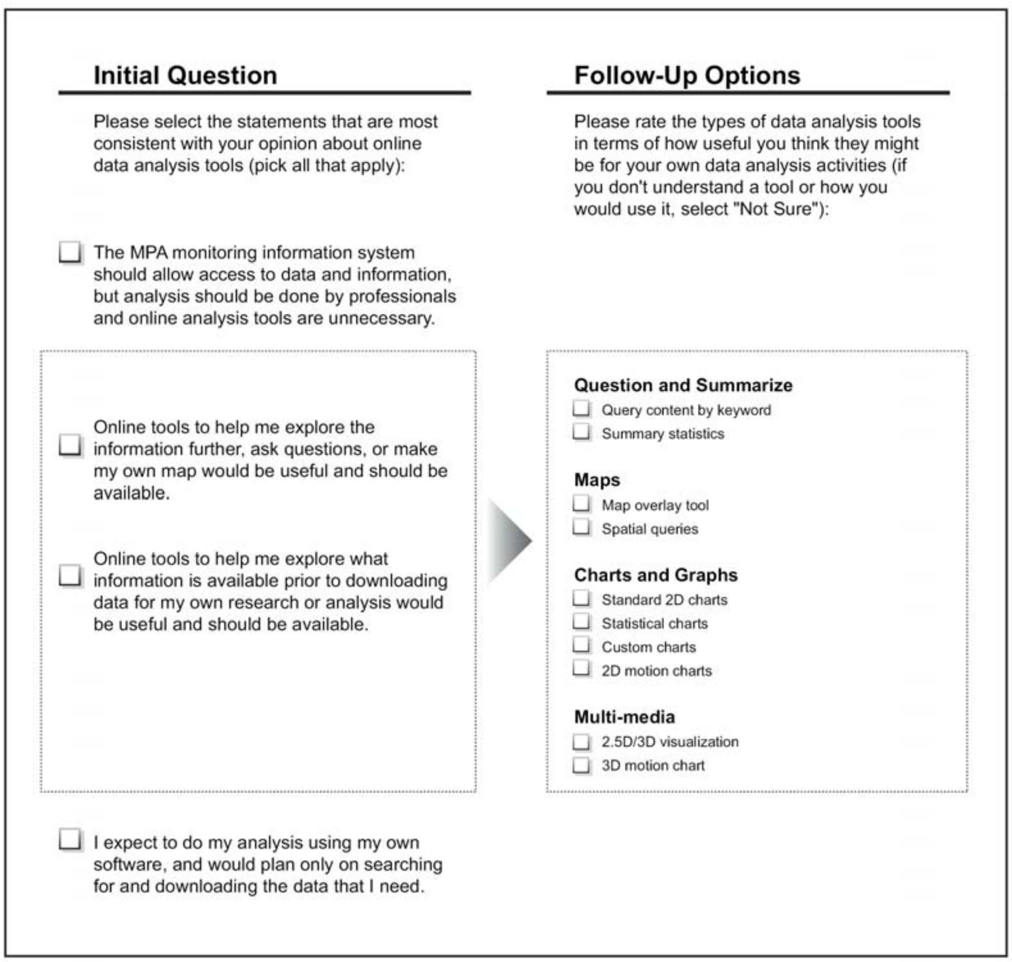


Figure 3-14. Interactivity gateway question; responses to the initial question options framed on the left lead to the respondent being shown the follow-up options framed on the right.

The initial question was available to all survey respondents and 73.6% answered it. Of these, a little over half indicated that they would like exploration and analysis tools (“Explore” in Figure 3-15); the most complex level of system interactivity) or have an interest in basic tools to browse and search for data (“Browse” in Figure 3-15). Overall, 75% of the people who answered the gateway question picked Explore and/or Browse. In contrast, 36.9% indicated they thought online tools were unnecessary (40.5% when calculated as the mean percent per archetypes), and only 19.4% (20.5% mean percent per archetype) selected the final option indicating they only need to download data to do their own analyses (Figure 3-15).

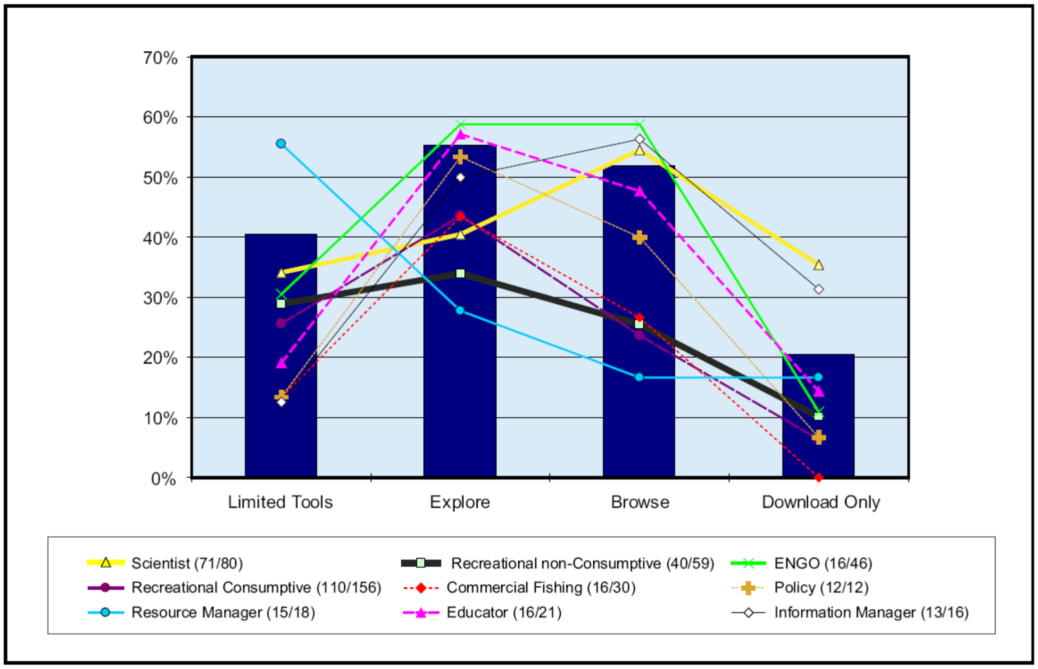


Figure 3-15. Percent of respondents who indicated a preference for each level of interactivity to work with data from the IMS. Columns indicate the percent of respondents averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey for each archetype. The number of respondents selecting at least one option in this category was 382/519 (73.6%).

Interestingly, two thirds of the resource managers indicated that data analysis should be “left to the professionals,” the highest percentage among all of the archetypes. In comparison to other archetypes, the greatest proportion of ENGO stakeholders specified an interest in both exploration and browsing tools.

Follow-on questions, shown to those who selected explore or browse, examined their relative interest in various tools having differing levels of complexity (Figure 3-14). The tools were grouped and presented in order from simplest to most complex as follows (see also Appendix A):

- Question and Summarize (keyword search; summary statistics);
- Maps (Map overlay; spatial queries);
- Charts and Graphs (2D charts; statistical charts; custom charts);
- Multi-media (2D motion chart; 3D visualization; 3D motion chart).

Results showing ratings for each tool type are illustrated in Figure 3-16. Overall, a map overlay capability was the most popular tool type, with 59.3% of eligible respondents citing this function as essential, followed closely by a keyword query tool (54.3%; Figure 3-16). 3D tools were both the least familiar (almost a third of respondents indicated they were not sure about these tools), and least frequently deemed essential (11%).

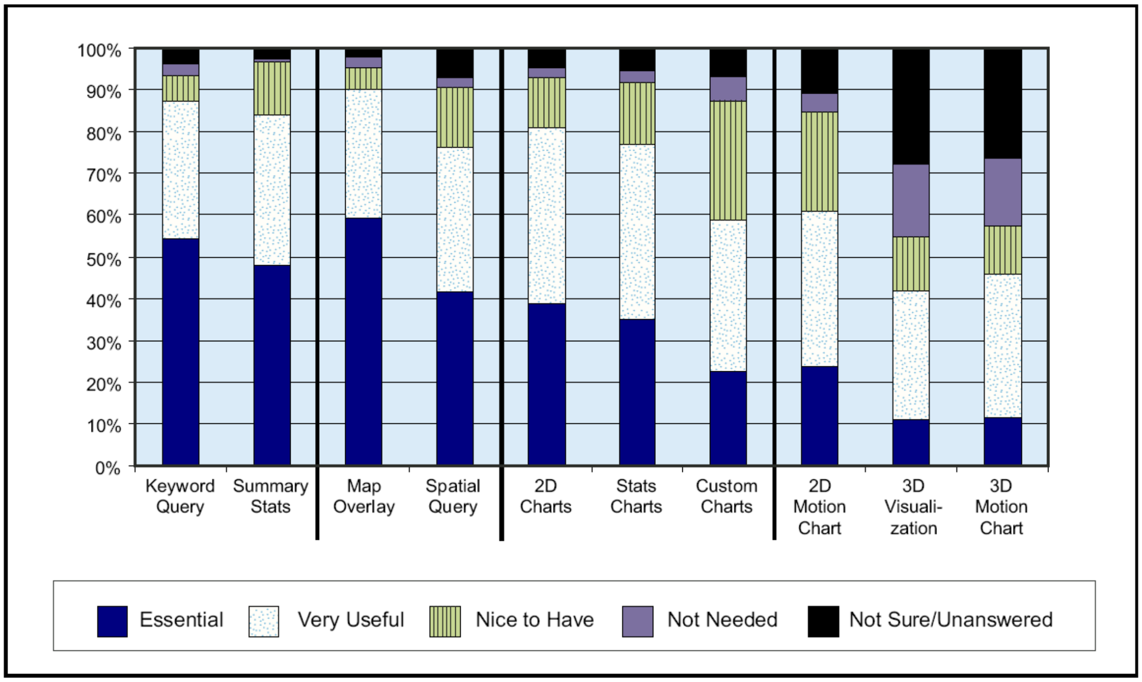


Figure 3-16. Relative rating of each tool type by those respondents indicating an interest in data visualization and analysis tools. The number of eligible respondents selecting at least one option in this category was 287/382 (75.1%).

These results show that basic keyword search and mapping tools are perceived to be the most critical. Search tools will be relatively easy to implement. The map overlay tool suggests the IMS will need to include a spatial component. Less interest exists in high-end visualization tools, although, at the same time, there was much uncertainty about what these tools can do.

Comments provided by respondents to this section revealed that people who have experience with more complex data analysis and visualization tools believe they should be included if they are targeted for specific uses and well-implemented. Sample responses include: "*Once available, these features become essential;*" "*These kinds of tools will take thought and time to develop into meaningful, simple ways of displaying data, but they can be very powerful visualizations and modern technology can be fairly easily used for this now.*"

In comparison to other archetypes (Figure 3-17), a greater percent of people in the commercial fishing, recreational consumptive, and policy-related archetypes showed an interest in the more complicated tools. Information managers, resource managers, educators showed a particularly high level of interest in map overlays. Resource managers showed a surprising lack of interest in statistical charts (Figure 3-17).

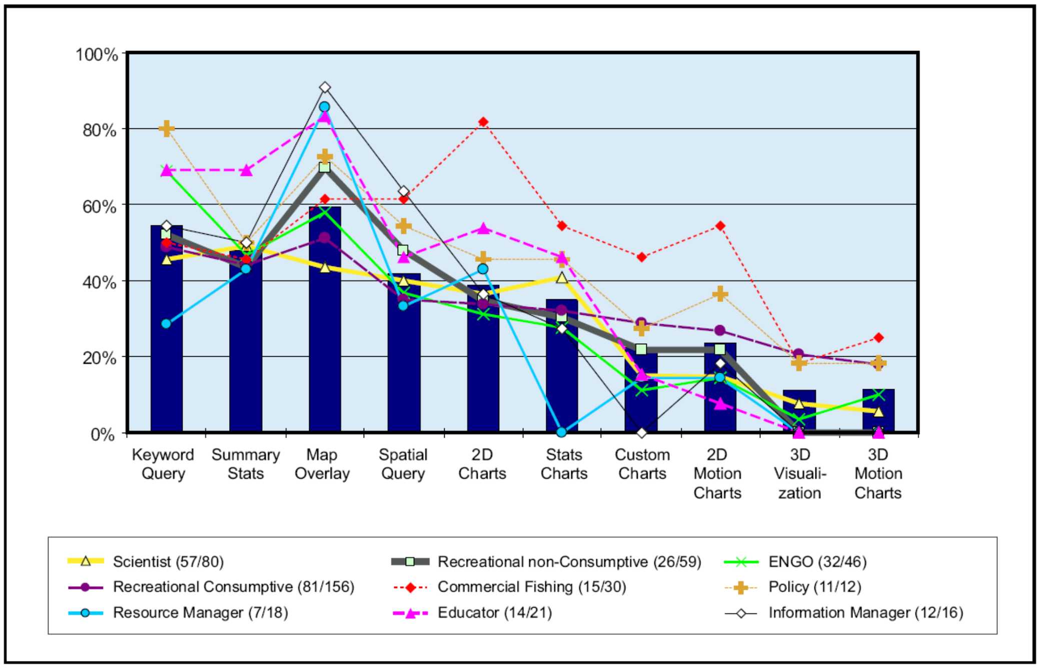


Figure 3-17. Percent of respondents who indicated a certain tool type was essential for the IMS. Columns show the percent of respondents who answered essential averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey in each archetype. Results are based on the responses of 219 people who selected essential for at least one tool type.

Particularly interesting comments offered in this section included:

- *"…balance' of tool set with respect to K-12 education, public education, NGO and other scientific uses…avoid deluxe offerings that can't be updated easily…try something out and gather metrics for a period of time…"*
- *"Tools to view results in context with regulations – simple map view tool, e.g., not necessarily linked to database, just showing results."*
- *"Here is a place for new innovation, open source, simple tools (non-scientific)."*

Interactivity: Communicate and Collaborate

The "Communicate & Collaborate" section of the survey (Section 12, Appendix A) provided a second source of information regarding respondents' opinions about levels of interactivity and online tools by asking the respondents to rate various tools and resources for communication. All respondents were eligible to answer all questions in this section and >80% responded to each question. The communication and collaboration tools examined are shown in Figures 3-18, 3-19, and 3-20, and include:

- Collaboration/Networking tools (contacts list; collaborative tools; discussion forums/wiki; blogs; decision support tools);
- Meeting/Conference Information Tools (announcements; meeting minutes; webinars);
- Interactive Educational Tools (Question & Answer for Students; Chat with a Scientist).

Among the various collaboration and networking tools, the one most often identified as essential was a searchable contact list of people involved in MPA monitoring (Figure 3-18). Despite the many good reviews of MarineMap (an on-line MPA planning tool designed to support the MLPA Initiative process, see www.marinemap.org) gathered by the survey (Section 3-6), the Decision Support Tools option was identified as essential by only 26.6%, although this was the second highest ranked collaboration/network tool.

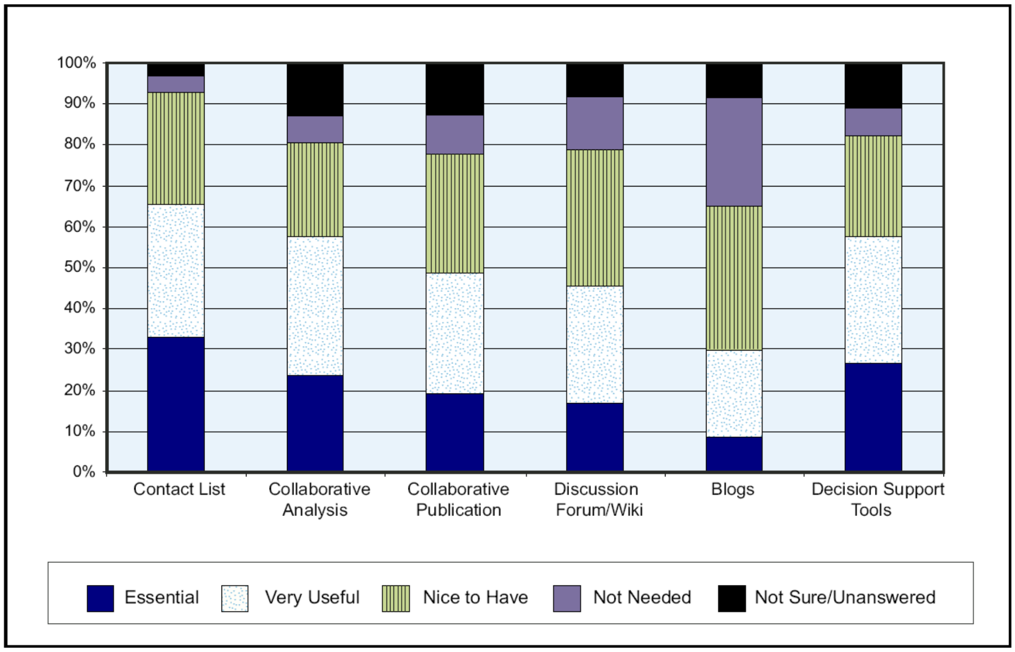


Figure 3-18. Distribution of all responses that indicated interest in specific types of communication and networking tools. The number of respondents selecting at least one option in this category was 419/519 (80.7%).

The social networking options (wiki and blog) were the communication tools most frequently ranked nice to have, but also the most frequently ranked not needed. This discrepancy possibly reflects differing philosophies or users' experiences with these kinds of newer communication methods.

Because the collaboration functions (analysis, publication) were most frequently ranked very useful, we analyzed the response by archetype (Figure 3-19). The distribution confirmed our assumption that scientists would be interested in these tools, but resource managers and information managers also indicated a high level of interest, as did, to a lesser extent, ENGO and recreational non-consumptive stakeholders. Comments provided in this section, reinforced this apparent interest in certain tools that facilitate intra-group communication and transparency. For example:

- "*collaborate commercial fisherman and scientists together*" and "*outreach to fishing members;*"
- "*I believe that one site can serve all by collaboration and integrating the best of the best tools into a central MPA portal;*"
- [Provide] "*...conflict information…what researcher with a financial interest in grants is also providing non-public information…*?" (see also Section 5.2).

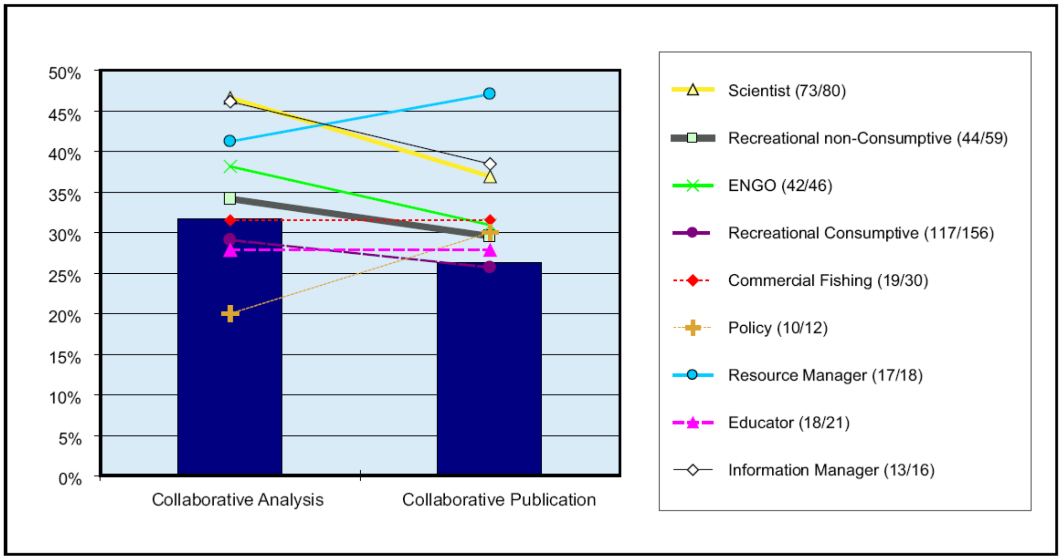


Figure 3-19. Percent of respondents indicating an interest in collaborative tools. Columns indicate the percent of respondents who answered very useful averaged across all archetypes; numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey in each archetype. The figure is based on the responses of 168 people who selected essential for at least one collaborative tool shown.

Although the audience for collaborative tools may be relatively small, creative deployment of such tools might enhance communication within and among various user audiences. From the perspective of IMS design, the costs and difficulty of implementing communication and collaboration tools vary widely; for example, a simple contact list is relatively easy to implement, while decision support tools require a larger investment in resources and time.

Among the various types of meeting/conference support tools, respondents most frequently rated public service announcements (Figure 3-20). None of the other types of tools was rated essential by more than a third of respondents (Figure 3-20).

The comments provided on conferencing tools called for easier access to webcasts of meetings: "S*ave and post videos of past Meeting/Conferences… so people can watch the meetings later in the day, week, or month*." "*All should be archived, with fairly detailed agendas to indicate what happened when so that it is easy to listen/look at pieces of interest*." Some made suggestions to ensure that conference proceedings were more accessible, such as by providing searchable transcripts of webcasts or a table of contents or index.

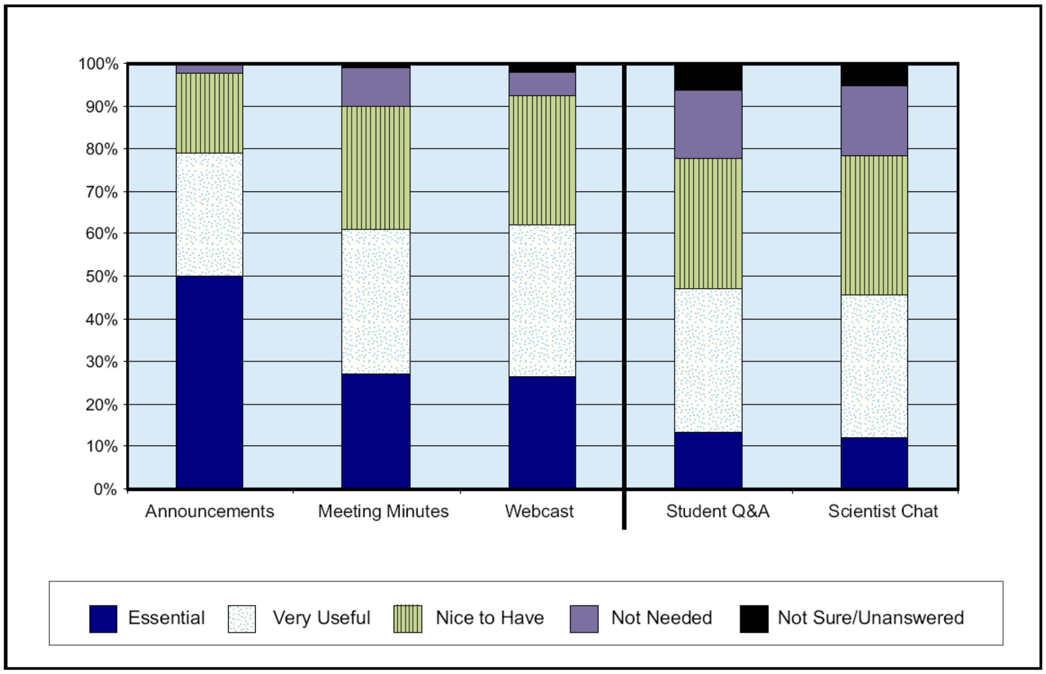


Figure 3-20. Ratings assigned by respondents to various conference support tools (left of line) and educational utilities (right of line). The number of respondents selecting at least one option in this category was 427/519 (82.3%).

The educational utilities (Student Question and Answer; Chat with a Scientist) were among the communication/collaboration options least frequently rated as essential (only blogs had a lower rating). More respondents rated these tools not needed than rated them essential (Figure 3-20). Comments in this section reflected the diverse opinions on including educational utilities: "*Teachers and students need to know what they can do in MPAs if they take field trips;*" "*The above forms of outreach and education are VERY time consuming and do not generally result in any tangible return to the program*."

Overall, the results on "Interactivity" suggest that less complex analytical tools were of greater interest to users than the "Communicate and Collaborate" tools, with the exception of public service announcements. More than three quarters of the respondents indicated that the map overlay, keyword search, summary statistics, and basic charting tools, as well as announcements, were essential or very useful. Respondents most commonly selected very useful or nice to have, rather than essential, for the "Communicate and Collaborate" functions. Several communication tools ranked higher, however, than the more complicated analysis tools (e.g., visualization tools). Several tools, of both categories, were considered to be not needed by more than 10% of the respondents, potentially indicating their uncertainty about how these tools might be useful (e.g., visualization tools, blogs).

These findings suggest that care should be taken in selecting utilities for the IMS. A utility to search for information, either via a map or keyword, is the most critical function. Additional data analysis or communication tools can be added after weighing the development costs against the benefits for various audiences. Section 5.3 provides additional discussion on interactivity relative to IMS design and development.

3.6 Human-Computer Interface

Several sections of the survey addressed the ways in which respondents interact with computers and other devices that access the internet. Findings from these sections will be used primarily for the IMS framework element related to interface (Table 1-2). Important considerations in designing the interface include: user preferences and opinions of existing websites; need for internet access by users who rely on languages other than English or have disabilities; and whether and how many users access the internet through devices other than the standard desktop computer. Results for Human-Computer Interface issues were derived from Sections 10 (website review), 11 (access/devices), and 13 (language/disability) of the survey (Appendix A).

Website Review and Experiences

All survey respondents were given the option of reviewing one or more websites. Respondents could review websites that they use routinely or were provided with a list of suggested marine-related websites. A total of 158 people chose to review at least one website. MarineMap was the website most commonly reviewed (20 reviews), followed by those of PISCO (the Partnership for Interdisciplinary Studies of Coastal Oceans, a long-term ecosystem research and monitoring program) and Reef Check (a non-profit organization) at 14 each (Table 3-12).

