Data Management and Communications Plan for Research and Operational Integrated Ocean Observing Systems

I. Interoperable Data Discovery, Access, and Archive

Part I: Overview

May 10, 2004

Contents

Preface

This Plan is the first in a series of documents that will address the Data Management and Communication (DMAC) requirements of the Integrated Ocean Observing System (IOOS), and other regional, national, and global observing systems. In the Preface, we provide some background information to assist the reader in placing the DMAC Subsystem into its proper context as a component of the larger IOOS. As such, the DMAC Subsystem will be developed, implemented, operated, and enhanced through the planning and governance structures described in the IOOS Implementation Plan.

IOOS

There is strong support in the U.S. Congress, the Executive Branch, and the U.S. Commission on Ocean Policy (USCOP) for development of a sustained, integrated coastal and ocean observing system that will make better use of existing resources, new knowledge, and advances in technology to achieve the following seven related societal goals:

- Improve weather forecasts and predictions of climate change;
- Improve safety and efficiency of marine operations;
- Provide more timely predictions of natural hazards and their impacts;
- Improve national security;
- Reduce public health risks;
- Sustain, protect and restore healthy marine and estuarine ecosystems; and
- Sustain, protect and restore marine resources.

Congress directed the U.S. marine environmental communities to come together to plan, design, and implement this observing system. The National Oceanographic Partnership Program[1] (NOPP) established the Ocean.US Office through a Memorandum of Agreement (MOA) in 2000. Ocean.US is charged with coordinating the development of the U.S. Integrated Ocean Observing System, based on concepts developed by national and international experts over the past dozen years. Ocean.US is overseen by an Executive Committee (ExCom) composed of representatives from those NOPP agencies that signed the MOA.

Ocean.US is preparing an IOOS Implementation Plan (www.Ocean.US), in cooperation with the participating NOPP agencies. The Plan is being developed in three parts: Part I—Structure and Governance; Part II—The Initial IOOS: Building on Existing Assets; and Part III—Improving the

[1]NOPP was established by Congress in 1997 (P.L. 104-201) to (1) "promote the national goals of assuring national security, advancing economic development, protecting the quality of life, and strengthening the science and education through improved knowledge of the ocean" and (2) "coordinate and strengthen oceanographic efforts to achieve these goals by 'identifying and carrying out partnerships among Federal; agencies, academia, industry, and other members of the oceanographic community in areas of data, resources, education, and communications."

IOOS – Enhancements and New Initiatives. Approval of all three parts of the IOOS Plan by the NOPP National Ocean Research Leadership Council (NORLC) is anticipated in the summer of 2004.

IOOS is envisioned as a coordinated national and international network of observations, data management, and analyses systems that rapidly and systematically acquires and disseminates marine environmental data and information on past, present, and future states of the oceans. The IOOS is being developed as two closely coordinated global and coastal components that encompass the broad range of scales required to assess, detect, and predict the effects of global climate change, weather, and human activities. The global component consists of an international partnership to improve forecasts and assessments of weather, climate, ocean state, and boundary conditions for regional observing systems. It is the U.S. contribution to the Global Ocean Observing System (GOOS). The coastal component blends national observations in the Exclusive Economic Zone (EEZ) with measurement networks that are managed regionally to improve assessments and predictions of the effects of weather, climate, and human activities on the state of the coastal ocean, its ecosystems and living resources, and the nation's economy. The coastal component encompasses the nation's EEZ, the Great Lakes, and the estuaries.

Existing and planned observing system elements that address both research and operational aspects of the seven IOOS goals will be integrated into the system. Evolution of an integrated system that is responsive to user needs will require an iterative process of selection, incorporation, evaluation, and improvement over time. Candidate technologies and capabilities may pass through a series of stages (research, pilot, pre-operational) prior to being incorporated into the operational IOOS, long-term research, or both. Detailed criteria for activities to successfully pass through each of these stages are presented in the IOOS Implementation Plan, Parts I and II.

A four-year cycle of planning, programming, and budgeting for IOOS implementation and development is described in Part I of the IOOS Plan. Ocean.US, in cooperation with NOPP agencies and Regional Associations (RAs)[2], will specify priorities for implementation and advancement of IOOS; formulate timetables; work within the Federal budget process to determine costs; and capitalize on unplanned opportunities. Research and development projects may be funded competitively through the NOPP process, or through mechanisms established by individual agencies in cooperation with Ocean.US. Operational elements are funded for extended periods of time based on demonstrated utility and performance.