Table 3-12. Websites Selected by More than One Person for Review

Website Name	URL	Number of Reviews
Marine Map	http://marinemap.org	20
PISCO	http://www.piscoweb.org	14
ReefCheck	http://www.reefcheck.org	14
CalCOFI	http://www.calcofi.org	8
BIOS	http://bios.dfg.ca.gov	7
MARINe	http://www.marine.gov	7
PacFIN	http://www.psmfc.org/pacfin	6
RecFin	http://www.recfin.org	6
Protect Planet Ocean	http://www.protectplanetocean.org	4
SiMON	http://www.sanctuarysimon.org	4
MLPA	http://www.dfg.ca.gov/mlpa	3
NOAA	http://www.noaa.gov	3
State of the Salmon	http://www.stateofthesalmon.org	3
CA Department of Fish and Game	http://www.dfg.ca.gov	2
Google Earth/Google Ocean	http://earth.google.com/ocean	2
Magic Seaweed	http://magicseaweed.com	2
MBARI Oasis	http://www.mbari.org/oasis	2
NCEAS	http://www.nceas.ucsb.edu	2
North American Fishing Club	http://www.fishingclub.com	2
Ocean Observing System (South Coast)	http://www.sccoos.org	2
Pacific Fishery Management Council	http://www.pcouncil.org	2
SCCWRP	http://www.sccwrp.org	2
Surfline	http://www.surfline.com	2

Respondents were asked to rate the websites on a variety of characteristics, ranking them from poor to excellent or not applicable/no opinion. The evaluated characteristics were:

- Data access/download
- Data upload
- Design, look and feel
- Ease of finding information
- Metadata/documentation
- Quality of content
- Tools or utilities
- Graphs and charts
- Mapping information

Respondents also were prompted to provide open-ended comments about each website.

Results of this survey section revealed which sites most effectively perform functions similar to the ones that the IMS might require. For defining website interface styles preferred by various users, however, the most relevant results were: a) who chose to review which websites; b) what attributes of the website they dislike; and c) open comments. The analysis below focuses on these results.

At least one individual from each archetype chose to review at least one website, except for the tribal respondents. The highest response rates were from information managers and students (68.8% and 66.7%, respectively). Scientists (47.5%), recreational non-consumptive stakeholders (44.1%), and resource managers (38.9%) were the next most responsive groups.

Following the review of specific websites, the survey asked respondents to rate problems they have encountered in using websites related to the website design or to information access. The rating options ranged from bothersome (least negative), to annoying, irksome, and finally prohibitive (i.e., that this issue would make them exit the site). The rank of "prohibitive" was of most interest, as it identified issues that would be sufficiently problematic to cause the loss of some audience members. The majority of the survey respondents (78%) elected to answer at least one of the questions in this series (See Section 11 of the survey in Appendix A).

In evaluating aspects of website design, respondents assigned the highest negative rating to sites that are too slow to load, with 48% assigning a prohibitive rank to this quality. The first three problems – a site that was disorganized, made excessive use of gadgets, or provided poor site navigation – were considered either irksome or prohibitive by about two thirds of respondents. The problem of a site being too busy, with too many options, was least negative, with only 8% assigning this quality a prohibitive rating. Fewer than ten percent of respondents either had no experience with these kinds of problems or were not sure.

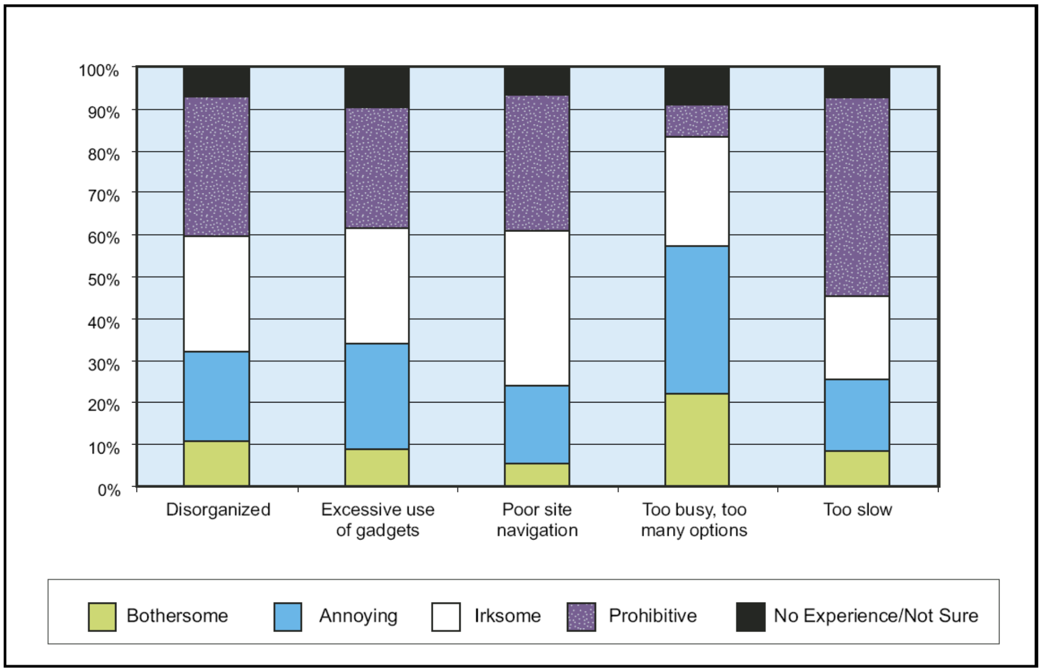


Figure 3-21. Respondent rating of website design problems. The number of respondents selecting at least one option in this category was 406/519 (78.2%).

These results have implications for the architecture of the IMS website. For example, if the applications selected to retrieve and display data are slow, the results suggest that the almost half of the audience will exit in frustration.

In considering the problems encountered with information access shown in Figure 3-22, the respondents most frequently assigned a prohibitive rating to "lack of useful data." The other problems with information access – unavailable or disorganized data, lack of documentation, or absence of ways to select data – were most often rated as irksome, with smaller percentages of respondents describing their experiences as bothersome or prohibitive. Fewer than 15 percent of respondents either had no experience or were not sure about each of the various information access problems.

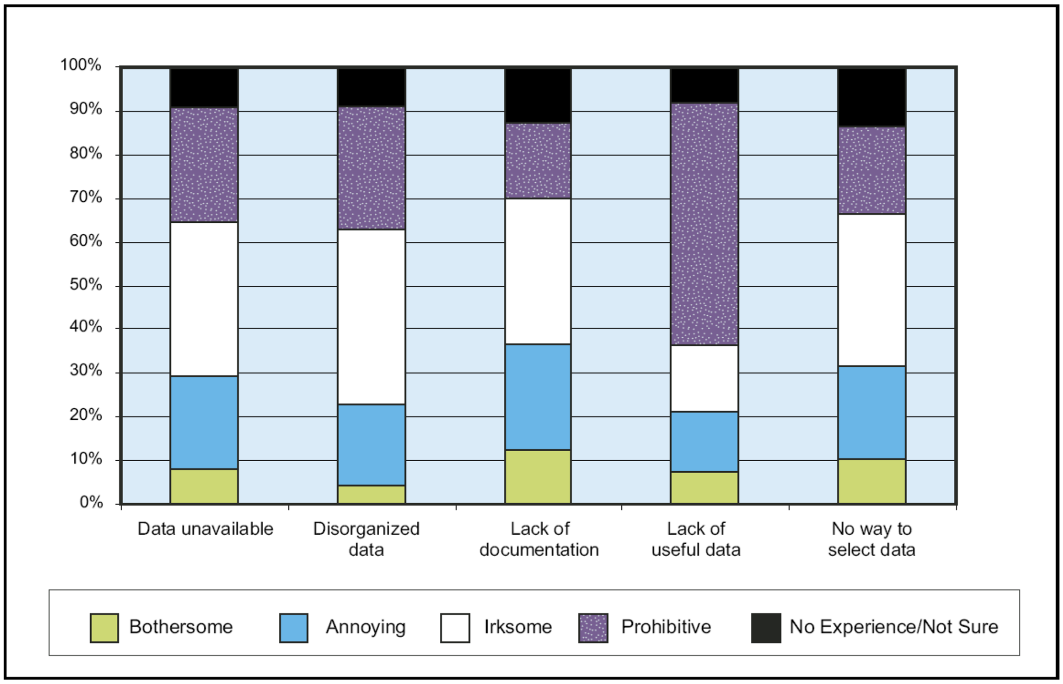


Figure 3-22. Respondent rating of website information access problems. The number of respondents selecting at least one option in this category was 406/519 (78.2%).

These results have implications for the IMS architecture, in particular for selection of the database and modular infrastructure for data retrieval. If the site provides data, it will not only need to enable efficient data retrieval, but also to inform the user about which data are available (scale, attributes) and how the user is filtering the data through selection criteria. The results suggest that respondents have had relatively little trouble finding or accessing data. In contrast, they have experienced some frustration once data are received, because the data are not what they expected or needed (Figure 3-22). Although only about 15% of respondents assigned documentation a prohibitive rating, it may be that poor documentation is an underlying contributing factor for other information access problems that were more often rated prohibitive, such as disorganized data, the lack of useful data, or inability to select data.

Website Accessibility: Devices

Several questions in Section 11 of the survey asked about the respondents' means of accessing the internet. All of these questions were open to all respondents. The first sought to identify the range of devices employed by potential users and had a response rate of 83.2% (432/519). Respondents could select one or more options from the following list:

- Desktop computer from my home or office;
- Desktop computer from the library or other public place;
- Laptop computer using wireless access;
- Mobile, hand-held device;
- Other (open comment).

A clear majority of potential users (91.7% of those who responded to this survey section) access the internet from a desktop computer in their home or office. Many (56.2%) also use laptops. A much smaller portion of the respondent pool indicated they use a mobile device (13.4%) or a publicly available computer (9.9%). Representation of the different archetypes among the group of respondents that selected mobile, hand-held device generally mirrored that of the larger survey population. Respondents who volunteered comments about mobile devices were primarily stakeholders who use these devices to obtain real-time information about MPAs. Comments to the "Other" category identified various ways to access the internet: "*cell phone with data,*" "*I-phone,*" "*dial-up modem (slow),*" "*SFSU library.*"

The survey asked the users if they had a specific need for information formatted for a mobile application. This question generated comments from 53 respondents, about half (47%) from recreational consumptive stakeholders who cited a need for MPA-related information (boundaries, regulations, enforcement information) for use by GPS, navigational software, and for electronic charts. Eight scientists also commented that electronic charting information for field research would be useful. Several comments also cited the need for data to be available for specific devices (I-phone, Blackberry) or other mobile device applications (pod casts, I-phone apps, media released, "*webcasts of data by scientists*").

The above findings show that most users will access the website from home or office. Respondents who indicated interest in mobile devices were mostly interested in information about MPAs associated with MPA management and navigation, rather than scientific monitoring data. A minority of respondents will likely access the website using a slow connection and/or from a public venue.

Website Accessibility: Language and Disability Options

The survey asked all respondents several questions about whether the site should support languages other than English or should include special functions to support people with disabilities (Section 12).

In response to the query about whether the site should support languages other than English, most of the 420 respondents said that other languages were either not necessary (37.1%) or that they had no opinion (35.7%). Languages identified by the 114 respondents who indicated the site should support other languages included Spanish (101 people), Vietnamese (18) or Chinese (Mandarin, Cantonese; 16), Korean (7), Japanese (5), Tagalog (4), and French (3). Others mentioned once or twice include Italian, German, Russian, Portuguese, Hmong, Filipino, Cambodian, and Kmer. Several respondents pointed out that Asian and Spanish-speakers are "*heavy users of marine resources*" and one respondent requested "*Monthly publication available in print and in English, Vietnamese, Spanish and Tagalog.*"

In response to the query about whether the IMS should have support for people with disabilities, of the 417 people who answered this question, 9% selected no, 52.3% selected no opinion, and 38.6% selected yes. Those who selected yes were asked follow-on questions that are compatible with web design guidelines for people with disabilities, including Perceivable (information and interface components must be presentable in ways that users can perceive); Operable (user interface components and navigation must be operable by the user); and Understandable. Most (86-89%) of the eligible respondents (those that selected yes) indicated that at least one of the options for each guideline category was important.

The results indicated relatively moderate support for alternative languages (27.1% of the people who answered the question) and for the use of features for the disabled (38.6% of those who answered the question). These features thus should merit consideration in subsequent phases of the IMS design and are relevant to the infrastructure element of the IMS framework (Table 1-2). However, including such features also would impact the overall cost of site design and maintenance. Also, the possibility exists that the ultimate home for the IMS may have strict standards for optimizing access to the website. This is true of websites managed by the state of California. Interestingly, responses from the website review suggested that the state-sponsored sites, while achieving maximum access, are not as user-friendly as might be desired for the IMS. One comment addressed this issue head on: "*Fortunately, the Monitoring Enterprise is NOT within the agency (CDFG), so does not have to follow the website design rules that the state has; this should free up … to be a more intuitive, effective site.*"

3.7 Institutional

The final portion of the survey (Appendix A, Sections 14 & 15) sought to identify possible data sharing relationships, the availability and compatibility of existing data systems, and other advantages or roadblocks for data sharing and system integration. Information gathered in this section of the survey is most applicable to the infrastructure element of the IMS framework (Table 1-2).

This section subdivided respondents into those who consider themselves potential collectors of monitoring information or of related information. The section's initial gateway question was posed to all survey respondents, and 240 respondents selected at least one of the first four options (Figure 3-23):

- I have been, or expect to be, involved with the collection of MPA monitoring data.
- I, or my institution, collect data associated with or in the vicinity of MPAs that could provide valuable information relative to formal MPA monitoring.
- I have been providing, or might provide in the future, information from consumptive uses (e.g., commercial or recreational fishing).
- I have been involved with an NGO/citizen's monitoring program that collects data that could provide valuable information relative to formal MPA monitoring.
- I have experience uploading content on the internet.
- I do not expect that I will ever provide data to the MPA Monitoring Enterprise.

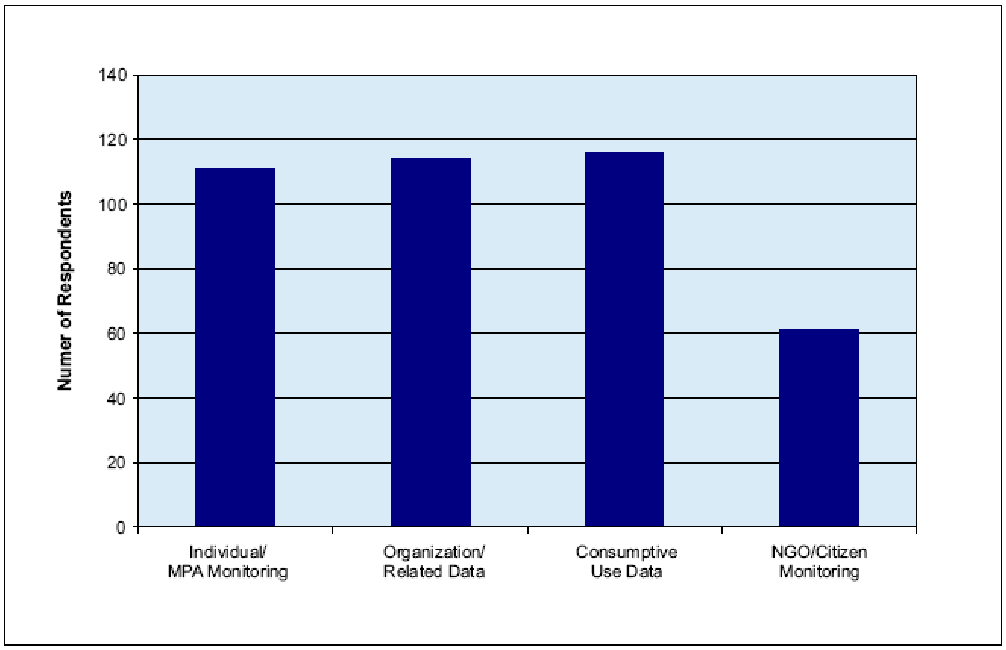


Figure 3-23. Number of respondents who indicated they were a potential source of monitoring information that could be provided to the IMS. Respondents could select more than one option. The number of respondents selecting at least one of these four options was 240/519 (46.2%).

Comparison across archetypes (Figure 3-24) shows scientists most often selected one of the first two options (35-46%), recreational consumptive stakeholders most often selected the third (41.4%), and ENGO stakeholders most often selected the last option (29.5%). Eleven of the 42 self-described citizens who took the survey cited interest in providing data, primarily in the consumptive use category (Figure 3-24). Representation of other archetypes in this analysis was too low to provide meaningful results.

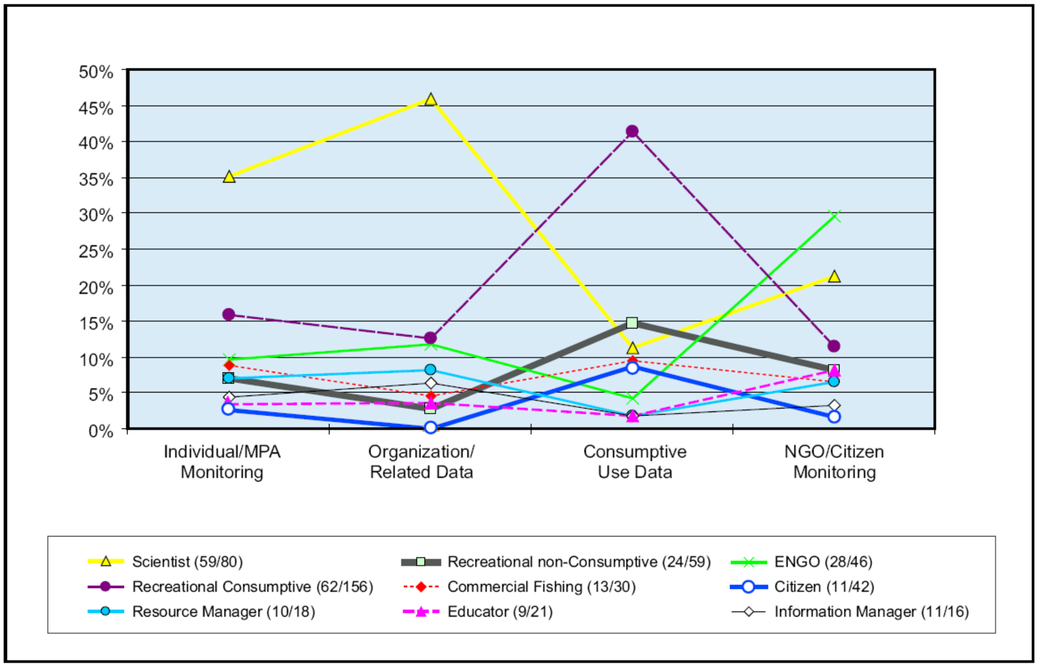


Figure 3-24. Percent of respondents in each archetype who indicated they were a potential source of monitoring information as a percent of the total of each data provision category. Numbers in parentheses show the number of people who answered this particular question/total number of people that took the survey in each archetype. The figure is based on the responses of the 240 people who selected at least one option as shown in Figure 3-25.

The 240 respondents who selected the options shown in Figures 3-23 and 3-24 were shown a follow-up question to explore whether they anticipated any issues that might limit or restrict data sharing:

- I will need to have a signed agreement to participate in data exchange.
- I or my institution would be willing to share data from our existing information management system.
- There is a firewall or other access restrictions to sharing data from my institution.
- I am not aware of any data sharing issues.
- I am not responsible for any data sharing issues.
- I have confidentiality or other access restriction concerns in submitting or sharing my data.

More than half of the eligible respondents indicated that they were unaware of and/or were not responsible for any data sharing issues (140 people). About 15% specified that they would need to have a signed agreement to engage in data sharing, while 9% anticipated there would be firewall or other access restriction problems. Comments offered about data sharing included: "*Confidential cultural resources information cannot be shared with public.*" "*Data collected by thesis students should be inaccessible until theses are completed.*" "*I cannot submit raw data because of disclosure issues, but can submit summarized data.*" "*Non-disclosure agreement needed for any sharing of confidential data.*"

A final question was posed to the 67 respondents that checked "I or my institution would be willing to share data from our existing information management system" from the question above (Appendix A, Section 15). Of these, 60 (89.6%) selected one or more of the following sharing methods: have their site as a link to the system (78.3%); send updates of their data (58.3%); serve as a data node to a distributed system (36.7%); and add metadata links of MPA monitoring data on their websites (30%). This information is directly applicable to the ultimate architecture of the IMS. (Note, however, that complete analysis of the different system integration options is beyond the scope of this UNA.) In total, 43 unique institutions were represented by these responses (Table 3-13).

Table 3-13. Organizational Affiliations of Respondents Indicating Willingness for Institutional Partnerships

Organization
BluePlanetDivers.org
Cabrillo National Monument
California Department of Fish and Game
California Polytechnic State University, San Luis Obispo Science and Ecosystem Alliance
California State University Monterey Bay
Central Coast Regional Water Quality Control Board
Channel Islands National Park
City of Laguna Beach
Delicasea Fish Live Seafood Commercial Fisherman
Ecotrust
Friends of Famosa Slough
Lawson's Landing Inc.
Long Marine Lab
Morro Bay National Estuary Program
MSI / MarineMap
National Park Service
NOAA MPA Center
NOAA National Marine Sanctuary Monterey Bay/Foundation
NOAA NMFS Southwest Fisheries Science Center
Orange County Coastkeeper
Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO)
Pepperdine University
Reef Check
Rough N Ready MWC
San Diego Coastkeeper
San Diego Freedivers
San Luis Obispo Coastkeeper
Sierra Club, American Cetacean Society, Cabrillo Marine Aquarium
Sierra Club, NCWN, RCWA, ARWPA, FOSC
Smithsonian Institution
South Laguna Civic Association and Laguna Bluebelt
Southern California Coastal Water Research Project (SCCWRP)
Surfrider Foundation
The Ocean Foundation
The Seadoc Society
Tijuana River National Estuarine Research Reserve
U.S. Fish and Wildlife Service
U.S. Minerals Management Service
Underwater Society of America
University of California at Santa Cruz
University of California Santa Barbara/Norris Rancho Marino Reserve
University of California Sea Grant Extension Program
University of California Sea Grant Moss Landing

These findings indicate that interest exists in providing information to the IMS, with results split among MPA monitoring, associated information, and consumptive uses. Considerations for the IMS architecture include potential data upload functionality (templates, etc.), as well as database standardization. The findings are also relevant to the themes of openness and confidentiality (see also Section 5.2).

4. Telephone Interview Results

Telephone interviews of 35 people were conducted between April 23, 2009 and June 1, 2009 using a script of questions based on the internet survey topics (Table 1-1). An initial quantitative analysis of the telephone interviews considered results according to archetype (assigned by Exa to each of the interviewees, Figure 4-1). Based on this information, we derived a set of themes that consistently arose in the interviews. Section 4 is organized around the survey topics, but also discusses the themes because they provide additional depth to the analysis. Section 5.2 discusses the themes in greater detail and integrates material from the telephone interviews with the online survey results, and Section 5.3 addresses related implications of the themes for the IMS framework elements.

A key difference between the telephone interviewees and the internet survey respondents was that many of the interviewees spoke for a group larger than themselves, and/or were asked to make generalizations about the larger potential user group. In addition, the interviewees were specifically asked their opinion about the critical audiences for the IMS and what priority they would assign to each audience (Section 4.1).

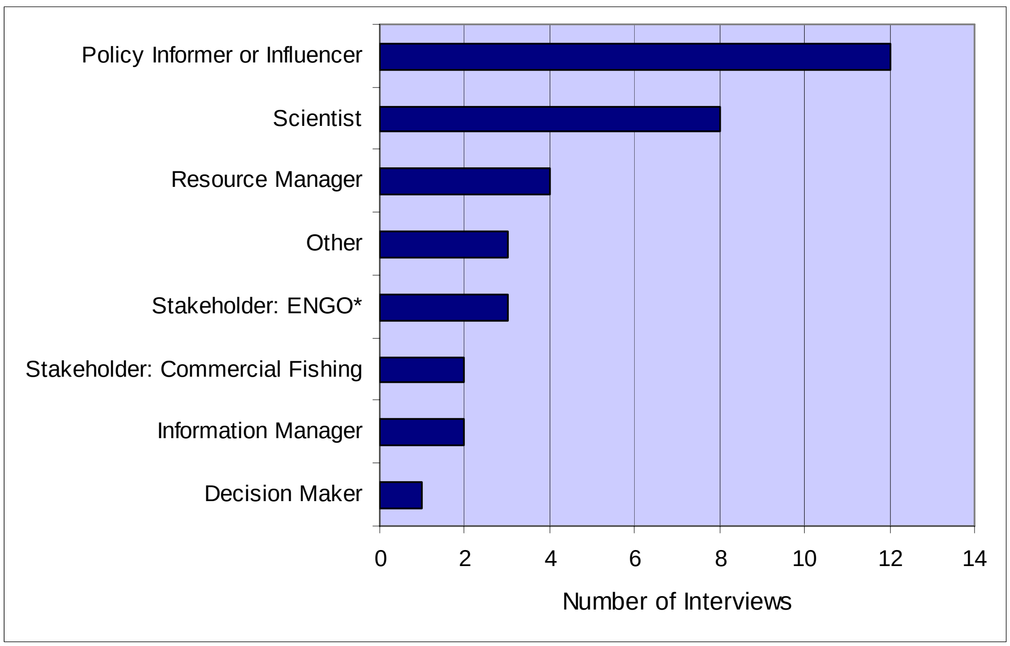


Figure 4-1. Distribution of telephone interviewees by archetype (*ENGO stands for Environmental Non-Governmental Organization).

4.1 Target Audience

All interviewees were asked about the expected target audience for the IMS, and 30 individuals (85.7%) provided an opinion. Descriptions of anticipated users fell into five general groups, consistent with a subset of the archetype classifications (archetypes were not specifically listed by the interviewers):

- Decision-makers
- Resource managers

- Scientists
- Stakeholders
- General public

A similar number of the interviewees (11-14) identified each of the above as a potentially important audience, and several prioritized the relative importance of each audience (Table 4-1). Opinions varied greatly among interviewees, even within a given archetype. One scientist indicated that scientists did not need the system, because they can exchange data and information through other existing channels. Another scientist, however, said the site should be primarily for scientists to exchange data.

Table 4-1. Number of Interviewees that Mentioned or Prioritized the Five Target Audiences

Priority Rank	Decision Maker	Resource Manager	Scientist	Public Outreach/ Education	Stakeholders
First Priority/Unranked[1]	14	14	13	11	13
Second Priority	5	4	5	7	10
Lower Priority	0	0	3	4	1
Not Needed	1	0	1	0	0
Total	20	18	22	22	24

[1]Fifteen of 30 interviewees who answered this question ranked their choices.

Ten interviewees mentioned the importance of engaging the general public through interactive content and highly simplified results. Three interviewees made a strong case for the importance of reaching the general pubic ("*taxpayers*") to inform them about the MPA monitoring process. These interviewees saw a link between stakeholder and public involvement and consequent support for decision-makers, resource managers, and the ongoing MLPA Initiative.

The above results suggest opinions vary widely about who is the primary audience for the IMS. Equally interesting were the differing opinions about which groups are lower priority or not needed: for example, whether the site should mainly support scientists *versus* those that felt that scientists were already experienced in data sharing, and so should be a lower priority. The inconsistent results provide no clear guidance for prioritizing the potential audiences for the IMS based on the user's role associated with MPAs. In fact, the interviews suggest that an alternate approach to defining user groups independent of role might help to provide a more rigorous method to prioritize user needs (hence, the development of User Personas presented in Section 5.1).

4.2 Information Content

Twenty-eight interviewees (80%) provided an opinion about what they see is the desired or critical information content of the system. The most consistent feedback (21 people) was the need for a highly synthesized metric, or simple graphical method, for representing the effectiveness of MPAs.

Among those citing the need for highly synthesized information, a subset (11) said that highly synthesized results were of no use (or worse, of potential misuse) without information about the context for those results. "*Highly synthesized information is important… but these simple results must be presented in some kind of context so that questions about the meaning of red and green are extrapolated…*" said one scientist, for example. By "context," most of these interviewees appeared to mean presenting summary metrics along with information about the original monitoring goals to "*…provide that linkage between the monitoring and the goals, the Plan and Progress*," as stated by one policy-related interviewee. Another suggested that relating monitoring information to "*the discrete objectives of individual MPAs is clearly important; …the original architect's intention for the site… demonstrates the results relative to the objective of each site*."

Not all interviewees shared this opinion about the need to provide the context for summary metrics, however. According to one interviewee who works in policy, for example, this information is already served by the CDFG website.

The next most common type of content, mentioned by 12 interviewees, was information about key species or habitats. Some responses were quite detailed, although rarely mentioning specific species or ecosystems, and were based on the interviewees' prior experiences with other MPA monitoring programs. Frequent suggestions we made of more general types of species or habitat metrics, such as "*measures of ecosystem health*" or "*abundances of target and non-targeted species*."

The other primary types of content mentioned by the interviewees included oceanographic conditions, human uses of MPAs (eight interviewees each), and information/updates on monitoring activities (six interviewees). Of those mentioning an interest in the IMS serving information about monitoring activities, two were scientists who suggested that it would be useful for scientists if the site identified who is doing what monitoring and where. The other four focused on uses of this information to inform a broader audience, including the general public, about current monitoring activities and anticipated results.

The timing and frequency of information release was a common theme that ran through responses to the question about content. For example, one policy person thought that "*infrequent update of monitoring activities*" was sufficient, because of the long time for results to be available. In contrast, one scientist believed regular updating of information content on the website would be critical to ensure the content would not get stale. However, this interviewee also thought that information on program activities would be sufficient to accomplish this objective. Timing in the interpretation and reporting of monitoring results also was addressed by one interviewee: "*An example is providing a time frame reference – how long would you expect some of these 'recovery' responses to take?*"

Although the interview format did not specifically pose questions about how frequently users expected to access the monitoring results, this issue arose in nine of the 35 interviews involving a varied set of archetypes. These interviewees anticipate that user needs for access to the MPA monitoring information will be infrequent (every five years was mentioned by four interviewees) and episodic.

The question of spatial scale was raised by the interviewers in 22 of the 35 interviews. This question paralleled the internet survey by asking interviewees about their interest in monitoring results at statewide, regional, local, or individual MPA scales. The interviewees were also encouraged to explore issues related to providing monitoring results at so many different scales.

Almost all (21) of the interviewees who responded to this question identified statewide as an important scale for presenting results. A subset (16) indicated that all monitoring scales were important. Those who emphasized a statewide scale framed this in terms of the network concept and decisions that would need to be made about made about adaptively managing the network. As one scientist put it: "*For future management (action), very likely changes to MPA will happen on regional and statewide scales*."

Those interviewees who indicated that presenting results for local MPAs would be important tended to focus on the fact that, as stated by one policy person, "*people will always have local interest.*" However, this same person also noted that the "*site should draw connections to statewide reserves*." One resource manager noted that "*in terms of analyzing the data, you have to compare one site to another*." Two interviewees gave feedback on how the spatial data could be organized on the site in response to questions about interactivity discussed below. Both suggested providing maps that allow users to drill down to finer and finer detail.

One or two interviewees mentioned each of the following content areas:

- Bathymetry
- Land use
- Anadromous fish
- Watershed information
- Stressors: oil spills, and non-fishing impacts and trends of stressors
- Fishery catch data
- Climate change data
- MLPA Initiative
- MPA information: access and restrictions; usage levels
- Monitoring protocols
- RFPs
- Contact information
- Links to other MPA programs worldwide
- Concept dictionary
- Definitions
- Educational material (videos/webinars)
- Underwater videos of surveys

Finally, three interviewees (one scientist, two policy) took a wider perspective on the information content issue. They felt that it was important that the site provide the critical audience (stakeholders, general public) with sufficient information to explain the larger story of MPAs, and the reasons behind monitoring. "*…a compelling story could be told on how people use these places, and how some are enhanced by using these MPAs, others are displaced*." "*The website needs to explain … what this monitoring data will show us and why it is important*." "*The … MLPA describes the economic heft of the coastal dependent economy: extractive uses, coastal tourism, etc…represents many billions of economic impact. The site should make clear the economic value of healthy ocean and coast and potential loss of money from a decline.*"

The above summary of desired content does not distinguish between content that would reside locally on the system versus content that could be accessed from another system or provided as links. And much of the specific feedback on information content is not directly relevant to IMS design, since these decisions will be made through a separate process. The interview results do, however, suggest that it will be important to report the MPA monitoring results within the context of the MPA goals and objectives, at relevant spatial scales, and with appropriate temporal frequency. Also, information on monitoring activities was cited as important for scientists, and/or as a way of informing the public about the monitoring process before data are ready to be presented.

4.3 Information Synthesis and Presentation

All interviewees were asked about the level of synthesis, or detail, of monitoring information that should be served by the IMS. Following the organization of the internet survey's gateway question, the interviewees were prompted for more detail depending on their answer about what level of synthesis they preferred. In particular, those who had special knowledge or experience managing detailed data, databases or raw data were further queried about their opinions related to data management issues.

Of 35 interviews, 29 expressed a clear opinion on their preferred level of synthesis. Their responses fell into three categories:

- Summary Reports / Synthesis (highly synthesized)
- Summary Data (moderately detailed)
- Raw Data (no synthesis)

Consistent with the internet survey feedback that highly synthetic content on the effectiveness of MPAs was important, 22 of the 29 interviewees said that either synthesis results and/or summary reports (e.g., annual reports) were the most important level of information, and in fact should be the primary product of the system (Table 4-2). Descriptive terms about this how this synthesis might be structured included: "*report card*", "*snapshot*", "*yes/no*", "*grades/colors/single metric*", "*summary report*", "*vital signs*."

Table 4-2. Number of Interviewees who Mentioned and Prioritized Each Synthesis Level

Priority Rank	Summary Reports/ Synthesis	Raw Data	Summary Data
Important	22	17	6
Lower priority	0	1	2
Not needed	1	4	0

Only one interviewee, a commercial fishing stakeholder, mentioned concerns about the site providing synthesized information because of potential errors or misinterpretation of rolled-up information. Eleven interviewees mentioned the need for error or uncertainty estimates to accompany the synthesized information, in the form of error bars, disclaimers, or other context information. One interviewee (a scientist) directly addressed the mechanism of converting raw data into summary information: "*…you need to have a strong communication link between the producers of data and the interpreters and producers of summaries. This is not a control function, but rather a conversation so that limitations and qualifications of data are clearly understood*."

Raw data was the second most commonly cited level of data synthesis that interviewees said the site should provide, although four interviewees specifically said raw data were not needed (Table 4-2). These four interviewees (two scientists, one resource manager, and one ENGO stakeholder) cited various reasons for excluding raw data from the site. These included: concerns about misinterpretation; potential restrictions placed on releasing raw data by funders or collaborators; the belief that raw data should be shared among scientists through other means than a public website; and the opinion that highly synthesized data was the only appropriate level of information for the site to serve.

A common theme among interviewees related to raw data, in addition to its scientific uses, is that it needs to be openly accessible to make the MPA monitoring transparent. Access to raw data, however, did not equate necessarily to providing the data over a website. For example, one interviewee suggested that "*it* *is appropriate to make data available to those who need it, without distributing it directly from a web-based download."* Making raw data available via CD or report appendices were some suggested ideas.

Providing raw data also raised the issue of data access limitations due to proprietary or publication interests. Five interviewees mentioned concerns regarding sensitive or confidential data. One interviewee noted that commercial fishing data would useful for the assessment of MPA effectiveness, but that confidentiality concerns would have to be addressed. One resource manager mentioned the need to assess the political context and sensitivity of some information, and suggested that protocols may be needed to: "*create a process for data release that preserves academic rights to original publication while maintaining transparency. Consider holding data in process of interpretation for a period of time before release of raw data; for selected groups require a binding agreement of data use and release or have a panel decide if they can have early access to raw data with a binding agreement; report what data is collected but mark 'not ready for release'*." At the same time, at least one interviewee was unsympathetic to such restraints saying "*Data should be available to everyone without proprietary limitations or long time delays*."

Of the 29 interviewees who expressed an opinion on data synthesis level, only eight specifically mentioned a level somewhere between highly synthesized and raw data. One resource manager provided an opinion about limitations on summary information while supporting the concept of tools for users to create their own summary information: "*The problem with synthesized products is that they are never quite exactly what you want. It may leave out some critical piece of information for making correlations. Therefore, it is helpful to add some flexibility on the user's end regarding which parameters and time periods to combine*."

Of the interviewees who mentioned data visualization (21), mapping was the method most commonly identified (16) as a desired method for the site, specifically a map on which the user can pick a region or an MPA and see the monitoring results for that site. Twelve interviewees mentioned the importance of showing summary information generically in a graphic or other visual presentation; six of these interviewees specifically mentioned the importance of showing results in the form of graphs or charts.

A few interviewees (6 out of 35) mentioned the need for the system to provide "tiered" levels of information or access. Although not directly asked, these interviewees suggested that multiple synthesis levels of data should be available, starting with the most synthesized, and then allowing the user to "drill down" through levels of detail, depending on their interest.

Managing or Serving Data

Interviewees who demonstrated they had specific knowledge about managing detailed data or serving raw data to groups of investigators were encouraged to further explore these issues as discussed below.

<u>Centralized vs. distributed data management</u> – Of the seven interviewees who mentioned a preference for a specific type of data management, three (two resource managers and an information manager, all associated with CDFG) preferred that the data be centralized, and four preferred a distributed data management approach (one scientist, one ENGO stakeholder, one resource manager, one "other"). The cited advantages of a centralized system included better control over the data, reduction of possible duplication, and also better integration with existing state information management infrastructure. The described advantages of the distributed system were that data producers would retain ownership of their source data, and that QA/QC and updates or modifications would be better facilitated with a distributed approach. Those who preferred a distributed system expected that the Monitoring Enterprise would provide access to the data as a "switchboard" or "clearinghouse." One interviewee (an ENGO stakeholder with information management experience) stated clearly the importance of developing a

database that can be distributed but functions as a centralized database: "*At a minimum, the data have to be stored in a relational database which is queryable - this is the absolute minimum standard*."

Standards – Nine interviewees mentioned the need for data standards, such as database tables, database fields, templates for data contributors, file formats, data communication protocols, and species taxonomy codes. Two recommended that the Monitoring Enterprise coordinate standards with existing monitoring programs and database managers (including California Department of Fish and Game [CDFG], State Water Resources Control Board), and four recommended adoption of existing databases standards rather than the creation of new ones. One interviewee felt strongly that the system should adhere to all CDFG system standards (hardware/software/programming), and suggested that the IMS would eventually end up within the CDFG.

Metadata and QA/QC Protocols – Eight interviewees mentioned the need for metadata and metadata standards; three recommended that all metadata be available on the site, even if the source data were not available for download. One interviewee expressed a desire for the metadata to be detailed enough for another researcher to conduct comparable survey in the same location. Five interviewees mentioned the need for QA/QC protocols.

The overall results of the interview segments on information synthesis suggest that users desire highly synthesized monitoring results and, to a slightly lesser extent, raw data. Information synthesis will need to include some estimation of uncertainty. Relatively few interviewees cautioned against serving raw data; however, because of the more limited audience for raw data, less expensive delivery approaches, such as on CD, might be worth exploring. Numerous suggestions were made to establish a map interface for viewing summary data. Relatively little interest appeared to exist in mid-level data summaries.

4.4 Interactivity

Interviewees who mentioned specific content and/or levels of synthesis of information were asked how they would envision accessing the information, and if they would be interested in querying or graphically viewing the data. Some interviewees were asked, in an open-ended way, if they had ideas about how interactive the site should be or, more generally, how important it would be to have tools or utilities to help users view or access the data.

Of the 35 interviewees, 22 provided input on the importance of different kinds of system interactivity for the site. Responses were categorized into four levels of interactivity, in increasing order of complexity:

- Static (no tools);
- Data download;
- Interactive query and view;
- Complex analysis or decision support tools.

The interviewee responses were bimodal with respect to the desired level of site interactivity. Nine said report links or other static presentations of results were sufficient, and that resources would be better used in other areas rather than in developing data access or visualization tools. "*Data synthesis is too complex for interactivity; do the analysis outside and then post results,*" said one scientist. According to another: "*…public funds should be used for the conveyance of synthesized information to a target audience; tools are of secondary importance*."

Of the 22 interviewees that indicated some level of data interactivity would be useful, capabilities for interactively viewing and querying the data in a limited way were mentioned most frequently (15) (Figure 4-1). Most (11) said this function was important. However, four specified that it would be nice to have, but was a lower priority than other levels of interactivity. A desire for interactivity with a mapping interface was mentioned specifically in three interviews. This group indicated that there was a "*value to putting 'cool tools' for data analysis and visualization for the public*." Another interviewer suggested that the "*Public will look for a graphical interface…*"

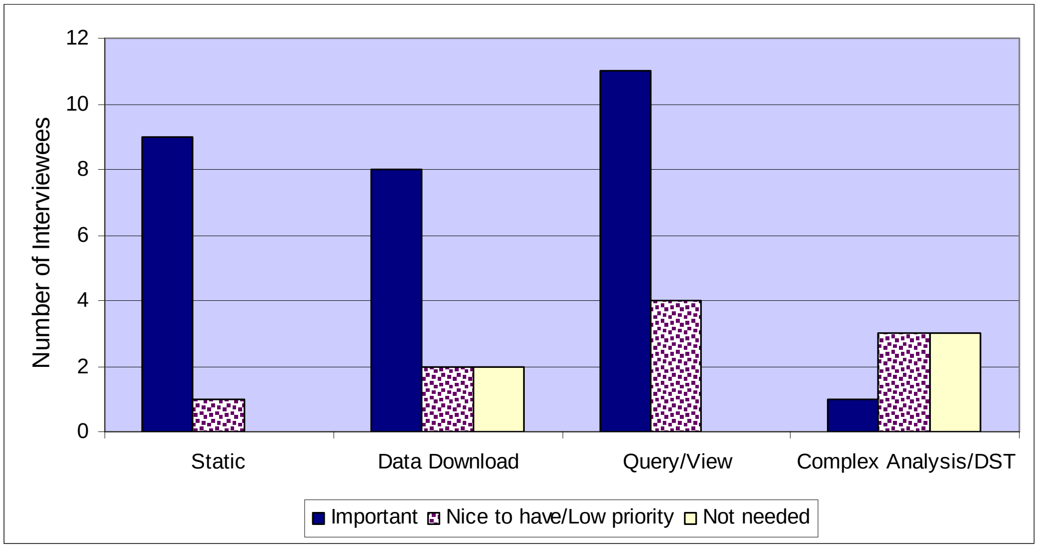


Figure 4-2. User needs for system interactivity as identified by telephone interviewees (22 of 35 individuals). DST refers to decision support tools.

Five interviewees suggested general or specific tools. In addition to being able to zoom in on a map interface, two specifically suggested it was critical to be able to query the database on key words. As suggested by one scientist "*…three species in seven MPAs…be able to query at this level with error bars*." A resource manager suggested tools that would conduct standardized assessments: "*…status and trends on a spatial basis, ability to zoom in on individual MPA.*"

Eight interviewees suggested that data download capability would important, and two others said that this type of function would be nice to have, but a lower priority. In contrast, two scientists specifically stated that raw data should not be downloadable due to concerns about data misuses, but they said summarized data would be acceptable for download.

Interest in complex on-line analysis or decision support tools was low, with only one person identifying this as an important need (Figure 4-2). Interviewees generally expected that people who need analytical capabilities would have their own tools and would prefer to conduct analysis using their own software.

One policy-related interviewee suggested a general approach to tool development, with the Monitoring Enterprise acting as a "*neutral partner*" for various clients that need specific tools: "*Data synthesis can be done as needed for specific clients that should pay for it…make custom-tailored information products, for example they can create data that would support decision support tools*." The idea of specialized tool development that might not be available for all users also was suggested by another interviewee, a resource manager, who suggested a "*… function to assist in the adaptive management process in putting together reports that are more detailed. While the web might not have all this functionality, the adaptive* *management portion could be used … for their purposes."* This suggests targeted partnerships between the Monitoring Enterprise and other vested partners.

Finally, without prompting, at least two interviewees suggested that intuitive data submittal tools (e.g., eBird) that allow input from citizen monitoring groups and/or the fishing community would be a way in which the IMS could facilitate partnerships and communication. Although the usefulness of these data is potentially circumscribed, such a function might go far in attracting certain users of the site. As one scientist pointed out, it is "*critical to have participation of fishing community in monitoring and evaluation, this leads to community-based management*."

Collaborative Functions

Ten of 35 interviewees mentioned collaborative functions. Of these, six felt that social networking tools such as blogs, discussion boards, and wikis were either unnecessary or inappropriate. They suggested that effort put into such tools was outside the main mission of the Monitoring Enterprise. In contrast, four interviewees said that high priority should be place on integrating newer technologies that are commonly used by younger users, including blogs and communication between stakeholders. "*The site should encourage open discussion…allow people to write their own opinions… blogs are conversational, insightful. Net meetings, chat rooms, it's all people having conversations*," said one policy-related interviewee. Another interviewee who works in policy suggested that the Monitoring Enterprise should "*try to take advantage of some of the existing social network tools…best if it looks like a modern communication interface*."

Only two scientists raised the issue of using the website for collaboration among scientists, but without specific information on what kind of tools they would require. In contrast, a resource manager said that "*Collaborative functions are a lot of work to maintain, might be best left to academics or other partners*."

Education/Outreach/PR

Sixteen of the 35 interviewees had specific comments about the educational, outreach, or public relations aspects of the website. Three suggested that providing good educational material is costly, and that it would be more cost-effective for the Monitoring Enterprise to link to existing materials or to form partnerships with entities that specialize in providing these types of materials.

Thirteen interviewees acknowledged the benefit of providing a site that is interesting and accessible to the general public and the education community, because it could increase the knowledge base and overall support for monitoring of MPAs. A majority of these (9) suggested either that this need did not require web-based utilities (FAQs, educational materials, etc.), or that any educational component must avoid an advocacy tone because of the need for the site to be credible and unbiased.

4.5 Human-Computer Interface

All interviewees were asked questions that solicited their views on what characteristics make a website effective *versus* ineffective or well- *versus* poorly-designed. Five of the interviews explored issues related to website-based management of marine-related data, because the interviewees had relevant expertise. Finally, all interviewees were asked about whether they might use mobile devices to find or receive information.

Website Design

Twelve of the 35 interviewees mentioned some aspect of the website usability or design. Most comments were very general, and included concepts like "*user-friendly*," or "*ease of use*." One interviewee recommended a structure that is "*deep and narrow*" rather than "*shallow and wide*."

Interviewees suggested two general models for the site design:

1) Simple, monitoring results only – Static website with limited information, updated only when new reports are available. As described by one interviewee: *"The first page should presents annual (or whatever time scale) data; these data should be available in a PDF of that information that can be downloaded and printed. The website should provide contact information for people that want to get involved, RFPs, etc."*
2) Multi-faceted, phased site – A more complex site, starting with some basic information summarizing the active monitoring activities, then building up to more complex data synthesis and mapping results reporting.

All interviewees who spoke to this issue agreed that the site design will be critical to encourage use. "Simple" was the most common design term, although different interviewees used the term in different ways. Three interviewees equated "simple" with "cost-effective" – and suggested a scaled-down site with limited interactivity. Six interviewees used the term simple to describe clear, well-organized, intuitive, or useful for the layperson. For example, one scientist with experience working with marine science-related websites equated the simplified synthesis of monitoring information to how the public accessed information about the weather. Three interviewees suggested that the site avoid using jargon, acronyms, or "*talking down*" to the audience.

One ENGO stakeholder with experience in websites suggested that a multi-faceted site was the optimum design: "*The IMS should have tiered levels of access… the most effective websites have multiple ways of access – map interface for more lay users; the next level for stakeholders with specific questions; more advanced user query (drop-down queries).*" One policy-related interviewee suggested that the site could initially be simple ("*FAQ is best for [stakeholders]*") but as the public got involved, the site could begin to serve additional needs. Another interviewee suggested that a help box with live response was most useful when searching for difficult information.

These results suggest two possible models for the website design, each of which has very different implications for the IMS framework issues (Section 5.3). The support for the streamlined model came from those concerned with the effective use of resources and a limited role for the IMS. The support for the more complex, multi-faceted site was from interviewees who were interested in engaging the potential user audience or who had deeper knowledge about website-based information management.

Reviews of Existing Websites

Twenty-three of the 35 interviewees provided feedback on their experiences with existing websites. Websites were chosen by the users themselves, not by the interviewers.

Several websites were discussed by more than one interviewee (Figure 4-2):

- MarineMap
- Google (specifically Google Ocean)
- California Cooperative Oceanic Fisheries Investigation (CalCOFI)
- Ocean Observing Sites (OOS)
- Recreational Fisheries Information Network (RecFIN)
- Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO)

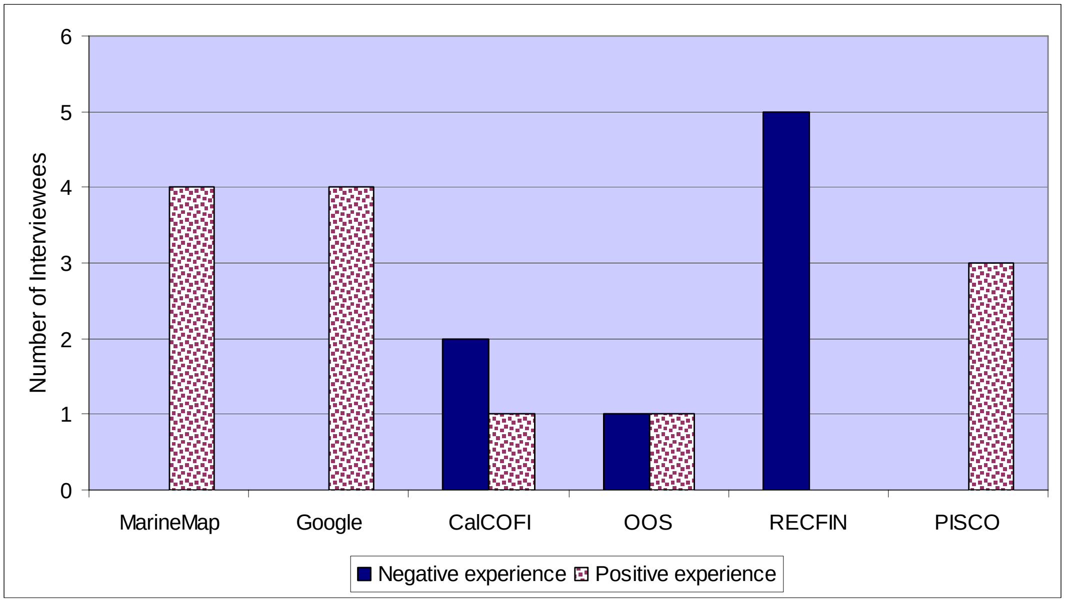


Figure 4-3. Number of telephone interviewees who reviewed each website.

Interviewees had positive experiences with MarineMap, Google, and PISCO websites and felt they were well-designed and easy to use. MarineMap, in particular, was cited as useful because "*all of the data are in one place that allows people to burrow down to different layers. In creating MPAs, allows people to do this interactively because it pulls everything together*." Two scientists and a manager thought finding information through the PISCO website was effective.

Interviewees had mixed opinions about CalCOFI and OOS websites. Those with positive opinions focused on the availability and quality of numerous data sets of interest. Those with negative opinions said information difficult to find or download. RecFIN was acknowledged as having a great deal of useful information, but all users, including experienced ones, had difficulty correctly finding and extracting the information they needed. In reviewing the OOS and RecFIN sites, one interviewer experienced with marine information websites commented that "*data clearinghouses … are difficult to see if you are getting raw or summary data…difficult to tell what the boundaries are*."