[2]RAs will be established, based on regional priorities, to design, implement, operate, and improve regional observing systems by increasing the resolution of the variables measured,; supplementing the variables measured by the national backbone with additional variables,; providing data and information tailored to the requirements of regional stakeholders,; and implementing programs to improve public awareness and education. Regional observing systems are needed to provide data and information on phenomena that are more effectively detected or predicted on regional scales that go beyond the jurisdiction of individual states.

DMAC

For planning purposes, IOOS is considered to be composed of three subsystems: the OBSERVING SUBSYSTEM (**collection of remotely sensed and *in situ* environmental measurements and their transmission from platforms - telemetry**); the DATA MANAGEMENT AND COMMUNICATIONS SUBSYSTEM (**DMAC – discovery and delivery of data within IOOS, and interoperability with other relevant observing systems**); and the MODELING AND ANALYSIS SUBSYSTEM (**evaluation and prediction of the state of the marine environment**).

Central to the success of IOOS (and other regional, national, and international ocean and coastal observing systems) is the presence of a DMAC Subsystem capable of supporting the wide variety and large volumes of data, the reliability and integrity requirements of operational data delivery, and the many other needs of the IOOS user community. The DMAC Subsystem is the primary integrating element of IOOS, and it will provide the linkages among other IOOS components, partner organizations, and systems in other disciplines (e.g., terrestrial, atmospheric). Because of the critical need for a basic data communications infrastructure to support existing and newly emerging IOOS observing systems, the DMAC effort was initiated early on in the IOOS planning process. In the spring of 2002, the Director of Ocean.US appointed the Data Management and Communication Steering Committee (DMAC-SC), including representatives from Federal and state government agencies, academia, and the private sector. The DMAC-SC was tasked with developing a detailed phased implementation plan for this IOOS subsystem.

The DMAC Plan, this document, presents a coherent strategy for integrating marine data streams across disciplines, organizations, times scales, and geographic locations. It has been divided into three main parts: Part I provides an overview of the requirements and technological considerations, and the strategies for addressing them. Part II presents the detailed DMAC System Implementation Plan in outline form. Part III, the Appendices, provides in-depth discussion of key technical topics.

The DMAC Plan focuses on enhancing the interoperability of existing IOOS components through development of a common Data Communications Infrastructure. The infrastructure will consist of standards and protocols for metadata, data discovery, transport, on-line browse, and long-term archive. Other important issues such as QA/QC, modeling and applications, security, data assembly, and telemetry will be addressed in future IOOS documents. It should be noted that this document is a plan, not a specification. The cost model presented herein includes support for systems engineering services to conduct a formal design analysis leading to a formal specification. The specification will guide planning and implementation decisions such as application of resources.

The DMAC Subsystem is an integral part of IOOS, and it is being developed in close coordination with other IOOS components. The longer-term implementation priorities and recommendations articulated in the DMAC Plan are being incorporated into the overall IOOS planning and budgeting processes described in Parts I and II of the IOOS Implementation Plan. The IOOS planning and budgeting processes will not have a direct influence on the Federal budget process until FY 2007. Annual updates to the DMAC Plan, developed in collaboration with the participating NOPP agencies and RAs, and with recommendations from a new DMAC- Standing Committee, will feed into future planning and budgeting cycles beyond FY 2007. In the interim time period (FY 2004-2006), efforts are being made to obtain support from participating NOPP agencies, RA's, and other sources of opportunity for the immediate, shorter-term priorities of the DMAC Subsystem. Implementation of these immediate priorities will lead to development of an initial DMAC Subsystem to support existing and emerging observing system activities at the local, regional, and national levels.

DMAC Plan Review

The DMAC Plan has undergone several levels of review, and new drafts were produced to address the comments received at each stage. The process was as follows:

- February-March, 2003 – Internal review by the DMAC-SC and members of the six supporting teams[3];
- April-May, 2003 – External review by over a dozen national and international technical and scientific experts;
- June 2003 – Public review by over sixty policy makers and technical experts who participated in a national workshop sponsored by Gulf of Maine Ocean Observing System and Ocean.US in Portland Maine; and
- September-November, 2003 – Formal Public review process: The draft DMAC Plan was posted on the Ocean.US/DMAC Web site and announcements of its availability appeared in several marine community newsletters. E-mail notifications were also sent to several hundred members of the marine community, including participants in the 2002 Ocean.US Community Workshop (Airlie House); the March 2003 Ocean.US Regional Summit; the National Ocean Research Leadership Council, the U.S. GOOS Steering Committee, and the ExCom.