Feedback on state-based websites was also mixed. One policy-related interviewee noted that the poor design of California agency websites made it difficult to find information. In contrast, another policy-related interviewee thought that the CDFG site worked well for finding information on meetings and reports, and a decision-maker described the CDFG site as useful for data access. One ENGO stakeholder said they found it frustrating to try and find information on the MLPA Initiative site.

Interviewees offered opinions about several other sites, including:

- Southern California Coastal Water Research Project (SCCWRP), described by a policy-related person as "*works well, a little technical*"
- Reef Check California's Nearshore Ecosystem Database (NED), described by a resource manager as "*good, but NED is a little slow*"
- the NOAA/NOS charting sites, described by a scientist as "*geographic-based and tabular search do not synch well*".

These results are consistent with those from the internet survey (Section 5.3). First, a well-designed site ensures that, if data are served, the user can easily identify what data they want, and understand what data they are receiving (i.e., raw *versus* summary data). Processing speed is important, and over-use of scientific jargon should be avoided. Finally, the feedback on the effectiveness of California agency sites will be important if the architecture of the IMS will be integrated at some level with state agency information and/or hosting.

Client Device

Nine interviewees provided input on the question of access to the site using a hand-held mobile device (i.e., Smart Phones). Only two suggested that information served to a mobile hand-held device would be useful; both represented stakeholder groups (recreational consumptive; commercial fishing). Of the other seven interviewees, five commented that they had no use for data sent to their phones, while two others could see other potential users, particularly younger ones, possibly finding this helpful.

4.6 Institutional

The institutional home for the IMS will affect the architecture and design of the system (Table 1-2). Nineteen of 35 interviewees provided an opinion about the institutional home (or host) for the system. Seven interviewees said that the IMS must be maintained at an “independent” or “neutral” location; three specifically mentioned the Monitoring Enterprise (Table 4-3).

Table 4-3. Preferred Institutions for the IMS Mentioned by Interviewees

Preferred Institutional Home	Number of Mentions[1]
State Agency[2]	5
Independent/Neutral (including Monitoring Enterprise)	7
University	3
Monitoring Enterprise	3
Not State	7
Not University	2
Mixed Opinion	3

[1]Many interviewees identified several options.
[2]Most commonly California Department of Fish and Game.
Table based on 19 interviewees.

Opinions varied widely when interviewees were queried about whether the system should be housed within a California state agency. Five interviewees thought that the logical place for hosting the IMS was a state agency, rather than a university or an independent site. Two noted that longevity was a key attribute of state-controlled sites: “*For long-term stewardship…plan to hand off to a state agency*.” The availability of data standards and an existing infrastructure were other cited advantages of a state host. In contrast, three interviewees said they were concerned that the inflexibility of public agency processes would hinder information dissemination and website development. They expected that an independent entity, such as the Monitoring Enterprise, would be more flexible and responsive to user needs. One policy-related interviewee expressed concern that a state agency might be less than forthcoming: “*…perceives that there has been difficulty getting data from the state, specifically CDFG*.” Two interviewees specifically mentioned trust: “*There is a lot of mistrust of the government controlling data*.”

Trust was also an issue for one interviewee who specifically advised against a university-based IMS. "*There is a general distrust of academics and CDFG*;" another respondent suggested that "*the public will perceive the University as less biased.*" Two interviewees also cited the problem of data accessibility for a university-controlled site: "*If responsibility is given to the UC system, there would have to be a contingency that the data need to be available*." "*Academics have issues with … intellectual property*."

One scientist with a "*mixed opinion*" thought that state ownership would be a "*perfect solution*" except that the agencies (and universities) are "*bound by rules*." Finally, another scientist summarized the topic by invoking one of the consistent themes of the results (Section 5.2): "*Doesn't matter where…as long as it is trustworthy and accessible*."

Overall, these results suggest that while a state home might offer longevity and a mechanism to deal with access issues (disabilities), it also might restrict flexible and better (intuitive) site design. Universities have less bias but potential problems with accessibility. An independent institutional home is likely to be more-trusted, and would have greater flexibility to develop *a more intuitive, effective site*. But this approach also has the potential to be less cost-effective if it involves 'starting from scratch.'

5. Summary of Interpreted Internet Survey & Telephone Interview Results

The two information-gathering methods relied upon in the UNA – the internet survey and the telephone interviews – yielded complementary results. Section 5 draws together and synthesizes this material into a unified picture of user needs.

5.1 Audience as System Driver: User Personas

The critical next task in developing the IMS will be to clarify the system's objectives, and characterizing the audience will be an essential element of this task because the IMS will based on a user-centered design. The goal of this section of the UNA is to provide the Monitoring Enterprise with tools to enable prioritization of the anticipated primary, secondary, and lower priority tiers of users. Such prioritization will assist the Monitoring Enterprise as it makes decisions about how and whether to meet various audiences' needs within the constraints of finite resources.

The UNA started by using the archetype convention to group users according to their association with, or interest in, MPAs. The internet survey results were analyzed and presented by the percentage of each archetype that responded to the survey questions; the telephone interview results were discussed relative to the archetype of the person that was interviewed. Although these analyses provided useful results, there are several limitations to using archetype as the sole tool for prioritizing audiences.

First, the potential audience for the IMS is broad and diverse. Because the IMS does not yet exist, it is impossible to know whether the survey pool of archetypes reflects the actual population of future users. Using the raw count of respondents by archetype does not necessarily provide a reliable method to prioritize the audiences. Some archetypes were underrepresented and some overrepresented in the survey results relative to the initial estimations developed through the invitees on the contact list.

Second, survey respondents and interviewees within each archetype expressed widely varying preferences and opinions. Consequently, no way exists to directly relate archetype preferences to IMS framework elements shown in Table 1-2. Our initial assumption, that individuals in a given archetype would have similar needs, was not borne out by the results of the survey.

Finally, diversity within each archetype classification adds another level of complexity. No person is exclusively defined by their job or choice of recreation; thus identification by archetype is a useful but blunt tool, in the case of such highly variable audience characteristics, to neatly classify and prioritize.

Defining a limited set of "User Personas" provides an alternative to the archetype approach that avoids the shortcomings described above. To develop user personas based on the internet survey and telephone interview data, we grouped users who have a specific set of information use and preference attributes, regardless of archetype. Each group was then assigned a classification that summarized these attributes. The user persona serves as an idealized user that humanizes and provides insight into each potential audience for the IMS. This insight should facilitate the ultimate design of the IMS.

User Persona Methods

We selected two survey topics – Synthesis and Interactivity – to define the user personas, because these topics provided a tractable way to scale the answers and were strongly linked with IMS framework elements. Although potentially adding depth and detail to some design issues, we did not use Content or Human-Computer Interaction because they do not drive major system design decisions. Similarly, while institutional home will have a great impact on IMS infrastructure, our information on this topic was limited to respondents who were associated with potential partner institutions rather than the full set of potential users.

As shown in Table 5-1, we derived Synthesis (Section 6, Appendix A) and Interactivity (Sections 12 and 14, Appendix A) indices based on the most common sets of results from all of the questions on these topics in the internet survey. These indices were then coded on a scale of 1-4 (Table 5-1).

Table 5-1. Survey Topic Indices for User Persona Assignments

Survey Topic	Level	User Persona Index Description
Interactivity	1	No tools and/or some basic keyword search or social networking tools
	2	Tools to search for and/or download data of primary interest
	3	Tools to display and analyze data on-line
	4	All levels of tools and utilities
Synthesis	1	Highly synthesized information only
	2	Summary data and information products
	3	Raw data or detailed database, and/or scientific reports
	4	All levels of data synthesis

The internet and telephone survey results were analyzed for the two metrics, and an index assigned to each respondent or interviewee. Only those for whom we could assign an index for both metrics were used in the analysis that follows, including 428 of the internet respondents (82%) and 15 interviewees (42.8%). Figure 5-1 shows the number of people in each of the 16 potential combinations of the four synthesis and interactivity indices. The relative size of the circle for each combination corresponds to the number of individuals having that combination. Indices representing more than 20 individuals were color-coded into the four user personas that shared similar attributes.

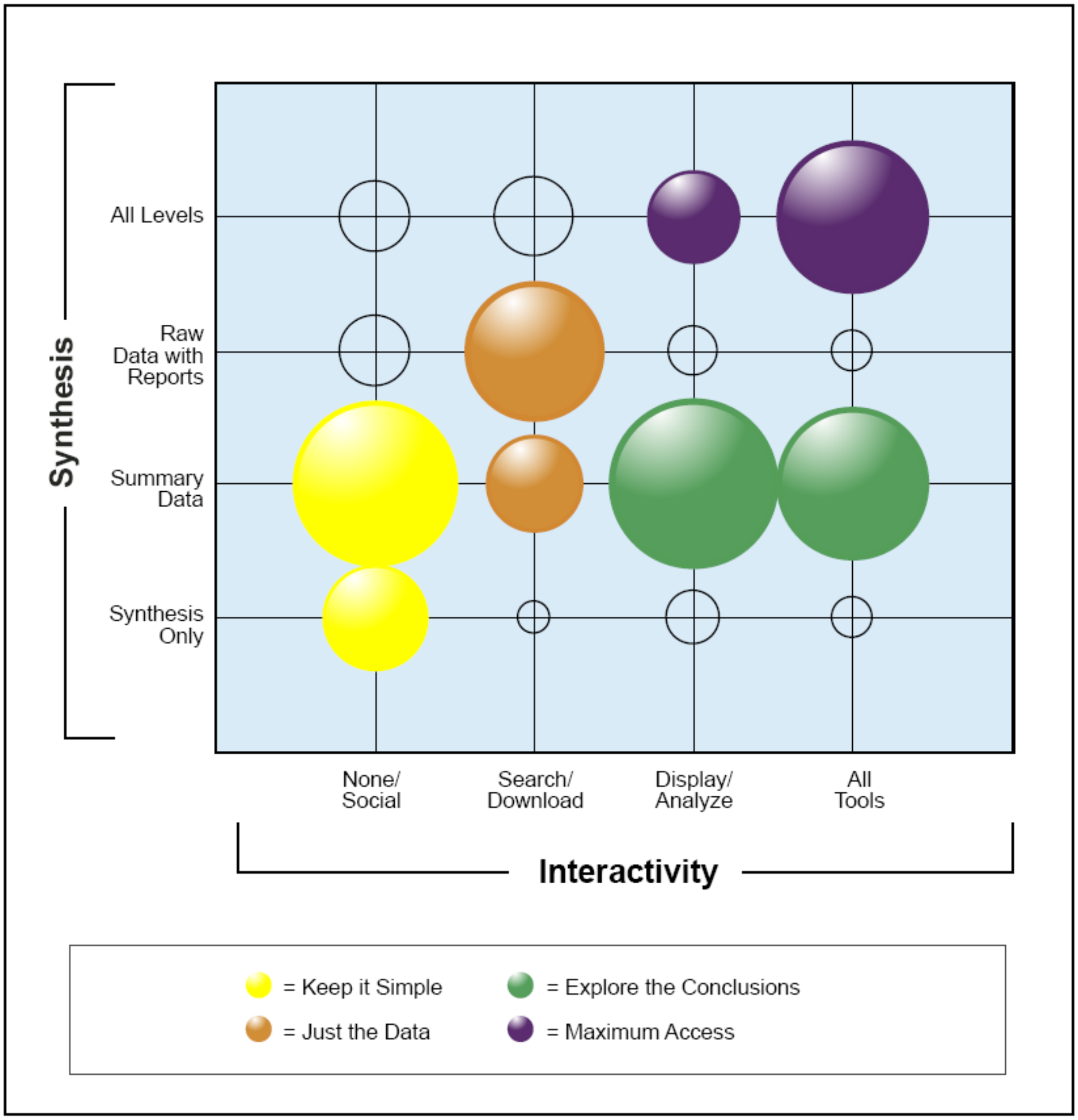


Figure 5-1. User persona assignments based on each of 16 potential combinations of the two indices. The relative size of each circle corresponds to the number of individuals (see also Table 5-2). Color-coding indicates four user personas categorized by similar attributes.

Table 5-2 describes the set of attributes for each color-coded user persona, and provides the number of respondents and interviewees that fell into each index and user persona.

Table 5-2. User Persona Names and Descriptions

Index Name	Index[1]	Description	Respondents	Interviewees	Total	Percent
Keep it Simple	1-1	Simple, cost-effective summary information, some interest in basic communication tools	24	2	26	5.9%
	1-2		65	2	67	15.1%
	1-3		13		13	
	1-4		13		13	
	2-1		3		3	
Just the Data	2-2	Searchable database, metadata, and download support important	25		25	5.6%
	2-3		41	5	46	10.4%
	2-4		17		17	
	3-1		8		8	
Explore the Conclusions	3-2	Mapping and graphing tools available for interacting with summary data and maps	61	6	67	15.1%
	3-3		7		7	
Maximum Access	3-4	All levels of data should be accessible, high interest in interactive tools	22		22	5.0%
	4-1		5		5	
Explore the Conclusions	4-2	Mapping and graphing tools available for interacting with summary data and maps	60		60	13.5%
	4-3		5		5	
Maximum Access	4-4	All levels of data should be accessible, high interest in interactive tools	59		59	13.3%
Total			**428**	**15**	**443**	

After the User Personas were finalized, each individual was identified as to both persona and archetype (Table 5-3). Although there was no obvious relationship between these two variables, some inferences are drawn based on the results presented in Table 5-3 and described below.

Table 5-3. Number of Respondents and Interviewees by Archetype in each User Persona

Archetype	Keep it Simple	Just the Data	Explore the Conclusions	Maximum Access	Other	Total
Recreational Consumptive	29	7	37	28	24	**125**
Scientist	7	27	22	11	15	**82**
ENGO[1]	13	5	17	9	3	**47**
Recreational Non-Consumptive	12	7	9	9	7	**44**
Commercial Fishing	2	3	3	8	6	**22**
Resource Manager	6	4	4	2	3	**19**
Educator	4	3	8	2	1	**18**
Information Manager	1	4	4	2	4	**15**
Policy	0	4	6	1	1	**12**
Decision Maker	1	1	2	1	0	**5**
Military	0	0	1	0	1	**2**
Tribal	0	1	1	0	0	**2**
Local Coastal Manager	5	1	1	1	0	**8**
Citizen	9	3	7	5	4	**28**
Student	1	1	1	2	1	**6**
Miscellaneous	3	0	4	0	1	**8**
Total	**93**	**71**	**127**	**81**	**71**	**443**

[1]Environmental Non-Governmental Organization

User Persona Definitions and Analysis

The discussion that follows builds-out each persona and describes their likely needs and preferences based on the survey data and enriched by the interview discussions. It also goes deeper into the evaluation of archetype distribution (Figure 5-2).

Keep it Simple (yellow) – The "Keep It Simple" persona category consisted of 89 respondents and 4 interviewees (20.8% of the population used to create personas; Table 5-2). This persona would advocate for a streamlined, cost-effective IMS that contains primarily static downloads of synthesis reports and key findings. They would not think it would be useful to include interactive data access and viewing tools, although basic communication functions (meetings, announcements, social networking tools) could be important. This group included people who wanted just highly synthesized data and/or reports as well as a larger proportion (71%, see Index 1-2 in Figure 5-1) who suggested that more detailed summary data tables, maps, and other pre-created information products would be useful.

There was a wide mix of archetypes in this persona category (Figure 5-2). The archetype with the highest percentage of individuals in the Keep it Simple persona was resource managers (37.8%, or 6 of 16; Figure 5-2), and local coastal managers (5 of 8, or 63% [not shown on Figure 5-2]). All but one of the managers fell in the 1-2 index (Figure 5-1), perhaps reflecting their interest in sufficiently detailed information (via tables and maps) to facilitate management decisions. Approximately a third of the recreational (32% non-consumptive and 29% consumptive) and ENGO-related (30%, including one telephone interviewee) stakeholders also fell in this category. No policy-related individuals and relatively few scientists fell in the Keep It Simple category.

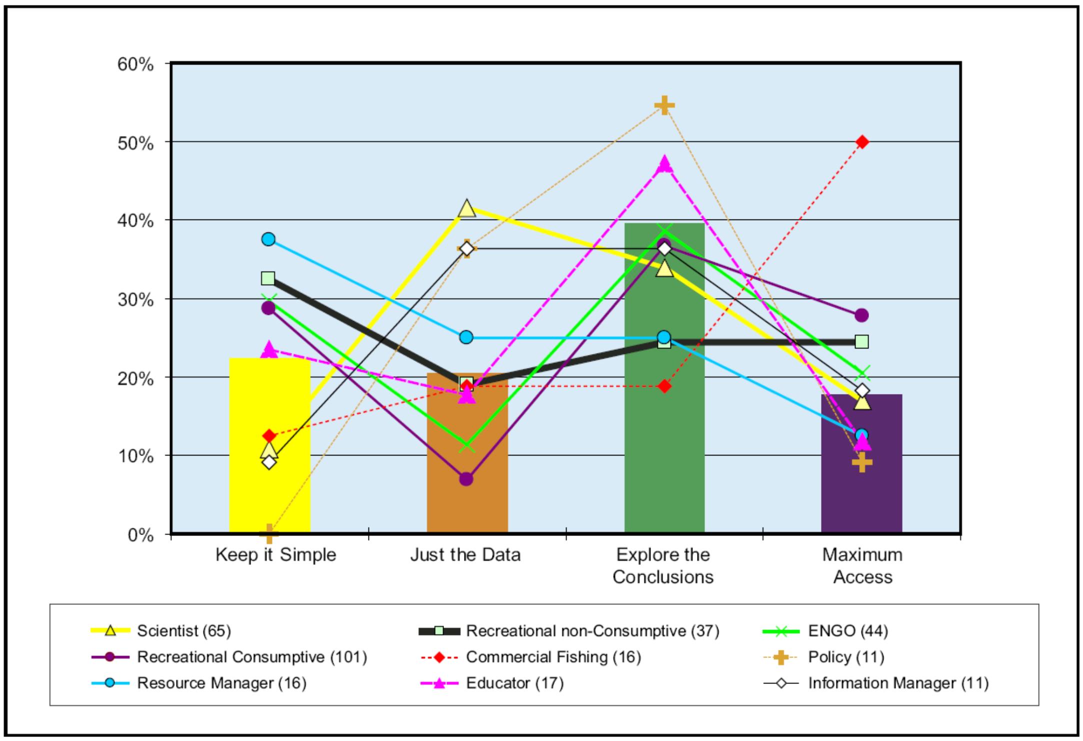


Figure 5-2. Archetype distribution of the four persona categories. Columns indicate the percent of each persona averaged across all archetypes; numbers in parentheses show the number of people (internet respondents and interviewees) for each archetype included in the persona analysis (442 total individuals).

Just the Data (light brown) – Sixty-six internet survey respondents and 5 telephone interviewees (15.4% of the population used to create personas; Table 5-2) fell into the "Just the Data" persona category. This persona is most interested in finding and getting access to raw data or databases, as well as summary data, and is less interested in highly synthesized information or online tools to interact with the data. The Just the Data persona commonly had specific opinions on data standards, use of metadata, and download formats for the data.

The three archetypes with the highest proportionate representation in this persona were scientists (27 of 65, or 42%), policy informers/influencers, and information managers (each 4 of 11 people, or 36%; Figure 5-2). This perhaps reflects the need of scientists and information managers for data to process using their own software, and the opinion of some policy-related individuals (all internet respondents) that providing data is a critical function of the IMS. Stakeholders (especially recreational consumptive [only 7 of 101] and ENGO [5 of 44]) were the most poorly represented archetypes in this persona.

Explore the Conclusions (green) – The "Explore the Conclusions" persona category was the largest in terms of the number of individuals, consisting of 121 survey respondents and 6 interviewees (28.3% of the population used to create personas; Table 5-2). This group combined those who wanted tools to view and analyze information with those who ranked all functionality high (including searching, download, social networking, etc.). The common element in this category was the interest in summary information, rather than raw data, detailed data tables, or databases. This persona is interested in a wide variety of online tools to view, search for, map, ask questions, and otherwise interact with the information.

The archetypes with the highest percentage of individuals in the Explore the Conclusions persona were policy informers/influencers (6 of 11, or 55%, including two interviewees), followed by educators (8 of 17, or 47%). Commercial fishing stakeholders had the lowest representation in this category (3 of 16, or 19%). Twenty-five percent or more of the people in each of the other archetypes fell within this persona.

Maximum Access (purple) – The "Maximum Access" persona category consisted of 81 internet survey respondents and no telephone interviewees (18.9%; Table 5-2). This person wants to ensure that all levels of data are accessible, and was the least discriminating in their choices of online tools for interacting with the data. This group was most likely to be concerned with transparency in the development of and access to monitoring information (Section 5.2).

Commercial fishing stakeholders were the most well-represented archetype in the category in terms of percent of individuals within an archetype (Figure 5-2), at 50% (8 of 16, no interviewees). After commercial, the two archetypes most frequently categorized as Maximum Access were recreational consumptive (28 of 101, or 27.5%) and non-consumptive (9 of 37, or 24.3%) stakeholders.

Implications of the User Persona Analysis

The costs for developing and managing an IMS that satisfies the needs of the four user persona groups increase from the lower left of the user persona matrix (Figure 5-1), or the "Keep it Simple" persona, to the upper right or "Maximum Access" persona. If a simple prioritization by numbers of users was used, then the IMS would be designed for the "Explore the Conclusions" persona; however, this approach is overly simplistic. There is an important role for considering archetypes; for example, many of the respondents and interviewees cited decision makers as a critical audience for the IMS, yet this group was underrepresented in the UNA results.

The user persona results could be used to design the IMS using a tiered, modular approach as suggested by six of the telephone interviewees and some comments in the internet survey responses. In this way, specific functionality could be narrowly defined and implemented for the likely users at different points along the path of the program development. For example, in the initial phases of monitoring data collection, the site could be designed for access by the "Just the Data" persona, as well as providing summary "what's going on" information on the front page.

Although the "Maximum Access" persona would appear to be most problematic as it is associated with the highest cost, certain themes emerged during the UNA (described in Section 5.2 below), especially ones related to this persona, that should be considered in the website design. For example, it appears that the desire for open access to raw data and flexibility in site interactivity was partially driven by the desire for accessibility and transparency of the monitoring and adaptive management processes.

5.2 User Need Themes

Internet survey respondents and telephone interviewees consistently raised certain themes that, although not directly related to an IMS framework element, should inform decisions in the next phases of the IMS development. This section synthesizes the themes that arose in the results reported in Sections 3 (internet) and 4 (telephone interviews). It also summarizes open comments from both the internet survey respondents and from the telephone interviewees. The last two questions of the internet survey (Appendix A) included two opportunities for open comments:

- Now that you have taken the survey, do you have any other suggestions or "out-of-the-box" ideas that have not been addressed for the MPA Monitoring Enterprise information management system? List up to three.
- What do you think are the top three most critical aspects that should, beyond all other possible information sources, utilities or functions, be implemented in the MPA Monitoring Enterprise information management system?

Similarly, the final question asked of each telephone interviewee was to list the top three considerations that they thought were of greatest importance in developing the IMS. Below we summarize the IMS thematic issues that arose most commonly and provide recommendations for addressing these issues in the IMS development. Then we highlight several interesting "out-of-the-box" ideas that might merit consideration in the IMS design

Transparency /Open Access

Although synthesized information was cited as critical by the majority of internet survey respondents and telephone interviewees, the theme of "transparency" was directly or indirectly raised by people who requested the availability of raw data. Twelve of 35 telephone interviewees mentioned the need for transparency; three comments in the internet survey responses spoke directly to this theme (e.g., "*Useful tools to satisfy customers that the system is truthful and transparent*."). The "Maximum Access" persona, which includes the highest contingent of commercial fishing stakeholders (in terms of percent per archetype), would argue that only through access to unfiltered raw data and a "*fully transparent record of methods, practitioners, dates and sampling plans*" (policy interviewee) would stakeholders be assured of the legitimacy of program results.

One solution might be for the IMS to ensure that all information, regardless of synthesis level, is provided with full documentation in plain, jargon-free language, placing the data in context with explanatory background information and/or uncertainty analyses. Metadata could be used for this purpose, but some standards are technical, not meant for a lay audience (as opposed to scientists and information managers), and would therefore be an insufficient solution. However, some technical experts may argue that explaining the scientific process in a simple, understandable way for the layperson is untenable.

At a minimum, providing background information on the "who, how, and where" of data collection could go far to address this theme. Another potential option is to provide access to raw data files by request, as suggested by one interviewee. This could be problematic due to proprietary, sensitive, or confidential data. This presents a clash of user needs that should be addressed head-on in the IMS design process, and/or through institutional agreements between the data providers and the Monitoring Enterprise.

Bias/Credibility

The related themes of bias and credibility were raised in both the internet survey comments as well as in the telephone interviews. Specifically, at least ten comments provided in response to the final internet survey section raised the point that the monitoring information needs to be credible and/or peer reviewed (e.g., "*There needs to be a way to assess the validity of the data and any assumptions made in its collection*."). Credibility of the information was directly raised by seven of the telephone interviewees.

One telephone interviewee drew a clear distinction between bias (related to spatial resolution, periodicity, patch size of sampling) and conflict of interest of the collector and interpreter. In the final comments in the internet survey, 15 respondents addressed the critical need for accurate, validated data (e.g., "*Above all, ensure the data in the system are of high quality and carefully checked before entering into the system*."). For assessing conflict of interest, the data might be included but flagged with the interest of the collector.

As with transparency, the issue of bias and/or credibility is most easily addressed with documentation of source, methods, and analyses in language readily and openly translated to the layman. Documentation of methods, quality assurance/quality control results, and peer review history could be included with the "metadata" record. This documentation should be formatted such that it shows, as suggested by one interviewee, that the information is legitimate (i.e., not biased in selection or presentation), with a clear record of the provenance (who collected it), so that the source can be weighed in the analyses and interpretation.