DMAC-SC members and Ocean.US staff presented numerous briefings on the Plan at regional, national, and international conferences and meetings, including those sponsored by: AGU, ASLO, AMS, JCOMM-ETDMP, IODE, NSF Cyber-Infrastructure and ORION, CAOS, GoMOOS, NVODS, and GODAE. They have also provided briefings to staff within their own agencies and organizations.

[3]The DMAC-SC was supported in their work by six teams Data Discovery and Metadata; Data Transport, Data Archive and Access; Applications and Products; Data Facilities; and User Outreach.

Section 1. Overview

INTRODUCTION

At the present time, there is no coherent data management and communications strategy for effectively integrating the wide variety of complex marine environmental measurements and observations across disciplines, institutions, and temporal and spatial scales. As a result, U.S. society is denied important benefits that might otherwise be derived from these data, such as improved climate forecasts and more effective protection of coastal marine ecosystems. Data are obtained by diverse means—nets are dragged; traps are set; instruments are lowered from ships, set adrift, or moored on cables and platforms; satellites scan the oceans from space; and laboratories are constructed on the seafloor. Measurements are made for a wide variety of purposes by individuals and sensors supported by many different kinds of institutions, including private industry, Federal, state, and local government, and non-governmental organizations. These data come in many different forms, from a single variable measured at a single point (e.g., a species name) to multivariate, four-dimensional collections of data that may be millions of gigabytes in size. These considerations, among others, led Congress to direct the U.S. marine data communities to come together to plan, design, and implement a sustained Integrated Ocean Observing System (IOOS).

Central to the success of IOOS, and other regional, national, and international ocean and coastal observing systems, is the presence of a Data Management and Communications (DMAC) Subsystem capable of delivering real-time and delayed-mode observations to modeling centers; model-generated forecasts to users; distributed biological measurements to scientists, educators, and planners; and all forms of data to and from secure archive facilities. The needs of end users must be a part of the implementation and operation of the subsystem, both as sources of specifications for subsystem design, and as agents of change to keep the delivery of products from IOOS relevant to national interests. At a minimum, the DMAC Subsystem will make data and products readily accessible, allow users to readily locate data and information products, and advise users on the specifications and limitations of data by providing essential metadata (descriptive information about the data) along with the data.

The information technology required to meet most of the needs of DMAC, while challenging, can be developed from existing capabilities through relatively straightforward software engineering. The greatest challenge facing DMAC is one of coordination and cooperation among IOOS partners and user communities. DMAC can succeed only if the participants actively use the data and metadata standards, communications protocols, software, and policies that will knit IOOS into an integrated whole. The creation of a successful IOOS DMAC will require a sustained effort, a commitment across the U.S. marine community, and continual coordination with our international counterparts.

THE VISION

IOOS is envisioned as a system of regional, national, and global elements that rapidly and systematically acquire and disseminate data and data products to serve the needs of government agencies, industries, scientists, educators, non-governmental organizations, and the public. The IOOS vision is one of *cooperative integration.* The member entities will continue in the independent pursuit of their missions, while participating in a well-ordered data and information infrastructure. If IOOS were a living being, the Data Management and Communications Subsystem would be its blood and circulatory system—the data used to produce the information products needed by IOOS are analogous to the oxygen and nutrients transported in the blood to feed the many highly specialized organs.

The following set of guiding principles addresses the DMAC vision:

Interoperability: DMAC will serve as a framework for interoperability among heterogeneous cooperating systems.[4] The cooperating systems will be free to evolve independently to address the needs of their target users. Software and standards needed to participate in DMAC will be available directly to partners, or provided through commercial and non-commercial sources. DMAC will also be interoperable with systems outside of the marine community that manage atmospheric and terrestrial data.

Open, easy access and discovery: DMAC will enable users from all over the globe to easily locate, access, and use the diverse distributed forms of marine data and their associated metadata and documentation in a variety of computer applications (e.g., Geographic Information Systems, GIS, and scientific analysis applications). Users will be unencumbered by traditional barriers such as data formats, volumes, and distributed locations. DMAC will integrate cooperating systems so that data discovery will be seamless, and multiple versions will be easily tracked. There will be a "free market" of ocean sciences information, including officially sanctioned IOOS data sets, as well as data and products from other sources.

Reliable, sustained, efficient operations: DMAC will provide high reliability with 24/7 delivery of real-time data streams from measurement subsystems to operational modeling centers and users with time-critical requirements. It will provide high reliability in the delivery of com-

[4]By "interoperable" we mean that systems can function cooperatively through seamless exchange of data, the blending of data derived by different methods/instruments, an ability for models to access all varieties of source data without detailed knowledge of its origins, or the ability of users to see derived information in a way that is not limited by knowledge of the data collection methods or processing.

puter-generated forecasts, estimates of state, and delayed-mode and real-time data to end users. DMAC will require sufficient bandwidth and adequate carrying capacity to support large exchanges of raw data and model outputs among high-volume users. DMAC will offer techniques that reduce the need for large data transfers, such as server-side subsetting and computation, to allow users with limited bandwidth to enjoy the benefits of IOOS. Feedback mechanisms will be built into the technical design of DMAC to ensure that problems are detected and rapidly addressed.

Effective user feedback: IOOS will provide a continuous, vigorous outreach process addressing all levels of users of marine data, emphasizing the benefits of participation in IOOS/DMAC, and helping to identify and remedy difficulties encountered by those who are participating. In addition, this process will identify and address changing user requirements that drive the development and growth of DMAC.

Open design and standards process: DMAC will commit to an open software design. All standards and protocol definitions will be openly published so that participating organizations may create functioning DMAC components based on these specifications. The standards development process will be open and inclusive, so that it fosters buy-in by all stakeholders. Existing information technology and scientific standards will be used in preference to development of new solutions, whenever possible. The standards and protocols will be of sufficient breadth and quality to guarantee interoperability of all observations and products. Institutions participating in IOOS will ensure that the data they contribute comply with these standards and protocols.

Preservation of data and products: Irreplaceable observations, data products of lasting value, and associated metadata will be archived for posterity in an efficient and automated manner.

CHALLENGES

The DMAC design faces a trio of competing characteristics:

1. **Loosely federated organizations:** No top-down corporate management structure exists to effectively manage major shifts in data management strategy (and the resulting dislocations) in order to achieve interoperability.

2. **Physically distributed repositories:** Data must reside and be managed at many distinct locations (some of which contain vast volumes of data).

3. Heterogeneous data: Classes of data range from huge satellite track records, to multi-dimensional model outputs, to Lagrangian drifters, to polygonal geographic regions and point measurements. Variables are from diverse disciplines and are unevenly distributed in time and space.

Though challenging, the technical requirements for DMAC that are imposed by characteristics 2 and 3, alone, could be addressed by relatively straightforward software engineering. Solutions that do not adequately embrace the loosely federated structure of IOOS, however, cannot succeed. Community building considerations must be central in the design of the DMAC Subsystem.

COMMUNITY BUILDING

For DMAC to succeed it must achieve acceptance and recognition by marine data provider and data user communities. Only "gentle" (non-coercive) tools of persuasion will be effective within the loosely federated structure of IOOS. Individuals working in pursuit of their organization-specific goals must perceive that participation in DMAC will lead to a net gain toward achieving those goals. Thus, the greatest challenges for initial acceptance of DMAC and for subsequent growth in its usage are in the areas of community outreach and organizational behavior (the factors that enable a community to agree upon and use standards) rather than in technology.

Organizational challenges exist at both management and technical levels of the system. The leadership of existing marine organizations and programs must understand that access to the benefits that will accrue from IOOS depends upon their willingness to commit their organizations to the development and use of the DMAC Subsystem. At times this may result in short-term inconveniences to their organizations. Technical staff involved in the development of information management systems will need to ensure that their systems conform to the interoperability standards set by DMAC. This may sometimes entail duplication of software functionality that is available through in-house systems.

INTERNATIONAL COOPERATION

IOOS is a component of the Global Ocean Observing System (GOOS). Integrating global data from GOOS (i.e., blending U.S. and international observations to produce global data products) is essential for IOOS to be able to generate high-quality assessments and forecasts of U.S. coastal conditions. Producing global assessments of coastal ecosystem health and sea level change effects, for example, will also require that IOOS observations and data products be fully integrated into GOOS along with other national and international ocean observing system efforts. These requirements for data integration must be reflected in a very high level of coordination between U.S. DMAC activities and related international data management and communications activities.