Misuse of Data/Uncertainty

The issue of open access to raw data commonly raised the theme of concern about the potential misuse of monitoring information. Fourteen of 35 telephone interviewees mentioned the concern that people may misuse or misrepresent the monitoring data, with most (12) acknowledging that misuse could or would happen, but that transparency was more important factor in making decisions about limiting user capabilities or data distribution. Two telephone interviewees (both scientists), in contrast, felt that it was important to place some limitation on user capabilities to avoid data misuse. They suggested restricting access to raw data entirely, or creating a distribution process that would require data requestors to speak to someone who can provide them with guidance and information about data caveats.

One telephone interviewee succinctly described the misuse problem as either a lack of awareness of how to interpret uncertainty in scientific results, or an incomplete understanding of the context of results. That individual saw a risk that highly synthesized reports might be overly simplified without acknowledging potential error or uncertainty in the results. There was also the perception that raw data could be easily misinterpreted (ascribing trends, causality, or covariance inappropriately).

Eleven telephone interviewees suggested that one solution for limiting data misuse is for error or uncertainty estimates to accompany the synthesized information, taking the form of error bars, disclaimers, or other context information. In addition, providing contextual information, rather than just the final synthesized results, was suggested by this group. Widely varying views were expressed as to what background material would be appropriate, ranging from simple error indicators to basic marine science educational material. One interviewee advised using the Heinz Center Reports ("State of Nation's Ecosystems") as a model of synthesis with clear pointers to data and qualifiers regarding the data quality and what it shows.

Timeliness of Information

At least six internet survey respondents made comments about the importance of the timeliness with which monitoring information would become available, expressing the concern that the scientific process would interfere with this need. Although the time to produce monitoring information is not dependent on IMS design, designing data and information flow paths to aid in monitoring information processing could be a useful function to speed delivery of information to users. Automated processing and posting of interim results is another potential IMS feature that could address the timeliness issue, although we are unable to accurately assess the viability of this option without more specific information about the content or processing needs of the monitoring data that ultimately will be collected.

Three telephone interviewees expressed concern that expectations for rapid posting of results to the IMS would be difficult given how much time it will take for monitoring to detect measurable change within the MPA network. In fact, five internet comments specifically expressed the need for near-term collection of baseline information to compare against subsequent monitoring results. Interim information updates that document what information is being collected, and for what purposes, is one content solution to address both the need for timeliness, as well as the concern to avoid premature conclusions.

Cost

Ten of the 35 telephone interviewees mentioned cost as a concern. The "Keep it Simple" user persona included those who felt that the website should maintain focus on the core, necessary functions, or even the bare minimum due to cost limitations, or the desire to avoid negative public perception about extraneous spending. At least eight internet survey respondents made direct comments about either the cost-effectiveness of the IMS ("*Don't waste resources going overboard on this IMS - too complex will fail, and the resources are needed for the monitoring itself*"), or transparency ("*Where will the funding come from*?").

The solution to this perception issue is to evaluate the IMS options using a cost/benefit approach. For example, the cost for serving static monitoring reports is relatively low (serving the needs of the "Keep It Simple" persona); but the cost of a poorly designed site that prohibits users due to difficulty in finding these reports has a higher (but less explicit) cost. Similarly, the cost of managing information that serves the needs of the "Just the Data" persona could be significant if a strict standardized database is created and needs to be managed; however, the cost of individuals responsible for synthesizing this information having to do their own data management is possibly within the same potential range of cost. A detailed cost/benefit analysis of the IMS options as described by the User Personas is outside of the scope of the UNA, but the results could be used to generate and compare various approaches' direct and indirect costs.

Summary Ideas

Here we report some of the more interesting and relevant "out-of-the-box" ideas from both the internet survey respondents and telephone interviewees.

- *If the system is put in place and running well, perhaps it can hold other similar data from coastal California.*
- *Experience and observation blog by/for MPA visitors.*
- *Short audio/video clips in different languages explaining simple "rules" for public understanding, perhaps by location.*
- *Short stories showing measurable successes to instill confidence in public as you go.*

- *Use it in part to identify data needs - maybe a table with MPAs on the vertical axis and data types (WQ, fish, habitat...) on the horizontal, and various symbols for the cells indicating 'robust data, some data, monitoring underway, and no data. The cells could be live links to the actual data available.*
- *Why not include a section for local involvement/volunteers community activities?*
- *Involvement of local people and infrastructure. The involvement of local expertise and human resources is essential to meeting all of the goals and objectives of the MLPA.*
- *Get info about the impending release of the MPA Monitoring Enterprise IMS into social media (when time comes).*
- *It would be cool from a consumer point of view to be able to look at an MPA sign on the coast and see a URL on the sign and put it in your PDA and get a dashboard of results for that MPA. Resource managers are slow to take on this sort of thing, but there is a podcast narration for travel on Highway 37 up the coast north of SF. On the ferry to the Channel Islands there is an LCD screen displaying progress across a multibeam image of the seafloor. As devices on the belt become more sophisticated, the delivery of public information needs to keep pace.*

5.3 Summary Findings and Implications for the IMS Framework Elements

This final section of the UNA summarizes the findings for each survey topic (Table 1-1) integrated across the internet survey and telephone interview results and, where appropriate, in relation to the user personas. It explicitly links these results to the IMS framework elements (Table 1-2) and provides related recommendations in the summary matrix tables that accompany each subsection below.

Content

The core content that the website must have, cited by both those who were interviewed by telephone and those who participated in the internet survey, was information related to whether the MPAs are effective and/or are achieving the goals set out for them. Internet survey respondents indicated strong interest in all of the monitoring data types, and telephone interviewees agreed with the need for specific metrics including ecosystem and species indicators. Other types of information were considered more commonly as very useful or nice to have rather than essential, including human use data (although stakeholders had a strong interest in this type of information) and educational resources. Although internet respondents ranked educational resources most often as nice to have, several interviewees suggested that an educational element was invaluable for placing the monitoring data in context because of data complexity; however, other interviewees thought that education was not the primary function of the Monitoring Enterprise.

While the actual content of the monitoring information will be decided in a separate monitoring planning process, the results provide some clear guidance regarding the IMS database framework element on content (Table 5-4). The actual content will largely dictate specific data management decisions (data organization, etc.). Nevertheless, the survey findings strongly suggest that whatever information is included, it must be tagged, documented, linked, or otherwise interlocked with "context" that is understandable to lay audiences. This context could take many forms as suggested by those surveyed: variability (spatial, temporal) and error; paired objectives and findings (i.e., objectives of an individual MPA); or a primer that explains the information being reported. Using metadata (that standardizes much of the contextual information) would fulfill this need only if it can qualify and explain data in a jargon-free manner. Finally, in order to address the users' perceptions of what information content is essential, this associated explanatory information should explain why certain monitoring data were collected (or excluded).

Table 5-4. UNA Findings and Implications for IMS Framework Elements: Information Content

SURVEY TOPIC	Content
TOPIC DESCRIPTION	Preferences for types of monitoring information; user needs for other related information.
INTERNET SURVEY RESULTS SUMMARY	More than 50% of users indicated that all the presented monitoring information content (ecosystem, species, and human use) was essential or very useful, with some specific choices (especially fish) considered essential by a majority of users.
TELEPHONE INTERVIEW RESULTS SUMMARY	Summary of "how MPAs are doing" is critical; results of monitoring metrics in context with MPA goals is important; include details on current monitoring activities; content should be phased based on the schedule of monitoring and reporting.
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Database
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	The data content will impact how complex the database has to be; in addition, different formats of information may require specialized software or other components. The results of the survey, however, suggest that the more important element is providing the context of the information (documentation and/or metadata) that explicitly defines what data are being collected and why.

Synthesis

The level of information synthesis provided by the IMS has a direct impact on IMS design decisions, therefore the results and the implications of these results are presented and discussed based on four levels of synthesis based on the user persona analyses. Highly synthesized information summaries and reports have been grouped together because they require a similar level of intensive data processing, although this analysis will most likely occur outside of the IMS with the results then linked to the site.

Synthesis Level 1 – Consistent with the expressed need for summary findings on whether MPAs are effective, the most commonly cited need was access to clear and concise syntheses of monitoring results (Table 5-5). It is most closely associated with the "Keep it Simple" user persona. The suggested format of this summary varied. Telephone interviewees commonly cited a desire for a single, clear and graphic status indicator variously described as a report card, red light/green light, dashboard, snapshot, yes/no, and vital signs. Both telephone interviewees and internet survey respondents voiced the need for program reports, separate from scientific publications.

Two primary products are associated with this synthesis level – highly synthesized programmatic results or indicators, and published scientific or programmatic reports. Unless a method to automatically calculate summary metrics is desired, the development of these highly synthesized final information products will be primarily outside of the IMS environment. The design of the user interface, however, is key to enable the users to find the information quickly, and understand the meaning and context of the associated results. Therefore, this synthesis level is primarily a function of interface design, requiring services of a creative, intuitive website designer. Useful modules that could benefit the user include bibliographic search capability for scientific and other programmatic reports; additionally this capability would require programming skills.

Table 5-5. Survey Topic Findings and Implications: Synthesis Level 1

SURVEY TOPIC	Synthesis Level 1
TOPIC DESCRIPTION	Information summaries, reports for non-specialists, key findings, or scientific publications of monitoring results.
INTERNET SURVEY RESULTS SUMMARY	Nearly 80% of respondents said that this level of synthesis was essential; resource managers were most interested in synthesized findings, while scientists, policy-related, and ENGO stakeholders were most interested in scientific reports.
TELEPHONE INTERVIEW RESULTS SUMMARY	Information summaries were considered of highest importance, but should be accompanied with some indicator of error. Regular monitoring reports (annual or lower frequency) are also important.
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Interface; Secondarily Database and/or Modules
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	Accessible presentation of intuitive information synthesis graphics or reports is an important design element, requiring attention and resources paid to the website interface. Simple linkage of reports is straightforward, although reports or other documents could be stored in a "library" database, with metadata and/or keyword functionality, requiring a bibliographic module of the database. Automating creation of summary graphics from the database would require programming, but this process is problematic both from an IMS as well as a scientific standpoint. Overall, the presence of this level of synthesized information has a small resource impact on the IMS once the design is finalized, and is of critical importance.

Synthesis Level 2 – Most users were interested in some level of summary data products, as this synthesis level is associated with all but the "Just the Data' persona. A wide range of possible mid-level summary "products" could either be created outside of the IMS (such as annual report graphics), or through a programming module that generates summary products by request and displays them on a map or chart, requiring access to a managed, standardized database (Table 5-6). This synthesis level, therefore, requires potential software and/or applications to create information views on demand. The findings and implications presented here assume some level of interactivity, since fixed synthesis products (static) would fall in the Synthesis Level 1 category in terms of the IMS framework.

Internet survey respondents ranked mid-level summary data products as just as important as highly synthesized information "snippets" and reports. Individuals classified as resource managers were most interested in information summaries and summary tables (17 of 18 ranking it as essential). Few telephone interviewees, in contrast, were advocates of this level of information, explaining that monitoring reports (that contain this level of synthesis graphics, tables, and maps) would suffice, although accessing information via a map interface was cited as useful by 16 of those interviewed by telephone (Table 5-6).

Both the internet survey respondents and telephone interviewees expressed a strong desire to view MPA monitoring information via a map interface, as well as to 'drill down' to find more specific results (for example, local MPAs). The IMS map options are too numerous to provide summary recommendations, although they range from free, open source solutions to high-end commercial software. All levels require skilled information managers educated in managing and providing access to spatial data. The other variable in determining the level of resources required to manage information access via a map interface is the level of detail of the information behind the map, and whether the information comes from an actively managed database, or if the maps draw from a fixed summary data set catered to map viewing (a hybrid approach). The first option maximizes transparency, while the second is more likely to provide improved clarity of information for the layperson.

Table 5-6. Survey Topic Findings and Implications: Synthesis Level 2

SURVEY TOPIC	Synthesis Level 2
TOPIC DESCRIPTION	Summary information products including tables, graphs, maps, photographs, and video.
INTERNET SURVEY RESULTS SUMMARY	Maps were of highest interest across all archetypes.
TELEPHONE INTERVIEW RESULTS SUMMARY	Few interviewees suggested that summary information (outside of regularly published programmatic reports) was important, with the exception of maps to be used to find results of specific interest.
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Components; Secondarily Interface, Architecture, Database
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	Presentation of results on maps will require selection of components used to make and serve the maps; these options have a high range of resource requirements depending on the interactivity of the maps. As with reports, creation of any summary data products outside of the IMS and then linking them to the site is a straightforward process with minimal resources required.

Synthesis Level 3 – Offering standardized data tables, statistical summaries, and/or databases that provide the data supporting summary studies and metrics would require storing the data in a consistent, normalized database (or related databases served at distributed nodes) to facilitate querying and statistical calculations (Table 5-7). This option, most closely associated with the “Just the Data” user persona, would require a relatively significant level of effort to manage the information to ensure quality and consistency of the information. This level was designed to address data management needs, rather than processing needs required for creating summary tables for Synthesis Level 2. Although the implications of the apparently subtle difference between summary and detailed data tables may not have been clearly understood by all those who were surveyed, many of those that did respond to the need for this level of data, and the follow-on questions, were clear about their need for this level of detailed data.

Table 5-7. Survey Topic Findings and Implications: Synthesis Level 3

SURVEY TOPIC	Synthesis Level 3
TOPIC DESCRIPTION	Detailed tables of data, statistical summaries, compiled databases.
INTERNET SURVEY RESULTS SUMMARY	Less than half (49%) of respondents indicated an interest in this level of detailed data. Resource managers were most interested in summary statistics, while information managers were most interested in databases.
TELEPHONE INTERVIEW RESULTS SUMMARY	Detailed data/statistics were not specifically addressed, although database management issues were discussed by several interviewees. Most critically, data and metadata standards were considered vital, preferably using existing standards rather than creating new ones. There was disagreement about whether a centralized or distributed system would be more effective, but the ability to query a normalized database was cited as a basic, critical element to a subset of the interviewees.
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Architecture, Database; Secondarily Components
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	The need for management of a standardized (if not centralized) database is a critical decision for the IMS, as it not only requires resources up front to design and implement, but long-term commitment to manage the data. The results are conflicting about the need for a centralized database (or managed data over multiple, distributed locations), but at a minimum, selection and application of data and metadata standards will meet many needs expressed for organized, standardized, and well-documented data. Once a decision is made about the level of data management required, the follow-up decisions will require more detailed assessment of existing database and metadata infrastructures that might be candidates for the Monitoring Enterprise to adopt.

Almost half of the internet respondents (48.7%) indicated their interest in this level of information (Table 5-7); resource managers were most interested in statistical summaries, and information managers were most interested in databases (6 of 9 individuals each). Use of databases by information managers fits with the assumed preferences of this archetype; the expressed need for more detailed data and statistical summaries by resource managers (both internet survey respondents and telephone interviewees) suggested that more detailed data could be helpful in understanding the variability of the resources they are managing.

The case for a managed standardized database was made by one resource manager: "*The data should be centralized with some controls on formatting and structures… Data should get compiled so don't have to do a lot of back-door data management*."). The advantage of standardized data is that analysis is not hampered by mixing and matching fields (chemical or biological names, etc.), sample design elements (spatial compositing, replicates, etc.), and units (coordinate systems, units of measure, etc.). Another approach is for the Monitoring Enterprise to serve as a "clearinghouse" that will point to distributed sources of information, but this architecture makes it more difficult to enforce data standardization.

In the final analysis, the need for a standardized database to facilitate analysis may be best decided by those charged with evaluating the data for the purposes of the Monitoring Enterprise. If there will be a need to do standardized, regular analyses with a subset of monitoring data (for example, to come up with the metric on whether the MPAs are effective), greater efficiency might result from centralizing the storage of at least part of the data required to conduct this analysis. Alternatively, the standards developed and required at each node must be rigorously followed so that the process of data compilation across multiple sources does not become too arduous and time-consuming.

Synthesis Level 4 – As discussed in Section 5.2, the expressed need for raw data, in part, reflects a desire for transparency rather than a functional need. This synthesis level is most closely associated with the "Maximum Access" persona. Only 27.4% of the internet survey respondents expressed interest in raw data, with the two most frequently represented archetypes being scientists (33 of 79 or 41.8%) and commercial fishing stakeholders (12 of 30 or 40%). Many respondents who expressed interest in accessing raw data actually had limited experience in downloading data, but still felt it should be available (Table 5-8). One solution may be to provide access to raw data sets through other means by request, as suggested by one telephone interviewee (e.g., CDs). The findings of this survey topic confirmed the importance for metadata and documentation that describes the context of the raw data.

Table 5-8. Survey Topic Findings and Implications: Synthesis Level 4

SURVEY TOPIC	Synthesis Level 4
TOPIC DESCRIPTION	Raw data; questions were also asked about associated download formats and metadata.
INTERNET SURVEY RESULTS SUMMARY	About 1/4 (27%) of the respondents indicated raw data were important, but 15-25% of this group were inexperienced in data download methods and formats. Those that suggested a need for raw data most commonly requested data in a simple format (spreadsheet, text file), drawn from a standardized database, and accompanied by a metadata file describing the content of that download.
TELEPHONE INTERVIEW RESULTS SUMMARY	Many interviewees thought raw data essential, although a minority expressed the opinion against serving raw data. Because of the more limited audience (scientists and the need for transparency in the process), some suggested alternative, more cost-effective methods of delivering raw data by request (i.e., via CD).
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Components; Secondarily Database
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	Because having raw data available was considered less of an important need overall, meeting those needs could readily be achieved through external delivery (CD or ftp site) upon request. The issues of proprietary data, mis-use of data, or other procedural aspects do not directly impact the development of the IMS. As with databases, standardized metadata is another critical element to document the source and provenance of the raw data.

Interactivity

Website interactivity also has multiple implications for the-IMS framework elements, thus results are presented and discussed based on four levels of interactivity based on the User Persona analyses.

Interactivity Level 1 – This level is best represented by those in the "Keep it Simple" user persona (Figure 5-1). Although 40% of internet survey respondents indicated that they thought online tools were unnecessary, approximately a third of these same people also indicated they were interested in specific kinds of online tools (Table 5-9), especially the ability to find information via a map, or searching for a report. Some also commented on the usefulness of such tools as Google Earth or MarineMap. Nevertheless, resource manager internet survey respondents, as well as many of the telephone interviewees (9 of 22), suggested that the site should be very simple and streamlined, suggesting that complex tools would be perceived as a waste of limited resources.

Table 5-9. Survey Topic Findings and Implications: Interactivity Level 1

SURVEY TOPIC	Interactivity Level 1
TOPIC DESCRIPTION	No tools, or limited tools (social networking, for example).
INTERNET SURVEY RESULTS SUMMARY	Forty percent of respondents indicated that online tools were unnecessary, although some of these also selected tools of interest. The archetype selecting this option most commonly was resource managers (10 of 15). Response to social networking tools was dominated by the ranks of nice to have and not needed.
TELEPHONE INTERVIEW RESULTS SUMMARY	Nine of 22 telephone interviewees who commented on tools suggested that a static site offering report and summary information links was sufficient, and the most cost-effective approach. Six of ten interviewees who mentioned social networking tools said they were unnecessary or inappropriate; four suggested that tools that reached out to youth should be a priority.
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Interface, Architecture
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	Considering that 40% of both the internet respondents and interviewees suggested that a static website with fixed information links might be sufficient, this IMS architecture should be seriously considered. This model of a website would likely be the most cost-effective approach, with the main effort geared towards intuitive, effective site design (interface). This model would require that users with need for data access be served through other means (ftp sites; emailed files, etc.). The response to social networking tools likely reflected the experience and backgrounds of the surveyed respondents and interviewees.

The seemingly contradictory views of some internet survey respondents and telephone interviewees are probably a function of what individuals considered a 'tool.' This interactivity level is most likely a surrogate for cost-effectiveness, rather than concerns that tools are unnecessary or undesirable. Still, the results indicated that the model of a static site (e.g., no database linkage, none or only few applications) is one option that should be considered. The main need that would not be met with this model is searchable, online access to monitoring data or interactive mapping. The social networking options had high rankings of both nice to have and not needed. This bimodal response appeared to reflect, at least in part, the users' experiences with these kinds of newer communication methods.

Interactivity Level 2 – This level is best represented by those in the "Just the Data" user persona (Figure 5-1). Only 20.5% of the internet survey respondents selected the option that would suggest they are primarily interested in downloading data to do their own analyses. The percent of scientists and information managers selecting this option ranged from 30-40% of each archetype (Table 5-10). The code necessary to allow downloading of data is straightforward. However, designing queries that allow searching and filtering data is of critical importance, because a poorly designed data query interface is one of the more common complaints of data access websites (as the website reviews indicated). This capability also requires an element of database management, so that the database design supports easy construction of logical queries for subsets of data. The resources necessary to accomplish this level of interactivity include: intuitive, effective site design; an effective data download module; and a logical database design that facilitates querying.

Table 5-10. Survey Topic Findings and Implications: Interactivity Level 2

SURVEY TOPIC	Interactivity Level 2
TOPIC DESCRIPTION	Basic search tool, query and download reports and/or datasets.
INTERNET SURVEY RESULTS SUMMARY	Many scientists/information managers chose this option, but overall only 20.5% of respondents cited that they would need to download data to do their own analyses.
TELEPHONE INTERVIEW RESULTS SUMMARY	Eight of 22 interviewees who commented on tools suggested downloading of data was important, but two were concerned about allowing access to raw data.
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Modules, Secondarily Interface, Database
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	The results suggested that the primary audiences interested in downloading data (independent of those citing that raw data should be accessible) were scientists and information managers. The code module necessary for downloading is straightforward; more important is an effective query interface to ensure that users know what data they are receiving, as well as an organized database structure that facilitates efficient querying.

Interactivity Level 3 – This level is best represented by those in the "Explore the Conclusions" user persona (Figure 5-1). The internet survey response rate for those interested in data exploration and analysis tools ranged from 52-55%. The tool type of greatest interest was a map overlay capability, followed by keyword searches (Table 5-11). Of the telephone interviewees who commented on basic tools (22), half thought visualization tools were important. These interviewee opinions contrasted with those recommending only the simple static model of the website (Interactivity Level 1); they felt that if users cannot access the data easily, the entire purpose of collecting these data would be defeated.

The consistency of the feedback between the internet survey respondents and telephone interviewees suggests that the most commonly cited tools (maps and keyword search) should be among the first functions developed for the IMS, assuming resources are available. Mapping options are discussed above (Synthesis Level 2). Keyword search capability is a basic website analytical tool, and can be easily catered to the IMS (assuming key words are integrated with the database design). One criterion for choosing among tools might be to only include targeted, cost-effective tools and utilities that support expressed needs for data interactivity and that are so effective that their cost is well justified.

Table 5-11. Survey Topic Findings and Implications: Interactivity Level 3

SURVEY TOPIC	Interactivity Level 3
TOPIC DESCRIPTION	Visualization and display tools including mapping of results, generating charts or tables in real-time through on-line queries.
INTERNET SURVEY RESULTS SUMMARY	The overall interest in browsing and data exploration tools ranged from 52-55% of internet respondents. A map overlay capability was of highest interest (59.3% selected essential), followed by keyword search (54.3%). More than 50% of most stakeholders (especially commercial fishing) indicated that rapid generation of charts and tables by user request would be essential.
TELEPHONE INTERVIEW RESULTS SUMMARY	Half of the 22 interviewees who commented on tools thought basic visualization tools were important. Viewing and selecting data from a map interface was the most consistent mentioned idea, and secondly, query/search tools.
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Modules, Secondarily Interface, Database
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	The consistency of the information between the internet survey and telephone respondents suggests that the most commonly cited tools (maps and keyword search) should be among the first priority functions, assuming resources are available. Both of these functions are also useful for data query and download (Interactivity Level 2).

Interactivity Level 4 – The internet survey responses for high-end data analysis or visualization tools, or specific tools such as decision support tools, showed that many respondents were unfamiliar with these tools (27-28%) and many others rated them as nice to have but not essential (Table 5-12). Both internet survey respondents and telephone interviewees expressed concern that overly complex tools would indicate poor use of limited resources. Tools for targeted purposes, however, had some support, especially among those familiar with MarineMap, as well as some respondents/interviewees from the CDFG who anticipated a future need for tools to support adaptive management decisions regarding the MPAs. One interviewee suggested a model in which tools could be developed for specific clients that funded the tools' development. As more of these kinds of data visualization tools become available, if used appropriately, they can be effective in conveying information to diverse users in an intuitive way. As stated by one internet respondent: "*these kinds of tools will take thought and time to develop into meaningful, simple ways of displaying data, but they can be very powerful visualizations and modern technology can be fairly easily used for this now*."

Table 5-12. Survey Topic Findings and Implications: Interactivity Level 4

SURVEY TOPIC	Interactivity Level 4
TOPIC DESCRIPTION	Detailed analysis tools or decision support tools that require accessing many different layers and/or types of information to reach a specific goal.
INTERNET SURVEY RESULTS SUMMARY	Many of the respondents (27-28%) were unfamiliar with the most complex interactive tools (e.g., 3D motion tools, etc.), and few (11%) indicated they were essential tools. Comments for more complex tools ranged from "avoid" to "good place for innovation."
TELEPHONE INTERVIEW RESULTS SUMMARY	All but one interviewee suggested complex analysis tools were not necessary or of low priority; one interviewee suggested a model, however, of developing tools on an as-needed, targeted basis for specific users.
IMS FRAMEWORK ELEMENT LINKAGES	Modules
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	Although not a highly rated priority, there may be room for targeted, well-designed tools developed for a specific purpose and/or specific client.

Human-Computer Interface

A well designed website interface will be essential for successful use of the IMS (Table 5-13). Thus one of the most critical decisions will be the choice of internet designer, platform, and process by which this will be accomplished. Findings from both the internet and telephone surveys suggested a bimodal opinion about the site being housed within a state agency. From an interface perspective, the reviews of many suggest that state websites are non-intuitive; however, one advantage of the state web design process is the availability of protocols for ensuring access for all users.