Data Management Needs For Research Observatory Networks

The National Science Foundation's Ocean Research Interactive Observatory Networks (ORION) program will establish a research observing network (the Ocean Observatory Initiative, OOI) as well as develop instrumentation and mobile platforms needed to enable the broad range of research envisioned for this system. This integrated observatory network will provide the oceanographic research and education communities with a new mode of access to the ocean. The network has three elements: (1) a regional cabled network consisting of interconnected sites on the seafloor spanning several geological and oceanographic features and processes, (2) relocatable deep-sea buoys that could also be deployed in harsh environments such as the Southern Ocean, and (3) construction or enhancements to existing facilities leading to an expanded network of coastal observatories. The scientific problems driving the need for this infrastructure encompass nearly every area of ocean science and will provide earth and ocean scientists with the opportunity to study multiple, interrelated processes over time scales ranging from seconds to decades; to conduct comparative studies of regional processes and spatial characteristics; and to map whole-Earth and basin scale structures.

In addition to the ORION program there are a number of observing networks funded by or proposed to NSF whose goals are to investigate Earth structure (EarthScope), terrestrial ecology (NEON), hydrology (CUAHSI), environmental remediation (CLEANER), and atmospheric structure (COSMIC). To fully realize the benefits of these observing networks, it is important that data systems established for each program are interoperable. This interoperability will facilitate investigations leading towards a better understanding of the entire Earth system.

Observing networks established with basic research and education as the primary mandates share many of the same data management and communication requirements that drive the design of the IOOS DMAC, but will impose significant further constraints. To keep pace with research the system must have sufficient flexibility to evolve in response to changing and often innovative data collection techniques. It must address rapidly evolving user requirements in the classroom, in the field, in the laboratory, and in the analysis and synthesis of data. Research demands will continually stretch the capabilities of the data system to ingest diverse data types at varying frequencies. To meet the needs of the research and education communities, such a data management and communications system should be smoothly embedded in an information infrastructure that simplifies the task of finding and using data from multiple sources, and that facilitates collaborative work. The research infrastructure must also include analysis and visualization tools, a range of numerical models and model-data fusion tools, workflow tools, and one or more digital libraries. This infrastructure will be integrated with a computational grid to provide the necessary resources to support analysis, visualization and modeling tasks.

As an information infrastructure suitable for ocean science evolves, a number of specific elements are likely to be required. Many of these can be most effectively implemented only through close coordination with other agencies and with the efforts to establish a US IOOS. In some cases, coordination with international efforts is also required. NSF expects to receive specific advice and recommendations on the requirements of this information infrastructure, from the ocean science community and from information technology experts, as part of the OOI implementation design phase.

One of these coordination efforts must be between the DMAC-SC and the International Oceanographic Data and Information Exchange (IODE) (http://www.iode.org). Established in 1961, the IODE seeks to enhance marine research, exploitation, and development among its 60 member countries by facilitating the exchange of oceanographic data and information. IODE coordinates activities among a worldwide service-oriented network of global, regional, and national data centers.

DMAC will also coordinate its activities at the international level with the Data Management Coordination Group of JCOMM—the Joint WMO (World Meteorological Organization) /IOC (Intergovernmental Oceanographic Commission) Technical Commission for Oceanography and Marine Meteorology. The WMO and IOC established the JCOMM (http://www.jcomm.net) in 1999 as an intergovernmental body of technical experts, with a mandate to prepare both regulatory and guidance material for its member states relating to marine observing systems, data management, and services. IODE and JCOMM are collaborating on the development of an Oceans Information Technology (OIT) Pilot Project (http://www.oceans-it.net), which closely resembles (and follows by about nine months) the evolution of DMAC. The DMAC-SC has initiated, and must maintain, very close cooperation and interchange with the OIT participants as both IOOS and OIT are implemented.

Section 2. Technical Analysis

IOOS DATA COMMUNICATIONS

The relationship among the DMAC Subsystem and other IOOS components and partners is depicted in Figure 1. Data flow within IOOS begins with the Observing Subsystem. Raw measurements from the Observing Subsystem elements are processed at various Primary Data Assembly and Quality Control sites to make them available to the uniform DMAC Data Communications Infrastructure through a DMAC Data Entry Point. Every IOOS data stream must have a DMAC Data Entry Point. The infrastructure consists of standards and protocols to support: (i) IOOS-wide descriptions of data sets (Metadata); (ii) the ability to search for and find data sets of interest (Data Discovery); (iii) the ability to access the data in an interoperable manner from client applications (Data Transport); (iv) the ability to evaluate the character of the data through common Web Browsers (On-line Browse); and (v) the ability to securely archive data and metadata and retrieve them on demand (Data Archive).

The DMAC Data Communications Infrastructure provides access to IOOS data for all IOOS components and partners. Bi-directional communications will exist between independent data management systems both internal and external to the IOOS framework—data management systems from regional and international entities and from distinct disciplines such as meteorology. The DMAC Data Communications Infrastructure also conveys data, metadata and data products to users' applications (programs) and to those entities both inside and outside of IOOS who generate value-added information products. The information products and data address the interests of U.S. society through the advancement of the seven IOOS goals (Preface). Ultimately, it is the needs of U.S. users that guide the selection of new measurements, infrastructure, procedures, and products through IOOS User Outreach mechanisms.

The DMAC Plan (this document) offers a detailed, phased implementation strategy specifically for the development of the DMAC Subsystem Data Communications Infrastructure and Archive.

THE OBSERVING SUBSYSTEM AND PRIMARY DATA ASSEMBLY/QUALITY CONTROL

IOOS Observing Subsystem elements are managed by regional, national, and international entities. The measurements are highly heterogeneous, originating from surveys (e.g., fish stock assessments), cruise measurements, laboratory measurements, satellites, and automated inputs from *in situ* and remotely sensed sources that include time series, profiles, swaths, grids, and other data structures. A wide range of telemetry systems, including the World Meteorological Organization's Global Telecommunication System (GTS), are used to transfer data from the measurement platforms to and among the locations at which Primary Data Assembly and Quality Control occur.