Feedback from both internet survey respondents and telephone interviewees suggested that the biggest pitfalls noted in users' experiences with other websites were sites that were slow to retrieve information, and sites that made finding and downloading data difficult. This result demonstrates the importance of both good site design and efficient programming.

There were two overall models for the site design raised in the interviews (Table 5-13): a) static site with monitoring results only, updated when new reports are available; and b) a multi-faceted, phased site, starting with some basic information on what monitoring is happening, then building up to more complex data synthesis and mapping results reporting.

A phased approach has many benefits, not least of which is that data, functions, and other IMS aspects can be prioritized and addressed as the monitoring programs proceed. At the simplest, information that informs the public about "what's going on" would be a relatively cost-effective starting point. More detailed information/tools can be developed as the database and analysis methods are developed. A site that is updated and avoids stagnant content more likely will be visited.

Table 5-13. Survey Topic Findings and Implications for IMS Framework Elements: Human-Computer Interface

SURVEY TOPIC	Human-Computer Interface
TOPIC DESCRIPTION	User experience and preferences, how they interact with the internet, ease of use of a site, including physical or cultural barriers.
INTERNET SURVEY RESULTS SUMMARY	Users need efficient (rapid) retrieval and display of information, with sufficient documentation to understand the data they are requesting. Feedback included that the site should support alternative languages (27%) and people with disabilities (39%). Serving users with mobile phones was less of an indicated need.
TELEPHONE INTERVIEW RESULTS SUMMARY	Two options for site design were advocated: limited tools with clear access to monitoring reports, and a multi-faceted, phased approach for building a database-driven site. Interviewees highlighted the importance of providing information summaries without the over-use of jargon. Websites that serve data were cited as frustrating if the data cannot easily be found, downloaded, or documented. Opinions about serving users via mobile devices were consistent with opinions about social networking tools (limited, but some advocating important for youth outreach).
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Interface, Modules
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	Interviewee results suggested that the IMS could take one of two approaches for site design: limited (findings and reports only), or multi-faceted, with a possible phased approach. The two most critical design issues affecting user experience were ease of finding data, and processing speed. If the system allows querying and download of data, there should be sufficient resources to ensure logical organization for searching for data, and sufficient documentation to describe exactly what data are being retrieved. Including options for alternative languages (other than for published reports) or access for those with disabilities also will add to the overall cost of the IMS, but may be necessary to ensure accessibility of the information.

Institutional

A key issue, and one that the results of the UNA cannot resolve, is the decision as to whether the website will be part of an existing infrastructure (state agency [CDFG or State Water Board] or University), an independent entity, or an integrated hybrid (Table 5-14). Feedback from the internet survey respondents and telephone interviewees is mixed on the advantages and disadvantages of housing the system with the state. Many voiced strong feelings that the site should not be hosted and/or managed by the state; even some state agency respondents suggested that the level of rules for site management would preclude the development of an attractive, well-visited site. There is also the issue of trust, as one policy-related interviewee commented: "*Nobody trusts the government, this should NOT be a government-sponsored or* *hosted site, would definitely be a failure. Central issue is trust – most agencies are circumspect, so an external and independent site is the best approach."*

The long-term stewardship of the data, however, is a credible and important issue, and an argument can be made that the state is a good choice for long-term residence of the data. In addition, compatibility with some state standards (database, metadata) might be part of a good hybrid solution. In addition to the state, various organizations exist that are potential partners for data sharing solutions. The architectural solution for the IMS framework is outside of the scope of this User Needs Assessment, but the information collected here should be of use to initiate partnering discussions.

Table 5-14. Survey Topic Findings and Implications for IMS Framework Elements: Institutional

SURVEY TOPIC	Institutional
TOPIC DESCRIPTION	Organizational issues that will impact how the website will be run and maintained, including partnerships, institutional agreements for data sharing.
INTERNET SURVEY RESULTS SUMMARY	Almost half of the respondents (46%) indicated interest in providing monitoring information to the IMS, split among MPA monitoring data, associated information, and consumptive uses, suggesting that data upload functionality (templates, etc.) would be necessary. Other respondents replied that their agencies would be willing to serve as links or nodes to the IMS.
TELEPHONE INTERVIEW RESULTS SUMMARY	Interviewees provided mixed opinions about the advantages and disadvantages of housing the IMS at a state agency, university, or independent location. Websites supported by state agencies were perceived by some interviewees as difficult to use, but would serve as a reliable long-term home for the data. Universities might have problems with open access of information.
IMS FRAMEWORK ELEMENT LINKAGES	Primarily Architecture
IMS FRAMEWORK IMPLICATIONS AND RECOMMENDATIONS	The lack of clarity on whether the IMS should be integrated with an existing system (state, university, or other) or an independent entity leaves this decision to be made in consideration of many external factors (many of which are independent of IMS issues). The architecture of the IMS (model of the way the IMS is set up; where the database, interface, and applications are stored, and communication pathways among these elements) should be one of the first issues to be addressed, potentially as discussions between possible partners (state, university, NGO, private).

5.4 Information Management System: Critical Path Recommendations

Results from the User Needs Assessment reflected the complex and variable nature of the potential user audience for the Monitoring Enterprise IMS. In this final section, we conclude by providing critical path recommendations integrated across UNA results that we consider to be important next steps in the IMS requirements, design, and development process.

- Evaluate the user persona and archetype constructs to prioritize the audience for the IMS, and from this, clarify and restate the system's objectives based on meeting the needs of the highest priority users.
- The "Explore the Conclusions" user persona category was the largest in terms of the number of individuals, included a representative cross-section of archetypes, and therefore might be a useful place to start. The common element in this category was the interest in summary information, and interest in a wide variety of online tools to view, search for, map, ask questions, and otherwise interact with the information.
- Internet survey respondents and telephone interviewees expressed interest in accessing the information from a local perspective regardless of whether their need for data was local, regional, or statewide; this indicates that a map interface will be a required component for the IMS. Evaluating options for what role the map interface will hold in the system architecture can provide a focal point for one aspect of IMS design.
- The keyword search is another useful tool that should be integrated into the design; additional tools should include only targeted or phased cost-effective tools.
- The separate process of developing the content of the monitoring information should be integrated into the IMS design and planning process as soon as possible so that decisions on data standards, analysis, and need for data sharing among program participants can be made.
- As part of the content, the database design should include systematic metadata-type information that is understandable to the layman that provides the "who, how, and where" of data collection, as well as a method for reporting error or uncertainty estimates.
- Once the basic database content has been defined, the issue of long-term stewardship should be addressed in tandem with considering partnerships for data and platform sharing solutions.
- Finally, the system requirements document should consider inclusion of a modular, phased approach so that data, functions, and other IMS aspects can be prioritized as the monitoring program proceed.

APPENDIX A

Service Status Update

Survey: Survey

Status: Launched (survey active)

1. Introduction

Copy page • Delete page •

Thank you for taking this survey for the Marine Protected Areas Monitoring Enterprise. The information collected in this User Needs Assessment survey will be used to inform the design of an internet-based information management system that will provide online access to MPA information and monitoring results. The survey responses will be used only for this project. Your name, email, and all other personal information will not be used for any other purpose. Your responses will be kept confidential, and will be summarized without any identification information. Final results will be available in a report from the Monitoring Enterprise later this year.

Throughout this survey * denotes a mandatory question and + denotes that a follow-up question will appear if this option is selected

I agree to take this survey with the understanding that this information will be used only for the purposes of this User Needs Assessment, and that my name and specific responses will not appear in the published report. *

- ○ Yes
- ○ No

2. Sign-In

Copy page • Delete page •

Please enter your email address to access this survey.

Email Address [ ]

If your email address is not working please click here to access the survey.

3. Contact Information

Copy page • Delete page •

Please update and ensure all contact information is correct.

Name *

Email Address *

Organization

Title

City

Postal Code

Phone Number

4. About You

Copy page • Delete page •

Choose the statement(s) that best describes your interest in MPA monitoring information (pick all that apply): *

- [ ] At least part of my job is related to the establishment, management, or monitoring of MPAs. +
- [ ] At least part of my job is related to marine science; education; marine resource use; policy; or management in California. +
- [ ] I conduct or participate in commercial or recreational activities that are affected by MPAs. +
- [ ] I belong to an organization that has interest in MPAs in California. +
- [ ] I am a student and/or interested citizen.
- [ ] Other +

Please select the best description of your professional affiliation. (pick one)

- ○ Government/Public Sector +
- ○ Education & Research +
- ○ Marine Industry +
- ○ Charitable, Non-Profit or Non-Governmental Organization (NGO) +
- ○ Other +

Please select the option which best describes your 'Government/Public Sector' work. (pick one)

- ○ Federal agency
- ○ State agency
- ○ County agency
- ○ City/Town
- ○ Tribal
- ○ Port/Harbor
- ○ Commission or Council [specify:] ______
- ○ Other: ______

Please select the option which best describes your 'Education/Research' work. (pick one)

- ○ Grades K-12
- ○ College or University (post- secondary)
- ○ Non-profit research institution (aquarium, research foundation, institute)
- ○ Adult or public outreach
- ○ Other ______

Please select the option which best describes your Marine Industry work. (pick one)

- ○ Aquaculture
- ○ Commercial fishing
- ○ Energy
- ○ Fishing or boating supplies or services
- ○ Marine tourism
- ○ Recreational fishing
- ○ Marine engineering
- ○ Marina
- ○ Port and commercial shipping
- ○ Private vessel sales or rentals (sail, power, PWC)
- ○ Scuba diving or snorkeling
- ○ Transportation
- ○ Watersports (kayaking, canoeing, surfing, windsurfing, kitesurfing)
- ○ Other: ____________

Please select the option that best describes your 'Charitable, Non-Profit or Non-Governmental Organization' work. (pick one)

- ○ Aquaculture
- ○ Boating (sailing, powerboating, PWCs)
- ○ Chamber of Commerce
- ○ Commercial fishing
- ○ Environment/Conservation/Access
- ○ Marine or coastal tourism
- ○ Marine trade association (other than aquaculture, boating, fishing, watersports or tourism)
- ○ Marine education
- ○ Philanthropy
- ○ Recreational fishing
- ○ Scuba diving, snorkeling or free-diving
- ○ Watersports
- ○ Other: ____________

Please describe your profession.

____

Please select the best description of the primary focus of your organization as it relates to MPAs (pick one):

- ○ Aquaculture
- ○ Boating (sailing, powerboating, PWCs)
- ○ Chamber of Commerce
- ○ Commercial fishing
- ○ Education/outreach
- ○ Environment/conservation/access
- ○ Marine or coastal tourism
- ○ Marine trade association (other than aquaculture, boating, fishing, watersports or tourism)
- ○ Recreational fishing
- ○ Scuba diving, snorkeling, freediving
- ○ Tribal
- ○ Watersports
- ○ Other: ________

Please select the best descriptions of your role within your organization as it relates to MPAs (pick all that apply):

- ☐ Make decisions about ocean or coastal policies or regulations
- ☐ Support decision-makers or public processes by conducting policy or scientific analyses
- ☐ Manage coastal or marine resources
- ☐ Enforce coastal or marine regulations
- ☐ Participate in the designation or administration of MPAs
- ☐ Advocate for changes to marine or coastal policies or regulations
- ☐ Develop software or web tools to support decision makers or public processes
- ☐ Manage MPA data or information
- ☐ Conduct fundamental scientific research
- ☐ Conduct applied research/engineering
- ☐ Develop or deliver education programs
- ☐ Provide funding
- ☐ Provide administrative or technical support to my organization
- ☐ Other: ________

Please select all that describe your marine-related interests or role(s) of organizations to which you belong. (pick all that apply)

- ☐ Beachgoing
- ☐ Conservation
- ☐ Diving, Scuba, snorkeling, or freediving – non-consumptive
- ☐ Diving, snorkeling, or freediving – consumptive (e.g., spear fishing)
- ☐ Education or outreach
- ☐ Environmental
- ☐ Fishing – commercial
- ☐ Fishing – recreational, boat-based
- ☐ Fishing – recreational, shore-based
- ☐ Ocean kayaking or canoeing
- ☐ Personal Water Craft (PWC)
- ☐ Power boating
- ☐ Sailing
- ☐ Tide pooling
- ☐ Watersports (swimming, surfing, windsurfing, kitesurfing)
- ☐ Wildlife viewing
- ☐ Other: ________

Please explain your interest in MPAs.

5. About You

Copy page • **Delete page** •

I have an interest in the monitoring of the following MLPA regions (pick all that apply):

- [ ] Central Coast
- [ ] North Coast
- [ ] North Central Coast
- [ ] San Francisco Bay
- [ ] South Coast
- [ ] All of California

I am interested in accessing information about (pick all that apply):

- [ ] One or more specific MPA(s), individually +
- [ ] The MPAs within a study region
- [ ] The full statewide MPA network

Which specific MPA(s) are you interested in?

How often do you use the following types of software?

	Never	Rarely	Sometimes	Frequently
Web-browser	○	○	○	○
Email	○	○	○	○
Word processing	○	○	○	○
Spreadsheet	○	○	○	○
Database	○	○	○	○
Graphics (illustration, production)	○	○	○	○
Image processing	○	○	○	○
GIS	○	○	○	○
Modeling tools	○	○	○	○
Statistical Software	○	○	○	○
Programming Languages	○	○	○	○

Are there any other types of software that you consider important?

6. MPA Monitoring Information

Copy page • Delete page •

The questions in this section apply only to monitoring data specifically collected to address the effectiveness of the MPAs in meeting MLPA goals and objectives. For more information, click on this hyperlink to learn more about the goals and objectives of the MLPA. *Please note that this survey is NOT being used to develop monitoring plans, priorities, or methods. Detailed MPA monitoring plans will be developed in separate, consultative processes involving stakeholders, scientists, resource managers, and decision-makers and using a public review process. The purpose of this section is to help us consider the broad types of data and information the information system may need to accommodate.*

I would like access to the following types of MPA monitoring information. (pick all that apply)

- ☐ Information summaries, reports for non-specialists or key findings as they relate to the goals and objectives of specific or multiple MPAs.
- ☐ Scientific reports that describe monitoring results and conclusions drawn from the collected data.
- ☐ Summary tables, presentations, graphs, images, videos, or maps. +
- ☐ Detailed tables of data, statistical summaries, compiled databases. +
- ☐ Specific data formats, including GIS, time-series, or raster files. +
- ☐ Raw field/laboratory data with associated calibration or validation information. +
- ☐ Other: ____________

What options for information summaries that are automatically generated would you be interested in accessing?

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Data Tables	○	○	○	○	○
Graphs	○	○	○	○	○
Photographs	○	○	○	○	○
Videos	○	○	○	○	○
Maps	○	○	○	○	○

What options for processed and compiled data would you be interested in accessing?

	Not Needed	Nice To Have	Very Useful	Essential	Not Sure
Databases	○	○	○	○	○
Statistical Summaries	○	○	○	○	○

What specific data formats are you interested in using?

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Time-Series Data	○	○	○	○	○
Raster Data	○	○	○	○	○
Photograph or Video Library	○	○	○	○	○
GIS-Formatted Data	○	○	○	○	○

Assuming you want to download raw data, select the option that best describes your expectations. (pick one)

- ○ I would prefer extracting data in one file, and would expect that the data have been standardized across all fields (units, nomenclature, etc.).
- ○ I would prefer downloading data files from the original sources, even though there may be a differences (in units, etc.), as long as the data standards are well-documented.
- ○ I would prefer being able to select a data standard that meets my own specifications.
- ○ I'm not sure.

Please select the most common data formats you have downloaded to use in the past. (pick all that apply)

- ☐ Text files (ASCII, comma-delimited, etc.)
- ☐ Spreadsheets
- ☐ Databases
- ☐ GIS file formats
- ☐ I don't generally download data.
- ☐ Other: ______

What level of metadata (methods, calibration, documentation, and validation information) should be included in the Monitoring Enterprise database? (pick all that apply)

- [ ] Standardized metadata, including data source, methodology, validation, exclusions, etc. for every dataset.
- [ ] Summary metadata/methods for every dataset.
- [ ] Methods described in monitoring reports or publications.
- [ ] Programmatic methods and standards documents.
- [ ] Other: ____________

7. MPA Monitoring Information

Copy page • Delete page •

What types of monitoring information and results are you interested in learning about for California's MPAs from the MPA monitoring information system?

Please remember that this survey is not being used for monitoring planning, but simply to identify possible types of information.

Ecosystem Types

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Sandy beaches	○	○	○	○	○
Estuaries	○	○	○	○	○
Shallow soft bottom	○	○	○	○	○
Deep soft bottom	○	○	○	○	○
Rocky intertidal	○	○	○	○	○
Kelp and shallow rocky bottom	○	○	○	○	○
Deep rocky bottom	○	○	○	○	○
Pelagic and water column	○	○	○	○	○
Submarine canyons	○	○	○	○	○

If there is a different type of ecosystem that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

____

Species Groups

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Marine Mammals	○	○	○	○	○
Birds	○	○	○	○	○
Fish	○	○	○	○	○
Plants/algae	○	○	○	○	○
Invertebrates	○	○	○	○	○

If there is a different type of species group(s) that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

Human Activities

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Commercial fishing & harvesting	○	○	○	○	○
Recreational fishing & harvesting	○	○	○	○	○
Aquaculture & hatcheries	○	○	○	○	○
Recreational boating	○	○	○	○	○
Scuba diving & snorkeling	○	○	○	○	○
Beachgoing, swimming, surfing	○	○	○	○	○
Education & research	○	○	○	○	○
Shipping & marine transportation	○	○	○	○	○
Energy development & seabed mining	○	○	○	○	○

If there is a different type of activity that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

Summary Information of Different Types

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Summaries focused on individual MPAs	○	○	○	○	○
Summaries focused on individual species	○	○	○	○	○
Summaries focused on ecosystem types	○	○	○	○	○
Summaries at the scale of the MLPA regions	○	○	○	○	○
Summaries at the statewide scale	○	○	○	○	○
Summaries over a specific time period	○	○	○	○	○

If there is a different type of summary that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

8. MPA Monitoring Information

Copy page • **Delete page** •

What other monitoring-related information would you like access through the MPA monitoring information system?

MPA Descriptive Information

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Boundaries/Size	○	○	○	○	○
Location	○	○	○	○	○
Type	○	○	○	○	○
Goals/Objectives	○	○	○	○	○
Uses & allowed activities	○	○	○	○	○
Regulations	○	○	○	○	○

If there is a different type of descriptive information that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

Publications About Monitoring

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
California-based monitoring objectives and protocols	○	○	○	○	○
Monitoring protocols from other states or countries	○	○	○	○	○
General aquatic and marine monitoring concepts and theories	○	○	○	○	○
Documents about MPAs in other states or countries	○	○	○	○	○

If there is a different publication about monitoring that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

Educational Resources

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
MPA science bibliography for students (K-12)	○	○	○	○	○
MPA laws/policies for students (K-12)	○	○	○	○	○
Instructional DVDs	○	○	○	○	○
Posters/brochures	○	○	○	○	○
Teacher guides	○	○	○	○	○

If there is a different educational resource that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

9. MPA Monitoring Information

Copy page • **Delete page** •

What additional data would you like to access in conjunction with MPA Monitoring data?

Other Environmental Data

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Seabed geology/bathymetry	○	○	○	○	○
Oceanographic (physical/time-series)	○	○	○	○	○
Water properties/water quality	○	○	○	○	○
Satellite or aerial imagery	○	○	○	○	○

Is there a different type of monitoring program or data type that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

Program

Type of Information

Is there any type of ancillary biological data that you would find helpful?

Is there any other type of ancillary data that you would find helpful?

10. Experience

Copy page • **Delete page** •

In this section, we are interested in your experience with other websites, both marine-related and others that you use regularly. We are also interested in how you access websites to get information.

Will you provide feedback on websites that you use?

☐ I can provide feedback on coastal/ocean or science-related websites. +

☐ I can provide feedback on non-marine website(s). +

☐ I choose not to provide feedback at this time.

Click here for a list of suggested websites to review

Please enter a name of up to three websites (URLs) and rate each of the attributes listed.

Website 1

Please provide a website URL or name.

	Poor	Fair	Good	Excellent	Not Applicable/No Opinion
Data access/download	○	○	○	○	○
Data upload	○	○	○	○	○
Design, look and feel	○	○	○	○	○
Ease of finding information	○	○	○	○	○
Metadata/documentation	○	○	○	○	○
Quality of content	○	○	○	○	○
Tools or utilities	○	○	○	○	○
Graphs and charts	○	○	○	○	○
Mapping information	○	○	○	○	○

Comments on Website 1:

Website 2

Please provide a website URL or name.

	Poor	Fair	Good	Excellent	Not Applicable/No Opinion
Data access/download	○	○	○	○	○
Data upload	○	○	○	○	○
Design, look and feel	○	○	○	○	○
Ease of finding information	○	○	○	○	○
Metadata/documentation	○	○	○	○	○
Quality of content	○	○	○	○	○
Tools or utilities	○	○	○	○	○
Graphs and charts	○	○	○	○	○
Mapping information	○	○	○	○	○

Comments on Website 2:

Website 3

Please provide a website URL or name.

	Poor	Fair	Good	Excellent	Not Applicable/No Opinion
Data access/download	○	○	○	○	○
Data upload	○	○	○	○	○
Design, look and feel	○	○	○	○	○
Ease of finding information	○	○	○	○	○
Metadata/documentation	○	○	○	○	○
Quality of content	○	○	○	○	○
Tools or utilities	○	○	○	○	○
Graphs and charts	○	○	○	○	○
Mapping information	○	○	○	○	○

Comments on Website 3:

Do you have specific feedback on good and bad aspects of websites that you have used that you would be willing to discuss in a follow-up phone conversation?

○ Yes +

○ No

Please briefly provide some of your ideas or suggestions that you would like to discuss.

Our records show that your phone number is: [%%269:Phone Number %%]

If a phone number is not shown or it is incorrect would you please enter/update your contact details.

11. Experience

Copy page • **Delete page** •

Please read through the list of potential problems you may have encountered when using a website, and rate how this might prevent or inhibit your use of the site:

Design Issues

	Bothersome	Annoying	Irksome	Prohibitive (will exit)	No Experience/ Not Sure
Too busy, too many options	○	○	○	○	○
Disorganized, unclear how to start	○	○	○	○	○
Excessive use of gadgets/plugins, distracting	○	○	○	○	○
Too slow to load	○	○	○	○	○
Poor site navigation, time-consuming to move around	○	○	○	○	○

Information Access Issues

	Bothersome	Annoying	Irksome	Prohibitive (will exit)	No Experience/ Not Sure
Lack of useful data	○	○	○	○	○
Disorganized data	○	○	○	○	○
Detailed data unavailable	○	○	○	○	○
Lack of documentation or metadata	○	○	○	○	○
No way to select or subset data	○	○	○	○	○

Are there any functions that you would like to see provided by the MPA monitoring information system that have not yet been mentioned?

○ I can't think of any

○ Yes (please describe) [ ]

I expect to access the MPA monitoring information system from (check all that apply):

- [ ] A desktop computer from my home or office
- [ ] A desktop computer from the library or other public place
- [ ] A laptop computer using wireless access
- [ ] A mobile, hand-held device
- [ ] Other: ______

Do you have specific need for information formatted for a mobile application? For example, marine navigation software.

- ( ) No
- ( ) Yes: (Please describe) ______

Are there any ways of accessing or receiving information that you would like to see provided by the MPA monitoring information system that have not yet been mentioned in the questions above?

- ( ) I can't think of any
- ( ) Yes: ______

12. Communicate & Collaborate

Copy page • Delete page •

This section of the survey asks questions about what kind of communication tools the MPA monitoring information system should contain.

Please rate the importance of these communication and collaboration functions of the MPA monitoring information system website.

Collaboration/Networking

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
MPA monitoring contacts list	( )	( )	( )	( )	( )
Collaborative analysis tools	( )	( )	( )	( )	( )
Collaborative publication tools	( )	( )	( )	( )	( )
Discussion forums/wiki sites	( )	( )	( )	( )	( )
Blogs	( )	( )	( )	( )	( )
Decision support tools	( )	( )	( )	( )	( )

If there are any additional Collaborating/Networking features that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

Meeting/Conference Information

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Announcements	○	○	○	○	○
Minutes	○	○	○	○	○
Web broadcasting/webinars	○	○	○	○	○

If there are any additional Meeting/Conference features that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

Educational

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Q&A for students	○	○	○	○	○
Chat with a scientist	○	○	○	○	○

If there are any additional Educational features that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

I think there should be support for languages other than English

- ○ No, online translators or community interpreters are sufficient
- ○ No opinion
- ○ Yes, which language(s)

I think the MPA monitoring information system website should have support for disabilities.

- ○ Yes +
- ○ No
- ○ No opinion

13. Communicate & Collaborate: Sup

Copy page • Delete page •

What support for disabilities should the MPA monitoring information system website contain? Note that these are guidelines not prescriptions. (pick all that apply)

Perceivable

- [ ] Provide text alternatives for non-text content.
- [ ] Provide captions and alternatives for audio and video content.
- [ ] Make content adaptable; and make it available to assistive technologies.
- [ ] Use sufficient contrast to make things easy to see and hear.

Operable

- [ ] Make all functionality keyboard accessible.
- [ ] Give users enough time to read and use content.
- [ ] Do not use content that causes seizures.
- [ ] Help users navigate and find content.

Understandable

- [ ] Make text readable and understandable.
- [ ] Make content appear and operate in predictable ways.
- [ ] Help users avoid and correct mistakes.

14. Data Analyzers

Copy page • Delete page •

These questions are for people who would like to build custom summaries (including maps and other graphics) of MPA monitoring information.

Please select the statements that are most consistent with your opinion about online data analysis tools (pick all that apply).

- [ ] The MPA monitoring information system should allow access to data and information, but analysis should be done by professionals and on-line analysis tools are unnecessary.
- [ ] Online tools to help me explore the information further, ask questions, or make my own map, would be useful and should be available on or through the MPA monitoring information management system. +
- [ ] Online tools to help me explore what information is available prior to downloading data for my own research or analysis would be useful and should be available on or through the MPA monitoring information management system. +
- [ ] I expect to do my analysis using my own software, and would plan only on searching for and downloading the data that I need.
- [ ] Other: ____________

Please rate the types of data analysis tools in terms of you useful you think they might be for your own data analysis activities (if you don't understand a tool or how you would use it, select 'Not Sure'):

Question and Summarize

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Query content by keyword	○	○	○	○	○
Summary statistics	○	○	○	○	○

If there are any other types of 'Question and Summarize' tools that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

____

Maps

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Map overlay tool	○	○	○	○	○
Spatial queries	○	○	○	○	○

If there are any other types of 'Map' tools that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

____

Charts and Graphs

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
Select variables and plot standard 2D charts	○	○	○	○	○
Frequency distributions or other statistical charts	○	○	○	○	○
Custom charts	○	○	○	○	○
2D motion chart tool (plots X/Y data through time or other variable)	○	○	○	○	○

If there are any other types of 'Chart and Graph' tools that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

Multi-media

	Not Needed	Nice to Have	Very Useful	Essential	Not Sure
2.5D/3D visualization tool (fixed frame of reference)	○	○	○	○	○
3D motion chart tool (plots X/Y/Z data through time or other variable)	○	○	○	○	○

If there are any other types of 'Multi-media' tools that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

If there are any types of Models or Applications that you are interested in please explain and rank from 1 to 3 below. (1 is Nice to Have and 3 is Essential)

Do you have specific ideas or suggestions for tools that you would be willing to discuss in a follow-up phone conversation?

○ Yes +

○ No

Please briefly provide some of your ideas or suggestions that you would like to discuss.

Our records show that your phone number is: [%%269:Phone Number %%]

If a phone number is not shown or it is incorrect would you please enter/update your contact details.

15. Data Providers

Copy page • Delete page •

These questions are for people who expect to or may provide data to the MPA monitoring information system.

Please select all statements that apply:

- ☐ I have been, or expect to be, involved with the collection of MPA monitoring data. +
- ☐ I, or my institution, collect data associated with or in the vicinity of MPAs that could provide valuable information relative to formal MPA monitoring. +
- ☐ I have been providing, or might provide in the future, information from consumptive uses (e.g., commercial or recreational fishing). +
- ☐ I have been involved with an NGO/citizen's monitoring program that collects data that could provide valuable information relative to formal MPA monitoring. +
- ☐ I have experience uploading content on the internet. +
- ☐ I do not expect that I will ever provide data to the MPA Monitoring Enterprise.

If you anticipate submitting monitoring data, please select all statements that apply:

- ☐ I will need to have a signed agreement to participate in data exchange.
- ☐ I or my institution would be willing to share data from our existing information management system.
- ☐ There are firewall or other access restrictions to sharing data from my institution.
- ☐ I am not aware of any data sharing issues.
- ☐ I am not responsible for any data sharing issues.
- ☐ I have confidentiality or other access restriction concerns in submitting or sharing my data. Please describe:

If your organization has an existing internet-based information management system, please check all that apply:

- ☐ We would be willing to add a link to the MPA monitoring information system website.
- ☐ We would be willing to serve as a data node to supply data to the MPA monitoring information system website, assuming reciprocal data sharing.
- ☐ We would be willing to send updates of our data to the MPA monitoring information system website as they are available
- ☐ We would be willing to add metadata links of MPA monitoring data to our own inventory.
- ☐ I'm not responsible for data sharing.
- ☐ Other data sharing possibilities: ______

Please rate your experience with using the following methods to upload data over the internet:

	Problematic	OK	Worked Well	Ideal	No Experience/ Not Sure
Uploaded files using my own formats	○	○	○	○	○
Uploaded data using a standardized template	○	○	○	○	○
Uploaded my data with standard metadata file	○	○	○	○	○
Uploaded data in a GIS format	○	○	○	○	○
Submitted metadata only for files stored at my location	○	○	○	○	○
Used a wizard that walked through the upload process	○	○	○	○	○
Published data via web services that are available for others	○	○	○	○	○

Do you have expertise with, or an opinion about, data transfer formats, metadata standards, file delivery formats, internet standards, or other information management approaches that you would like to discuss in a follow-up phone call or discussion group for this project?

- ○ Yes +
- ○ No

Please briefly provide some of your ideas or suggestions that you would like to discuss.

______

Our records show that is your phone number: [%%269:Phone Number %%]

If a phone number is not shown or it is incorrect would you please enter/update your contact details.

16. Final Thoughts

Copy page • Delete page •

(Optional) Now that you have taken the survey, do you have any other suggestions or “out-of-the-box” ideas that have not been addressed for the MPA Monitoring Enterprise information management system? List up to three:

1

2

3

(Optional) What do you think are the top three most critical aspects that should, beyond all other possible information sources, utilities or functions, be implemented in the MPA Monitoring Enterprise information management system?

1

2

3

'Thank You'/Redirect Page

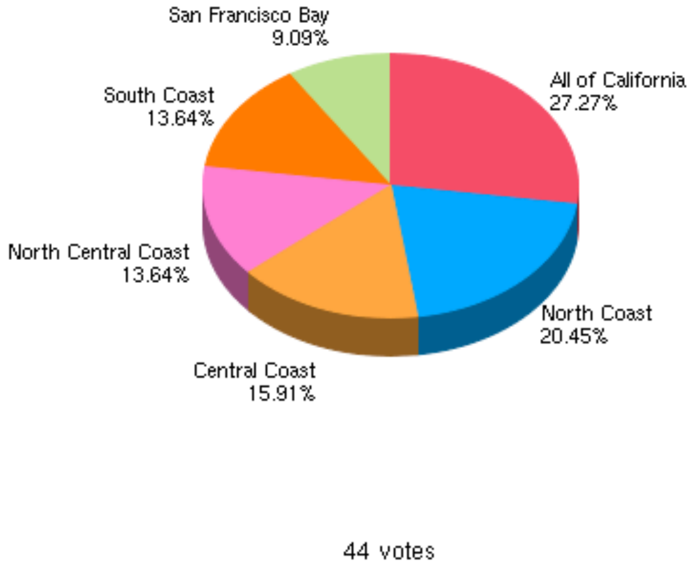
Thanks [%%258:First Name %%]!

Your response is very important to us.

The final results of this User Needs Assessment will be published on the MPA Monitoring Enterprise website later this year. You will receive an email when the report is available for viewing or download.

Shown below are the various MPA Region's that survey takers are interested in.

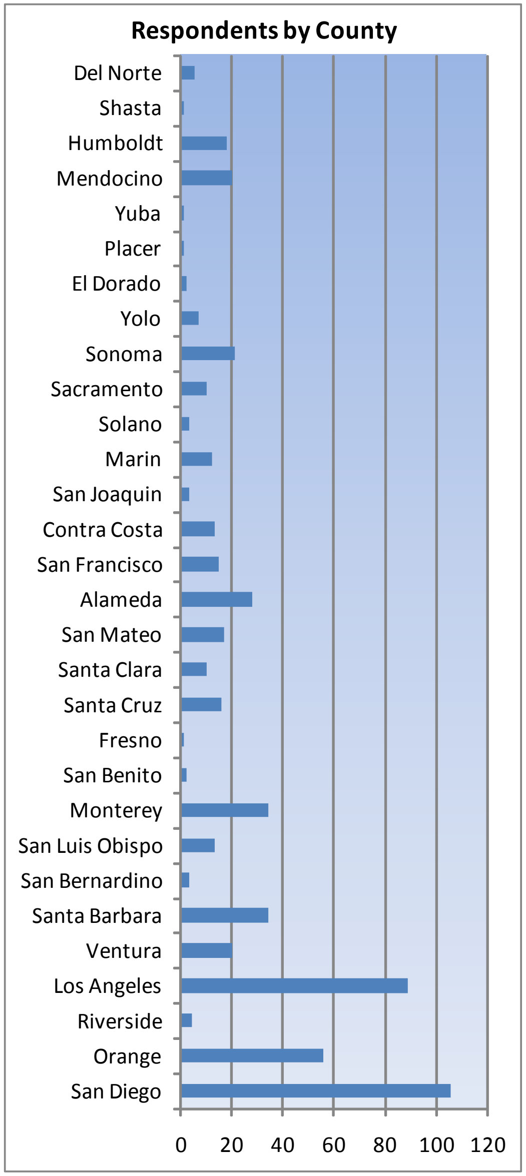
Shown below are the various MPA Region's that survey takers are interested in.

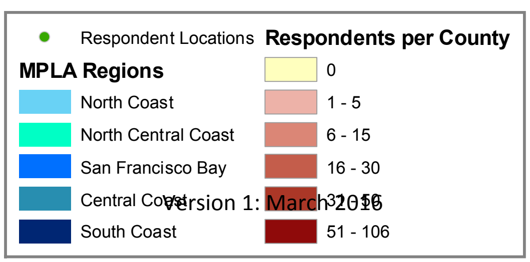


Want to show the this graph elsewhere? Here is the html for the image:

APPENDIX B

Survey Respondents: Geographic Distribution





Version 1: March 2016

Appendix C: Data Portal Inventory

Data Portal	URL	Description	Data Types	Geographic Focus	Owner/Funder
DataONE	https://www.dataone.org/	Data Observation Network for Earth (DataONE) is the foundation of new innovative environmental science through a distributed framework and sustainable cyberinfrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data.	Geospatial, Geographic, Atmosphere/Climate, Biology/Ecology, Geophysical, Sociology	USA	National Science Foundation
PISCO	http://www.piscoweb.org/data/data-access-and-applications	Over the last 10 years, PISCO has successfully built a unique research program that combines complementary disciplines to answer critical environmental questions and inform management and policy. Activities are conducted at the latitudinal scale of the California Current Large Marine Ecosystem along the west coast of North America, but anchored around the dynamics of coastal, hardbottom habitats and the oceanography of the nearshore ocean – among the most productive and diverse components of this ecosystem. The program integrates studies of changes in the ocean environment through ecological monitoring and experiments. Scientists examine the causes and consequences of ecosystem changes over spatial scales that are the most relevant to marine species and management, but largely unstudied elsewhere.	Biological, Chemical, Physical, Intertidal, Subtidal, Oceanography, Population Connectivity, MPA, MARINe, NMS	West Coast	Consortium: Oregon State University, Stanford University's Hopkins Marine Station, University of California Santa Cruz and University of California Santa Barbara
MarineBIOS	https://www.wildlife.ca.gov/Conservation/Marine/GIS/MarineBIOS	The California Department of Fish and Wildlife offers an interactive map for referencing relevant marine resource planning data. This tool, which is built on the latest version of BIOS, is a great place for looking up the boundaries and regulations of marine protected areas or investigating the attributes of benthic and intertidal habitat information.	Geospatial, MPAs, Habitats, Hydrography, Natural Resources	California	California Department of Fish and Wildlife
West Coast Ocean Data Portal	http://portal.westcoastoceans.org/	The West Coast Ocean Data Portal is a project of the West Coast Governors Alliance on Ocean Health (WCGA) to increase discovery and connectivity of ocean and coastal data and people to better inform regional resource management, policy development, and ocean planning. The Portal informs priority West Coast ocean issues such as tracking sources and patterns of marine debris, adaptation to sea-level rise, understanding impacts of ocean acidification on our coasts, and marine planning.	Marine Debris, Human Use, Biological, Physical	West Coast	West Coast Governors Alliance
National Marine Protected Areas Center	http://marineprotectedareas.noaa.gov/dataanalysis/mpainventory/	The Marine Protected Areas Inventory (MPA Inventory) is a comprehensive geospatial database designed to catalog and classify marine protected areas within U.S. waters. The Inventory contains information on over 1,700 sites and is the only such comprehensive dataset in the nation.	Ecology, Geospatial	USA	National Ocean Service
State of California Geoportal	http://portal.gis.ca.gov/geoportal/catalog/main/home.page	A primary goal of the California Geoportal is to improve access to California's geographic data portfolio, and expand the creative use of those data resources. The California Geoportal strives to increase information transparency, and is committed to creating an open environment for accessing important government derived geographic data.	Geospatial, Geographic, Atmosphere/Climate, Biology/Ecology, Geophysical, Sociology	California	California Department of Technology
Marine Cadastre	http://marinecadastre.gov/	MarineCadastre.gov was developed through a partnership between the U.S. Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) Office for Coastal Management and the U.S. Department of the Interior's Bureau of Ocean Energy Management (BOEM). MarineCadastre.gov is an integrated marine information system that provides data, tools, and technical support for ocean and Great Lakes planning. MarineCadastre.gov was designed specifically to support renewable energy siting on the U.S. Outer Continental Shelf but also is being used for other ocean-related efforts.	Geospatial, Biology, Ecology, Geographic, Human Use, Geophysical	USA	National Oceanic and Atmospheric Administration and Bureau of Ocean Energy Management
Knowledge Network for Biocomplexity	https://knb.ecoinformatics.org/	For scientists, the KNB is an efficient way to share, discover, access and interpret complex ecological data. Due to rich contextual information provided with KNB data, scientists are able to integrate and analyze data with less effort. The data originate from a highly-distributed set of field stations, laboratories, research sites, and individual researchers. The foundation of the KNB is the rich, detailed metadata provided by researchers that collect data, which promotes both automated and manual integration of data into new projects.	Geospatial, Ecology, Environmental,	Global	University of California Santa Barbara
University of California Digital Library	http://www.cdlib.org/uc3/merritt/	Merritt is a new cost-effective repository service from the University of California Curation Center (UC3) that lets the UC community manage, archive, and share its valuable digital content.	Various	Various	University of California
Central & Northern California Ocean Observing System	http://www.cencoos.org/data	CeNCOOS is the Central and Northern California Ocean Observing System, we are part of a national framework of integrated coastal observing systems covering all coastal areas throughout the United States. This national framework is called the Integrated Ocean Observing System (IOOS).	Geospatial, Oceanographic	California	Integrated Ocean Observing System
Oceanographic Decision Support System	http://odss.mbari.org/odss/	At MBARI, we utilize the ODSS for situational awareness, experiment planning, collaboration and data analysis. In the ODSS, ocean scientists collaboratively design their experiments, communicate with other participants, track asset locations and command robotic vehicles at sea. Data from these assets are retrieved when communication is possible, stored, then made available for download or playback from a single portal for further analysis either by human or machine learning methods. The latter is particularly useful in generating scientific publications post-facto the experiment.	Geospatial, Oceanographic	California	Monterey Bay Aquarium Research Institute
Ocean Data Portal	http://www.oceandataportal.org/	Ocean Data Portal provides a "one-stop shop" approach to oceanographic data held by the IODE global network of 80 National Oceanographic Data Centres (NODCs), as well as to resources from other participating systems. The ODP provides infrastructure to integrate collections and inventories of marine data and allow for the directory, evaluation and access to data via web services to meet the needs of the whole ocean community.	Geospatial, Oceanographic	Global	International Oceanographic Data and Information Exchange (UNESCO)
Northeast Ocean Data	http://www.northeastoceandata.org/	NortheastOceanData.org is an information resource and decision support tool for ocean planning from the Gulf of Maine to Long Island Sound. The website provides user-friendly access to maps, data, tools, and information needed for regional ocean planning. It is used by a broad range of government entities, non-government organizations, and ocean stakeholders.	Geospatial, Aquaculture, Human Uses, Biology, Ecology, Energy, Water Quality	Northeastern USA	Northeast Regional Ocean Council
Mid-Atlantic Ocean Data Portal	http://midatlanticocean.org/data-portal/	The Mid-Atlantic Ocean Data Portal is an online toolkit and resource center that consolidates available data and enables state, federal, and local users and the general public to visualize and analyze ocean resources and human use information such as fishing grounds, recreational areas, shipping lanes, habitat areas, and energy sites, among others.	Geospatial, Administrative, Fishing, Marine Life, Maritime Industry, Recreation, Energy, Security	Mid-Atlantic USA	Mid-Atlantic Regional Council on the Ocean
Data.Gov Ocean	https://www.data.gov/ocean/	Discover and access data, information, and decision tools related to our ocean, coasts, and Great Lakes. The National Ocean Council provides this site to support regional marine planning efforts across the country and to create a convenient place for anyone to find out more about our marine, coastal, and Great Lakes environments.	Geospatial, Oceanographic, Energy, Climate, Safety, Ecosystem, Disasters, Finance, Education	USA	U.S. General Services Administration

March 2016

Data Portal	URL	Description	Data Types	Geographic Focus	Owner/Funder
Massachusetts Ocean Resource Information System	http://maps.massgis.state.ma.us/map_ol/moris.php	MORIS, the Massachusetts Ocean Resource Information System, is an online mapping tool created by the Massachusetts Office of Coastal Zone Management (CZM), the Office of Geographic Information (MassGIS), SeaPlan, Applied Science Associates (ASA), Charlton Galvarino, and PeopleGIS. MORIS can be used to search and display spatial data pertaining to the Massachusetts coastal zone.	Geospatial, Biology, Coastal Hazards, Geology, Fisheries, Oceanography, Sociology, Infrastructure	Massachusetts	Massachusett's Office of Coastal Zone Management
Sanctuary Integrated Monitoring Network	http://sanctuarysimon.org/	SIMoN is an integrated, long-term program that takes an ecosystem approach to identify and understand changes within sanctuaries managed by the Office of National Marine Sanctuaries. SIMoN provides resource managers with the information needed for effective decision-making and promotes an unparalleled basic understanding of the complex and unique marine processes within the California Current ecosystem.	Geospatial, Ecology, Water Quality, Sanctuary Characterization	California	National Oceanic and Atmospheric Administration
EcoAtlas	http://www.ecoatlas.org/	California EcoAtlas provides access to information for effective wetland management. The maps and tools can be used to create a complete picture of aquatic resources in the landscape by integrating stream and wetland maps, restoration information, and monitoring results with land use, transportation, and other information important to the state's wetlands.	Geospatial, Wetlands	California	San Francisco Estuary Institute
Nearshore Ecosystem Database - Reef Check CA	http://ned.reefcheck.org/	Reef Check California is a network of informed and involved citizens, scientists and organizations that support the sustainable use and conservation of our nearshore marine resources. NED is your tool to input data from Reef Check California Surveys and view/query data collected at Reef Check California sites.	Geospatial, Ecology	Global	Reef Check California
MarinExplore	http://marinexplore.org/	Marinexplore.org is the easiest way to explore, discover, and share public ocean data.	Geospatial, Oceanographic, Energy, Climate, Ecosystem	Global	Planet OS
Biological & Chemical Oceanography Data Management Office	http://www.bco-dmo.org/	The Biological and Chemical Oceanography Data Management Office (BCO-DMO) staff members work with investigators to serve data online from research projects funded by the Biological and Chemical Oceanography Sections and the Division of Polar Programs Antarctic Organisms & Ecosystems Program at the U.S. National Science Foundation.	Geospatial, Oceanographic	Global	National Science Foundation
University Corporation for Atmospheric Research	http://cdp.ucar.edu/	Fostering deeper understanding of the atmosphere, Earth, and Sun - The University Corporation for Atmospheric Research is a nonprofit consortium of more than 100 North American member colleges and universities focused on research and training in the atmospheric and related Earth system sciences.	Geospatial, Atmospheric, Earth Sciences	Global	National Science Foundation
REEF	http://www.reef.org/db/reports	REEF was founded in 1990, out of growing concern about the health of the marine environment, and the desire to provide the SCUBA diving community a way to contribute to the understanding and protection of marine populations. REEF achieves this goal primarily through its volunteer fish monitoring program, the REEF Fish Survey Project. Participants in the Project not only learn about the environment they are diving in, but they also produce valuable information. Scientists, marine park staff, and the general public use the data that are collected by REEF volunteers.	Ecology	Global	REEF
MPA Watch	http://www.mpawatch.org/site/mpamap	MPA Watch trains volunteers to observe and collect unbiased data on coastal and marine resource use. Volunteers will be trained to collect valuable data on ocean users and their activities, such as surfing, kayaking, fishing, boating, running, etc. Specifically, the MPA Watch volunteers will observe and record both consumptive and non-consumptive offshore and onshore activities in and around MPAs, which will improve our understanding of how people are using these new MPAs.	Human Uses	California	MPA Watch

Data Portal Inventory

March 2016

Data and Information Management

Workshop Summary

Southwest Fisheries Science Center, 110 Shaffer Road, Santa Cruz, CA
14 October 2015

WORKSHOP OBJECTIVES

- Engage with partners on the state, regional, and national levels and draw on their expertise and lessons-learned to inform the development of the Data and Information Management plan.
- Present the updates required for the MPA monitoring Data and Metadata standards, and formalize those changes to ratify the statewide standards.
- Analyze the needs of the main user groups, and determine the minimum viable product requirements for the data management system.

WORKSHOP OVERVIEW

Stated Goal: *Create a system that leverages existing resources as much as possible, and integrates with existing work without duplicating effort*

This workshop is one component of a larger process for developing a Data and Information Management Plan for California's MPA monitoring data and results. As we turn our attention to long-term monitoring, we need to develop a comprehensive, strategic and feasible plan that articulates an approach that leverages existing public and private technology solutions, and serves as a roadmap for building the partnerships and data infrastructure in the future.

During this workshop, the Advisory Team together with additional technical, scientific, management, and policy experts (see page 10), discussed several of the main components of and key concepts within the plan:

- Defining the Terms: Data, Information, and Assets
- User Stories
- Data and Metadata Standards
- Data Discovery and Integration
- Data Visualization
- Funding Scenarios

For a complete outline of the plan, see Appendix A.

Key Point: *Initial comments focused on how the California Fish and Game Commission (FGC) would use data and information management system to help make decisions. State representatives suggested that the FGC would likely not use the site, but stakeholders would.*

WORKSHOP SUMMARY

Defining the Terms: Data, Information, and Assets

The definition for data offered to kick-start the conversation was: *raw data collected as part of MPA monitoring projects.* Participants engaged in a rich conversation about the broad spectrum of information that the word "data" encompasses, and quickly came to the conclusion that there is an artificial divide between data and information. Thus, we should bring information – *any product derived from the raw data and metadata* – under the umbrella term of "data" (e.g., derived datasets). Information also includes products that come from data, such summaries or syntheses created to communicate a specific story or result. It can fall along the same spectrum as data, however, because the products could be used as datasets to generate further derived datasets or products.

Participants also suggested we ensure the definition is broad enough to encompass data produced through socioeconomic monitoring (e.g., survey results, fisheries revenue) and geospatial data (e.g., map layers). Because OceanSpaces will reference, and perhaps even include, data beyond MPA monitoring data (e.g., outside the state-funded projects, enforcement data, contextual information), participants agreed that the definition should not be limited to "MPA monitoring data". By broadening the definition, we can demonstrate leveraging additional resources, and that we are not precluding other types of data. For the formal definition of "data", the words field or measurement would provide a neutral starting point.

Scripts

Not a data product themselves, but participants identified the need for preserving understanding how raw data are transformed or changed to create secondary or tertiary datasets. We should consider how these scripts can be stored with or linked to within the data management system. Metadata should be included to describe how these scripts interact with the raw data to create derived datasets.

Assets

In addition to data and information, MPA monitoring and other research generated *assets*, such as imagery (e.g., photographs, video). These assets are not data themselves, but require further processing to generate datasets. Currently, these assets do not exist alongside the datasets derived from them. However, there is interest in comparing imagery across temporal and geographic scales, so participants agreed that there is value in establishing this link. This category will require further discussion with imagery creators and users.

Suggested Resource: http://science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products/

Plan Requirement: *To properly describe data within this plan, we should adopt a spectrum of data levels that map to the data storage system.*

User Stories, from discovery to visualization

User Persona #1: Data Contributor

This user group was well represented in the workshop participants. The basic interactions with the data management system will involve uploading and possibly downloading data. Because data submission is a contract requirement for project leaders, there is no need to establish a goal or benefit of this for data contributors. However, additional services - such as data provenance and linking data with publications and other products - is perceived as a benefit. These services would need to be handled by Ocean Science Trust or on the back-end (technology solution) as project leaders do have time to update the system themselves. BOEM and others agree that counting publications and data use metrics are valuable to demonstrating the broader impact and utility of monitoring programs. At this time, provenance is still very complex, but building on the work of DataONE should be kept in mind as we design the management system.

Data contributors see a larger benefit in contributing data and having the system automatically output the data in formats and products. This would provide the impetuous for greater participation in contributing data and engagement with the system more broadly. There was broad agreement that linking reports and visualizations to data would be a major benefit to this user group. Finally, an important overall goal of the management system should be to ensure it is easy to use, allows for server-to-server transfers, and can provide useful outputs.

Suggested goal: *Data contributors are often required to engage with the system as part of their contracts. Voluntary participation in the community can be encouraged through opportunities for data sharing, gaining access to other data, and additional services, such as data crosswalks, producing stock visualizations, and linking to related reports and other products.*

User Persona #2: Resource Manager/Decision-Maker

This user group will likely interact with a map-based data discovery tool, and then access high-level reports or visualizations. Resource managers an decision-makers want to know that raw data are available and linked to the relevant reports, but will likely only interact with the reports and visualizations. Moving forward, many MPA monitoring products will be launched first on OceanSpaces rather than through MPA monitoring-focused symposia. There is broad agreement that OceanSpaces should house all products including videos of presentations, conference slides, technical reports, summary reports, etc. This provides this user group access to all relevant information, explore products, and understand data and report provenance. MARINe sees the what, when, and where as the most important benefits for this user group. In the MARINe database, 90% of their queries involve data availability at specific temporal and spatial scales. The State specifically agrees that this is important, particularly in the event of environmental disasters such as the Refugio oil spill. This broad applicability and value is especially important to demonstrate to state and federal decision-makers. This user group is likely trying to answer a specific question, and they should leave with an answer or a solidified next step.