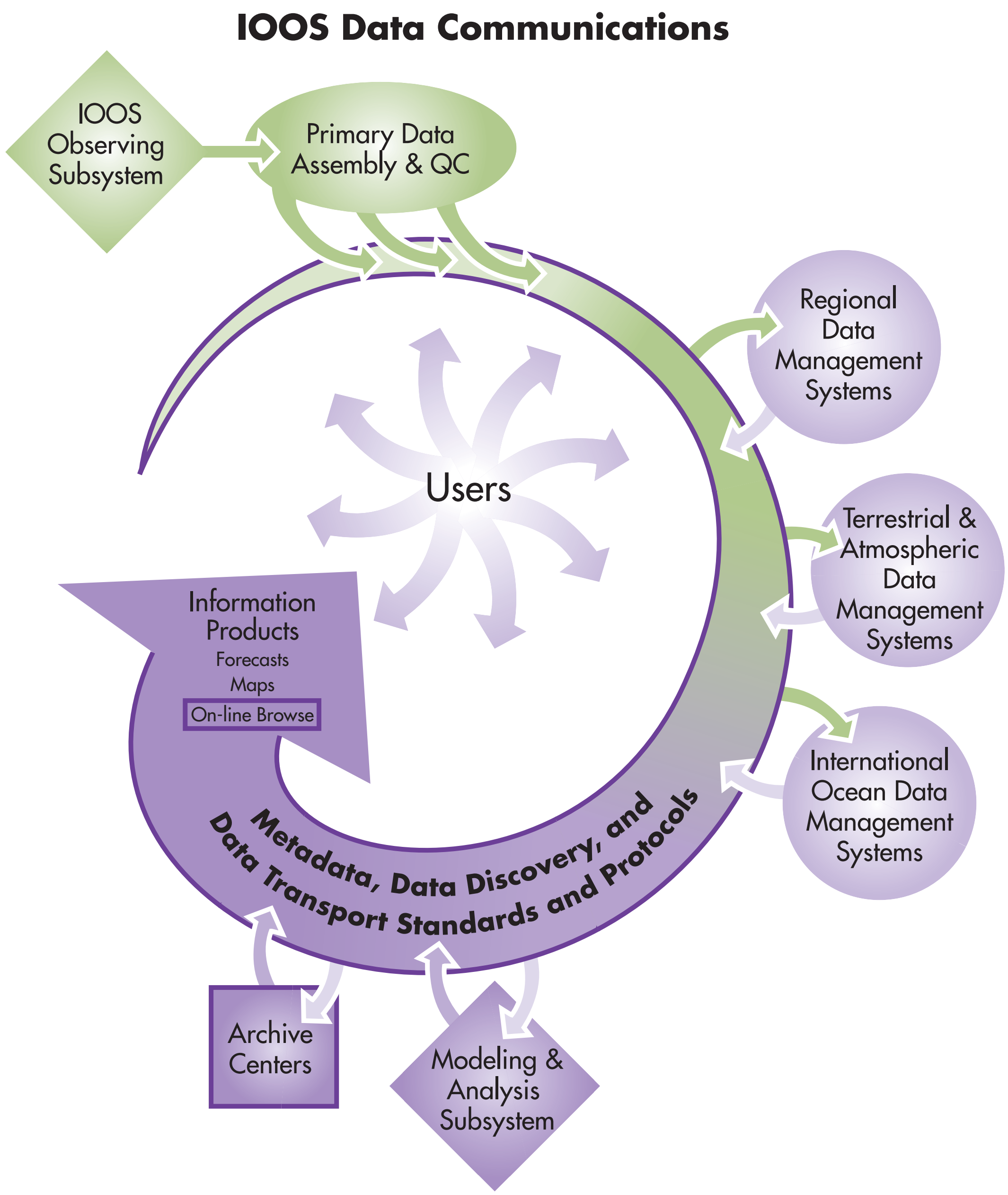

Figure 1. Solid outlines indicate the elements of the IOOS Data Communications framework, which are detailed in the DMAC Implementation Plan. The arrows flowing outward from users indicate the feedback and control mechanisms through which users ultimately direct the functioning of all parts of the system. Note that the National Data Management Systems are included in the concept of Primary Data Assembly and Quality Control.

Primary Data Assembly and Quality Control (PDA&QC) processes lie at the interface between the IOOS Observing Subsystem and the DMAC Subsystem. In general, some form of PDA&QC is required before IOOS data can be used. The character of this activity varies greatly by data type. It includes such processes as hand entry of numbers from log sheets, conversion from raw instrument voltages to physical units, calibrations, and the quality assessment of measured values against neighboring measurements or climatological norms. It is in this activity that myriad individual measurements are assembled into "data sets" (named collections of data) that may be referenced and queried as a whole. The activity may be conducted as a part of a data management strategy for a particular measurement type, where examples include NOAA/NDBC management of mooring data and NASA/PODAAC management of ocean satellite data. Alternatively, it may be managed by an IOOS Regional Association, for example assembling regional ecosystem measurements; or it may be found in conjunction with operational forecast modeling and state estimation, where a prime example is the U.S. GODAE Server (which operates in close association with the Navy's Fleet Numerical Meteorological and Oceanography Center).

Because of its location at the boundary between two subsystems, the responsibilities for PDA&QC are shared between the Observing Subsystem and the DMAC Subsystem. As a general rule the DMAC Subsystem's responsibility for procedures and standards of scientific quality control (QC) are limited to providing mechanisms to ensure that QC flags are reliably associated with the corresponding measurements. The standards and procedures for quality control will be developed by the relevant marine science communities. The DMAC Plan (this document), however, also includes a management responsibility to ensure that all IOOS data steams undergo primary data assembly and quality-control processing to make them available using DMAC standards and protocols at a specified DMAC Data Entry Point. In some cases this responsibility may involve support for specific facilities, such as the U.S. GODAE Server. In the future we may see the growth of "intelligent" instruments that can perform data assembly and quality control functions at the instrument subsystem level[5].

The telemetry systems that convey data from sensors to primary data centers/sites and the standards and procedures for data assembly and quality control lie outside the scope of this DMAC Plan. It is recognized, however, that these areas require careful planning within IOOS, and these topics will be addressed in future IOOS plans. The intent of the current DMAC Plan is to provide a foundation for these community building and planning activities.

[5]Delin, K.A., 2002, The Sensor Web: A Macro-Instrument for Coordinated Sensing, *Sensors*, 2, 270-285)

THE DMAC SUBSYSTEM – A DATA COMMUNICATIONS INFRASTRUCTURE

Metadata Management

(see Appendix 1 for a more detailed discussion)

Metadata are a critical component of the Data Communications Infrastructure, required for all key infrastructure functions: discovery, transport, on-line browse, archive, and access. A sustained commitment to the creation and management of metadata is a requirement to support the ability of users to locate and use data. Sustaining this commitment will be a challenge to IOOS Leadership of the sort discussed earlier in this document under Community Building. Certain classes of metadata (variable names, units, coordinates, etc.) are indispensable to any utilization of the data, and must be tightly bound to data transport as an integral part of the data delivery protocols. We refer to this class of information as "use metadata." Other types of information, such as descriptions of measurement and analysis techniques, help to place the data in context and are essential to overall understanding and usefulness of the data. We refer to this class of information as "descriptive metadata."