Plan Requirement: *The system should easily show the what, when, and where for resource managers and decision-makers, and guide this user group to high-level reports and other products. The reports and products should also clearly link back to the data on which they are based.*

User Persona #3: Stakeholders

This group includes a broad range of users, including scientists who are not data contributors. This user group will likely interact with a map-based data discovery tool, and then access high-level reports or visualizations. They may also have a specific area or species of interest, and will download and analyze raw data based on what they see in the synthesis reports and visualizations. Thus, linking reports to the datasets on which they are based is important. This group should leave the system with a general understanding of who, what, where, and when; a sense of the scientific rigor behind monitoring results; and information they can take to an FGC meeting.

Honing in on Statewide Data & Metadata Standards

Data contributors

Existing and potential data contributors span a broad spectrum of researchers, including academic, government (e.g., federal, state, tribal), and citizen scientists. Participants agreed that crowdsourcing is an important potential data source that should be added as a sub-category of citizen science.

Data types and metadata standards

As stated in the current standards, EML should be used for ecological and socioeconomic data. Participants agreed that ISO is a good standard to adopt for geospatial data – described as data with a spatial component within each record. ISO is very broad, and overlaps with EML allowing for easy to develop crosswalks to convert from one metadata type to another. When determining which data standard to require for each type of data, we need to consider the industry standard, and thus what is most convenient for the data provider and data user. Standards for data produce through traditional knowledge are being developed through a parallel process.

Example: *PISCO has ecological data with geospatial grids. This would be considered both ecological and geospatial, and could be represented by EML in a single package or by ISO and EML if the ecological and spatial files are stored in separate records.*

Assets (imagery and video) do not currently have a linked metadata standard. Audubon core was suggested as the new standard for describing biodiversity multimedia resources. The USGS just finished processing data in different formats, which included camera tow sleds. Data have been integrated into same data portal provider as CeNCOOS. Further discussions around multimedia resources should include representatives from the USGS.

Plan Requirement: *EML will be the metadata standard, unless data have a geospatial component for each record, which requires ISO.*

Data contributors and the standards

One of our goals is to move away from project leaders providing data on hard drives or just creating a data library that houses data in flat files. Contracts for baseline monitoring in the Central, North Central, and South Coasts were very general because no data and metadata standards existed when the programs launched. In contrast, the contracts for baseline monitoring in the North Coast require data to be uploaded to OceanSpaces adhering to specific Data and Metadata Standards. These standards should remain consistent statewide moving forward, with small tweaks to account for lessons learned as monitoring continues. When possible, the standards should provide a cohesive statewide structure allowing for contribution to this system, and for performance evaluation statewide (keeping in mind data collection varies among projects, and will need further preparation for inter-region integration). Enforcing these standards and continually investing in data integration projects can help move us toward the goal of having dynamic visualizations. Participants generally agree on the benefits and difficulties of statewide analyses; however, stakeholders are not interested in statewide syntheses. Based on use metrics from the State of the California Central Coast report on OceanSpaces and the Reef Check data portal (NED), stakeholders are most interested in their local coastal area, and are not particularly interested in results above the regional scale.

Raw data and scripts

Participants agreed that within the data standards, there should be a plan for handling scripts associated with data transformations and analysis. These should be stored in a system that allows updating and for linking to the associated data tables. GitHub is a possible platform for storing code, and then linking from the data system to the laboratory's GitHub repository. However, based on State needs, this code should also live in a flat file that can be uploaded with the data package.

State-funded vs. non-state-funded data

The data system will be designed to differentiate between state-funded data, and supplemental and contextual data. This will likely require tiered data and metadata standards for non-state-funded data. State-funded data will remain the primary focus, but participants agreed that allowing others to contribute data is valuable as it increases participation and inclusiveness. However, these non-state-funded data need to meet minimum tier standards. The idea is to encourage participation, "say thank you", but ensure standards are met for the quality control of data displayed. As with citizen science groups Ocean Science Trust has worked with in the past, steps can be taken to help less-technology savvy groups to shape their data collection processes to ensure data are rigorous and will be displayed in the data management system. The State may also need to develop specific guidance for citizen science groups, outlining the quality of data and characteristics needed to be considered rigorous enough for integration with other long-term monitoring results that inform adaptive management.

Plan Requirement: *Investigate a tiered Data and Metadata Standards framework, acknowledging more rigorous requirements are possible for State-funded data than non State-funded data.*

EXAMPLE FUNDING SCENARIOS

Stated Goal: *Data management and development should benefit performance evaluation of the MPA network and inform ecosystem goals. It should focus on developing the requirements for a minimum viable product, with the possibility for phases of new functionality in the following years.*

Two budget scenarios were presented and there was broad agreement that scenario A would not get us to a minimum viable product. Scenario B, which started with $300,000 in year one and $200,000 in year two, would be much more realistic. From the State's point of view, the goal of evaluating the performance of the MPA network should drive investments in system development and data management. The State is committed to a flexible data management system that can adapt to new management and community priorities, nimbly implementing aspects of the plan over time. To this end, there is broad agreement that a phased approach will best serve this process. The plan will identify which components will be implemented in phase one, focusing on identified priorities that realize 20% of the functionality while providing 80% of the benefit.

To frame this discussion, it is important to keep in mind that the State already considers OceanSpaces a success – it fulfills the requirements for storing MPA monitoring data and results. The ability to store \ data and make it publicly accessible, while linking to relevant reports is the basic need. Other needs identified during this workshop go above and beyond, and thus are strictly meant to broaden the use and relevance of MPA monitoring data.

OST currently stores MPA monitoring data on a 30GB storage allocation as part of the webserver that hosts OceanSpaces. However, there are geospatial datasets that require larger, local storage, at about 250GB per region. These are stored on a network storage device. These data, due to size limitations, are presented to Ocean Science Trust on hard drives and stored on the NAS. Ideally, both the main body of MPA monitoring data, and the larger geospatial datasets will be readily available for download in one system. This could require up to 50 GB per year of data ingestion.

The line items on the budget scenarios were in accordance with participants expectations:

- Infrastructure, including storage, software, hardware, compute cycles, network, backups
- Personnel, including Program Manager (full-time), Data Manager (short-term hire, initially), Web Developer* (on-demand contract)
- Products, including map-based discovery, data visualizations (basic, advanced) – this component could result in significant cost increases, depending on the complexity, infrastructure selected, and the level of site utilization.

**To clarify, web development costs would focus on developing a display that would live within OceanSpaces and allow access to the data on a server, using a decoupled front-end to back-end model.*

Example: *KNB would host the database and server infrastructure. KNB is based on a decoupled REST API model. KNB would host the database, and a REST API based user interface would be designed to live within the structure of OceanSpaces; allowing for access specifically to MPA monitoring data.*

DATA DISCOVERY AND INTEGRATION

All of our primary users – data contributors, resource managers, decision makers, and stakeholders – have expressed interest in viewing and having access to supplemental and contextual data alongside the ecological and socioeconomic data produced by MPA monitoring. Participants broadly agreed that this would be helpful. However, additional discussion is warranted around how to display this information and what supplemental and contextual data to include.

Data Discovery

Care should be taken in formatting the sampling points for display on a map-based discovery tool. Displaying points, polygons, or transects can cause confusion based on spatial scale and user familiarity with monitoring or the system itself. Focusing on the principle goal of displaying who, what, where, and when, it is likely a first iteration of the tool would display where sampling data are available.

Example: *Google maps can aggregate a large number of points into different sized circles in relation to spatial scale. A larger circle indicates that more sampling occurs in the area. Considering the large number and density of monitoring sites along the California coast, the tool cannot clearly display all sites when viewing the entire coast or even an entire MPA region. A solution like this that aggregates sampling sites, but illustrates spatial sampling effort, may provide the best user experience.*

Participants agreed that users may also want the ability to draw a bounding box, which then displays all sampling done within that area. This can then be downloaded as a simple list of data available or the actual data packages can be downloaded in bulk. Filtering options (similar to what already exists on OceanSpaces) will also be necessary on the map and any additional search or discovery options. Filters based on the fields in EML metadata provide a good starting point. These would have to be a pre-determined and limited set to provide the best value and limit confusion.

Focusing on the question of “why would people come to OceanSpaces for MPA monitoring data?”, the clear answer should be “because it would be all in one place”. The real value in displaying the data all in one place is that these data are what we need to evaluate if and how MPAs are working. The most cost-effective first step is to focus on core MPA monitoring data along with any long-term data that exist. Once the data are easily discoverable, we can work to tie-in key questions, such as ‘are MPAs working’, who’s doing work, and how results contribute to our understanding of ecosystem health. Would build on what OceanSpaces is doing well now with covering projects. Baseline reporting for each of the regions required context, which could be a helpful starting point for additional data to include.

Plan Requirements: *Do not lose rich project information OceanSpaces has/doing well.*

Appendix D: Workshop Summary

Plan Requirement: *Data discovery will include State-funded and non State-funded data with differentiation through a filtering mechanism.*

Participants had specific ideas about types of supplemental and contextual data to display, and agreed that even getting PISCO, MARINe, and State-funded data listed would be a success. Other data types included that would accomplish having ecology and environmental context in one place are:

- Data from ecological monitoring of various ecosystems
    - Rocky intertidal
    - Rocky and soft-bottom subtidal (including kelp forests)
    - Sandy beaches
- Data on specific metrics or indicators, such as birds
- Environmental data
    - Chlorophyll
    - Sea Surface Temperature
    - Salinity
- Geospatial data
    - Benthic Habitat Maps
    - Locations of ASBSs
- Water quality monitoring data

Participants also stated that only displaying principle MPA monitoring data on the map would be best, and then link to supplemental and contextual data in a "more info" section once a user selects data of interest. Phase 1 involves laying the groundwork for and implementing links to other data portals. Phase 2 could then incorporate a display of what is available from the different data sources. PISCO data could prove to be a good pilot dataset with which to test the new system. Differentiation between funded or non-funded data could be indicated by a logo or image.

Participants suggested a variety of data discovery mechanisms that should be considered:

- Map
- Thematic
- Data Contributor
- Temporal
- Spatial
- Taxonomic
- Ecosystem
- Keyword

Suggested Resource: https://search.dataone.org

Plan Requirement: *Phase 1 – Ingest data and make discoverable via a map. Phase 2 – Ingest, display, or link to supplemental and contextual data. Phase 3 – Visualize MPA monitoring data (possibly with supplemental and contextual data).*

Further discussions with data portal owners, such as SiMON and CeNCOOS will be needed.

Data Visualizations

A special meeting of advisory team members held in August explored the range of visualization options, from discovering data on a map to visualizing results in graphs or infographics.

Stated Goal: *Visualizations should initially focus on State-funded MPA monitoring data.*

From the State's point of view, any visualization tools would be built for citizen scientists, stakeholders, and fishermen, who want to look at their area or species and see a trend through time without judgment on good or bad. Other participants agreed that users would generate simple graphs to summarize data and report to commissioners or managers. We could provide static graphs and figures that exist in technical reports, which directly address metrics and questions in the MPA monitoring plan. However, making these graphs interactive was not widely seen as a good starting point. Consensus grew around options like finding the top ten species of interest and creating basic graphs about species abundance. Creating a value judgment or deeper analysis prior to the ten-year MPA management review should be avoided. Some participants suggest that because trend graphs may be the most useful for decision-making, these should be a priority.

Once a framework for these visualizations is in place, it is easier to use it for other plots. One important consideration is where data is pulled from for these plots, and is it updated regularly. These considerations will be minimal for the principle MPA monitoring data, as it will all be housed internally. For supplemental data in the future, the infrastructure will have to account for visualization requirements, different data schemas, and updating data over time.

The implementation of visualizations will likely require a phased approach that is blocked by the requirements of transforming datasets to share a common schema. Phase 2 would likely include bringing in MPA monitoring partner data, as it will already be in a compatible format. The success of visualizations can help increase buy-in from data sources outside of monitoring partners. The value add of having data visualized can encourage participation and voluntary restructuring of data. Further discussions with the advisory team and technical experts will be needed to finalize the requirements for this topic.

OTHER

Technological Requirements

Previous advisory team conversations coupled with the results of this workshop will help us create the technology requirements for this system.

Participants were also initially interested in the selection of a model for this plan. The resulting discussion highlighted that we see the need to provide a hybrid live data storage model, that ensures the longevity and durability of MPA monitoring data for the State, and allows for advanced interactions with data, and linking with other data portals without duplicating data. The storage infrastructure will also need to be able to differentiate between State-funded MPA monitoring data, and other supplemental or contextual data stored in the system.

The contractual requirements for data contributors will need to be in place to ensure proper schema and data formatting to allow integration into our system. The Statewide Data and Metadata Standards will be in place for the North Coast and for long-term monitoring beginning in the Central Coast.

Advisory Team Members

- Tanya Haddad, West Coast Governor's Alliance, Oregon Department of Land Conservation & Development
- Matt Jones, National Center for Ecological Analysis & Synthesis
- Steve Lonhart, Monterey Bay National Marine Sanctuary, NOAA
- Jennifer Patterson, Central & Northern California Ocean Observing System
- Paulo Serpa, California Department of Fish & Wildlife

Other Experts

- Carol Blanchette, University of California Santa Barbara
- Mark Carr, Partnership for Interdisciplinary Studies of Coastal Oceans, University of California Santa Cruz
- Cyndi Dawson, California Ocean Protection Council
- Jan Freiwald, Reef Check California
- Rani Gaddam, University of California Santa Cruz
- Lisa Gilbane, Multi-agency Rocky Intertidal Network, Bureau of Ocean & Energy Management
- Dan Malone, University of California Santa Cruz
- Pete Raimondi, Multi-agency Rocky Intertidal Network, University of California Santa Cruz

Ocean Science Trust Staff

- Benét Duncan, Associate Scientist
- Erin Meyer, Program Manager
- Jim Wicker, Program Manager

Appendix A: Outline of Data Management Plan

I. Introduction

A. Need

B. Vision/Goals

C. User Needs

D. Planning

II. Overview of MPA monitoring Data

III. Data and Metadata Standards

A. Statewide Standards

B. Developing metadata crosswalks

IV. Data System Architecture

A. Proposed Data Workflow

B. Discovery and Information Discovery

C. Visualizing and Communicating Results

V. Implementing this Plan

A. Funding Requirements

B. Personnel Requirements

VI. Adapting this Plan

Data and Information Management

Visualizations Meeting Summary

MBARI, 7700 Sandholdt Rd, Moss Landing, CA 95039
27 August 2015

MEETING OBJECTIVES

- Engage with a subset of the Advisory Team to explore and define what data visualizations are, and learn from their experience creating visualizations for different types of data and audiences.
- Evaluate the primary user groups (for MPA monitoring data and results) that will utilize visualizations and what types of information would be most valuable to them.
- Explore the technology requirements and costs of visualizing MPA monitoring data.

MEETING OVERVIEW

This meeting is one component of a larger process for developing a Data Management Plan for California's MPA monitoring data. As we turn our attention to long-term monitoring, we need to develop a comprehensive, strategic and feasible plan that articulates an approach that leverages existing public and private technology solutions, and serves as a roadmap for building partnerships and data infrastructure in the future.

During this meeting, Ocean Science Trust met with a subset of the Advisory Team to discuss several of the main components of and key concepts around data visualization:

- What are visualizations
- User stories
- Visualizations goals and priorities
- Technology
- Funding requirements

MEETING SUMMARY

What do we mean by data visualizations?

Deciding what we wanted to include in the category of data visualizations provided the starting point for our discussion. It was clear from the beginning that the label "data visualizations" can have multiple meaning to different groups of people. For our purposes, we decided to define them in two functional categories:

- ***Map-based data discovery:*** This is a fundamental way of enumerating the data available for a specific geographic area. Visualizing sampling areas, points, or other locations on a map can be coupled with thematic, temporal, or taxonomic filters. The goal of this is exclusively to discover what data are available and where. Participants agreed that developing and implementing map-based data discovery is the only requirement of meet the need for data visualizations of MPA monitoring data.
- ***Data visualizations:*** This category of visualizations includes displaying measurements in a visual form (e.g., heat map of fishing pressure, sea surface temperatures, trend graph of fish counts). The goal of this is to communicate specific, management-relevant trends and/or data of most interest to targeted user groups. Participants agreed that this category will require a significant level of funding to accomplish and should be limited by the realistic uses of visuals for the data types available.

What to visualize?

Participants initially wanted to engage on the topic of who would create the visualizations and a clear message about aligning user needs with what the data and system can offer. As part of the MPA monitoring program, data providers are required to submit deliverables, like technical reports, to fulfill their contractual obligations. Participants suggested that data providers are the best suited to create visualizations and understand the limits of their data. Data providers can be tasked with creating specific outputs, either datasets prepped for data visualization, or the visualizations themselves. If the data provider does not create visualizations, then close collaboration with the data provider will be necessary. In addition to understanding the source for visualizations, we need to ensure we have well-developed questions that are reasonable to ask.

It was suggested that additional deliverables be added to contracts with those conducting monitoring to include data conditioning or visualizations. For example, data integration and synthesis is included as separate project within the South Coast MPA Baseline Program. It is important to examine user needs and find alignments, where possible, with the data products already paid for within the MPA monitoring program. Some visualizations could be adapted from technical reports already created. Interactive graphs can be synthesized from static graphs in technical reports if that new presentation of data would address a user need.

It is clear that understanding how to create useful visualizations from MPA monitoring data will require collaboration between the data providers and Ocean Science Trust to align user needs with data provider created products. Besides project-to-project questions, the MPA monitoring program has high-level questions aimed at assessing the effectiveness of the MPA network. These questions will need to be addressed to inform the official MPA network management reviews, which occur every ten years.

User Groups and Goals

Based on previous audience prioritization exercises conducted during the creation of the MPA Monitoring Enterprise User Needs Assessment and the development of OceanSpaces, we discussed the target audiences for data visualizations. The participants agreed that visualizations would need to focus on a specific user group to be successful. They also agreed that highly complex visualizations are not necessary.

We presented this list of users:

- Decision-makers
- Resource Managers
- Policy Informers/Influencers
- Scientists
- Information Managers
- Educators
- Stakeholders
- Students
- Citizens

Participants agreed the user group most likely to utilize visualizations would be in the combined resource manager/decision-maker group. This group would likely want summary level information to help answer a question. They would likely want to know the what, where, and when of monitoring (utilizing map-based discovery) and then get an initial understanding of the trends related to a particular MPA or species. They would then likely download the data and conduct further analyses themselves, or contact the data provider with additional questions. For this user group, simple trend graphs or inside MPA vs. outside MPA type visualizations would provide an appropriate level of information.

Regardless of the visualization created they should align closely with questions related to MPA effectiveness and user needs. Initial visualizations should remain simple, and benefit to the target user group should be tracked.

Technology and Funding

Due to the complexity of visualizations as a topic and the limited time for this meeting, there was not time to delve into specific technologies and funding scenarios. It was clear from the comments made throughout the discussion that visualizations represent a much larger cost to develop than other basic data management tasks. And due to the complexity and non-standard nature of MPA monitoring data from project-to-project, significant planning and funding would be required to make useful visualizations. It was also clear that the basic data management architecture would need to be prioritized, with visualization functionality being added in a phased approach. Based on these concerns, we allotted a specific line item in the data management plan budget for visualization technology development, and added this discussion topic to the data management workshop.

MEETING PARTICIPANTS

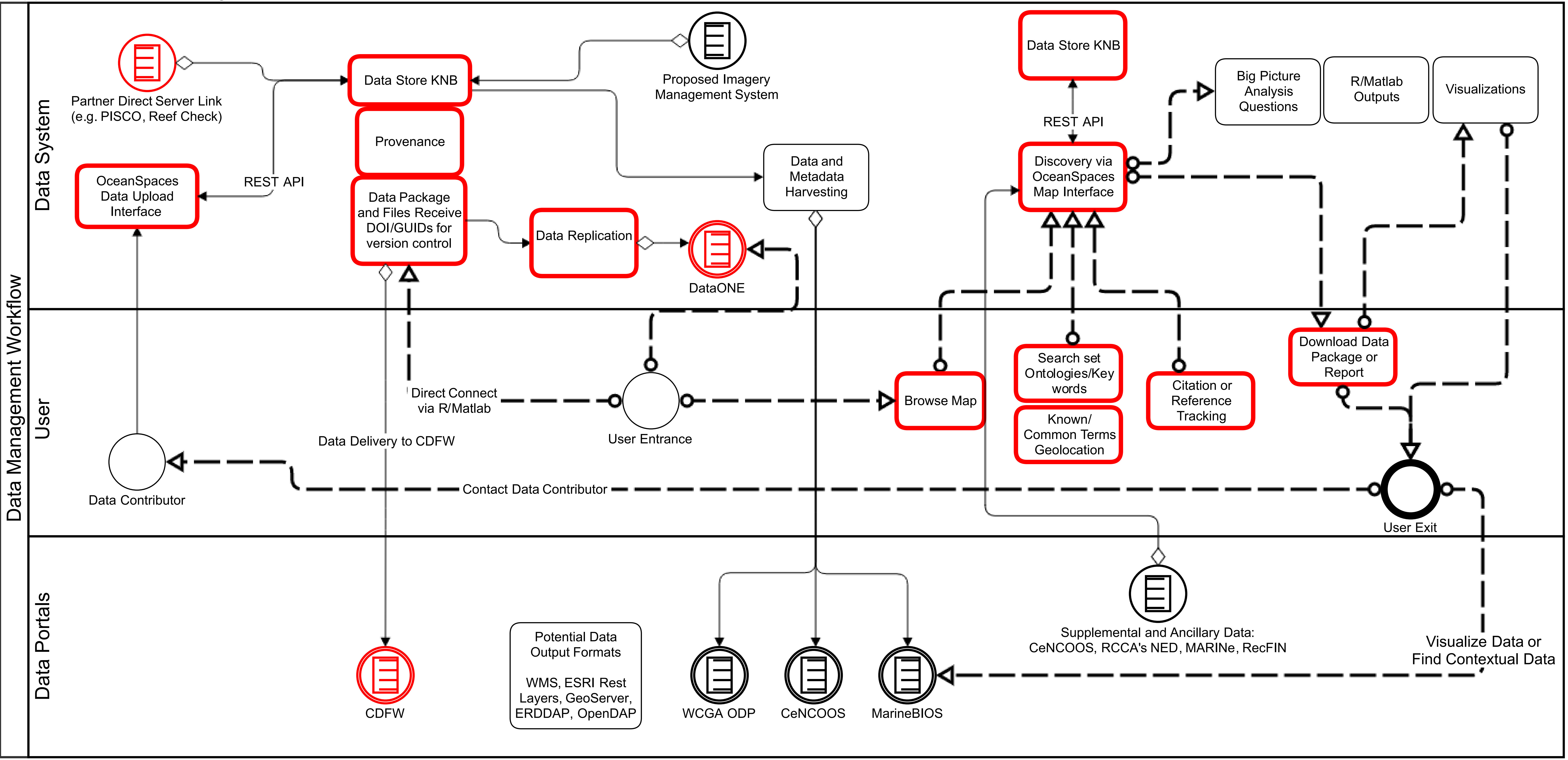
Advisory Team Members

- Steve Lonhart, Monterey Bay National Marine Sanctuary, NOAA
- Jennifer Patterson, Central & Northern California Ocean Observing System
- Paulo Serpa, California Department of Fish & Wildlife

Ocean Science Trust Staff

- Erin Meyer, Program Manager
- Jim Wicker, Program Manager

Appendix F: Data Management Workflow



Appendix G: Organization Conversation Log

Date	Organization
10/21/2014	Partnerships for Interdisciplinary Studies of Coastal Oceans/Oregon State University
10/20/2014	Partnerships for Interdisciplinary Studies of Coastal Oceans/Oregon State University
	Partnerships for Interdisciplinary Studies of Coastal Oceans/University of California Santa Barbara
	Partnerships for Interdisciplinary Studies of Coastal Oceans/University of California Santa Cruz
	Partnerships for Interdisciplinary Studies of Coastal Oceans/University of California Santa Barbara
	Partnerships for Interdisciplinary Studies of Coastal Oceans/University of California Santa Cruz
11/3/2014	West Coast Governors Alliance/West Coast Ocean Data Portal
	Central & Northern California Ocean Observing System
	Bureau of Ocean Energy Management
	SoundGIS
11/14/2014	Bureau of Ocean Energy Management/Multi-Agency Rocky Intertidal Network
1/13/2015	DataONE
	Long Term Ecological Research Network/DataONE
	Oak Ridge National Laboratory/University of Tennessee, Knoxville/DataONE
3/6/2015	West Coast Governors Alliance/West Coast Ocean Data Portal
4/2/2015	California Water Quality Monitoring Council
4/8/2015	Partnerships for Interdisciplinary Studies of Coastal Oceans/Oregon State University
4/12/2015	Partnerships for Interdisciplinary Studies of Coastal Oceans/University of California Santa Barbara
4/14/2015	National Center for Ecological Analysis and Synthesis/University of California Santa Barbara
4/14/2015	Partnerships for Interdisciplinary Studies of Coastal Oceans/University of California Santa Barbara
4/23/2015	Multi-Agency Rocky Intertidal Network/University of California Santa Cruz
4/23/2015	West Coast Governors Alliance/Oregon Department of Land Conservation and Development
4/30/2015	California Department of Fish and Wildlife
	SoundGIS
5/20/0205	Monterey Bay Aquarium Research Institute
5/20/2015	Central & Northern California Ocean Observing System
5/22/2015	National Oceanographic and Atmospheric Administration/Monterey Bay National Marine Sanctuary
6/3/2015	The Nature Conservancy
6/8/2015	National Oceanic and Atmospheric Administration/Monterey Bay National Marine Sanctuary
6/16/2015	California State University Monterey Bay
8/12/2015	Partnerships for Interdisciplinary Studies of Coastal Oceans/Oregon State University
9/22/2015	Partnerships for Interdisciplinary Studies of Coastal Oceans/University of California Santa Cruz
12/21/2015	California State University Monterey Bay

This log is a record of the scoping conversations for the Data Management Plan. It includes conversations leading to the formation of the Advisory Team, and those had in addition to official Advisory Team calls.

March 2016