As mandated by Executive Order[5], "each [Federal] agency shall document all new geospatial data it collects or produces, either directly or indirectly, using the standard under development by the Federal Geographic Data Committee (FGDC)." While there are many IOOS members to whom this mandate does not directly apply, the breadth of participation found in FGDC makes it a natural choice for DMAC[6]. The FGDC developed the Content Standard for Digital Geospatial Metadata (CSDGM) that provides a common set of names and definitions of compound and individual data elements used to document digital geospatial data. The content of FGDC records encompasses the elements of many other metadata formats including most of the content contained in the Directory Interchange Format (DIF) metadata records in common use by international IOOS partners. The scope of FGDC, however, is far broader than marine data. A focused activity to determine the precise information that will define DMAC-standard metadata records, along with mechanisms for extensibility, are initial tasks identified within Part II of this phased DMAC Implementation Plan. Controlled keywords (standardized topic names) and controlled vocabularies (standardized technical terminology) need to be selected or developed. The breadth of scientific disciplines that will participate in DMAC guarantees the existence of overlapping terminology, and therefore tools and

---

[5]Executive Order 12906 (April 11, 1994)

[6]It should be noted that the International Organization for Standardization (ISO) developed a standard for geospatial metadata. This standard, ISO 19115, was formally accepted in May 2003. It is anticipated that the next version of FGDC Content Standard for Geospatial Metadata (CSDGSM) will be in a form compatible with the international standard.

techniques to perform translation among these controlled vocabularies are needed. "Parent-child" hierarchies of metadata must be supported, since marine data are often managed as collections of observations that require description both as inventories and as individual observations.

Data Discovery

Data Discovery will initially be implemented as a process of locating data of interest through searches of metadata. (The ability to search within the data itself—so-called "data mining"—will be incorporated into DMAC data discovery as the technologies for doing so mature.) The data-discovery capabilities provided by the DMAC Subsystem will complement and extend the search capabilities that are widely available today through commercial Web search sites such as *Google* and *Yahoo*. Data discovery will initially be accomplished through electronically searching IOOS metadata catalog(s). Typically, search parameters include geospatial location, temporal information, keywords, controlled vocabulary items, and in some cases "free text." The speed and reliability of searching are improved when the catalogs are centralized. Mirroring of smaller catalogs at centralized locations may be appropriate, and is an area for further exploration.

In marine data management today, metadata records are often managed independently from the data. This division is certainly true of metadata used for data discovery, and it is also emerging as a requirement for "use metadata." As a practical matter, the initial solutions to Data Discovery within DMAC must build upon this architecture. The separation of metadata record management from the associated data management, however, leaves two significant challenges to be solved: (1) assuring that links between metadata records and points of data access (see Data Transport) remain valid over time; and (2) assuring that changes made to data sets are reflected in corresponding changes to the metadata records. In addition, the issue of what constitutes a "data set" from the data discovery perspective when the data are assembled on-the-fly from distributed sources, must be addressed. These challenges are made more acute because a given data set (replicated) may be made accessible by multiple data-providing organizations. Changes to the data (new versions) may be made independently by these organizations. Part II of the DMAC Plan lays out steps to investigate and resolve these technical problems as the DMAC Subsystem matures.

The data-discovery capabilities of the Data Communications Infrastructure will both permit humans to formulate queries directly to the catalogs, and support machine-to-machine queries. DMAC will provide at least one "portal," Web pages through which end users can search for IOOS data. In the mature phases of DMAC, search entry points will also exist at many alternative locations that relay their queries via machine-to-machine communications to the DMAC search service. Advanced data discovery techniques, such as the "Semantic Web" (www.w3.org/2001/sw/), will be monitored and considered for incorporation into the system as they mature.

Data Transport

(see Appendix 2 for a more detailed discussion)

The concept of a "web service" is fundamental to the ability of DMAC to connect quasi-independent systems. The term "web service" is used in many contexts today; in the DMAC Plan we intend the term to mean reusable software components that are based upon the eXtensible Markup Language (XML) and provide a standardized means for computer systems to request data and data processing from one another. Web services make data and software capabilities available on one computer, accessible to other computers via the Internet. Web services use the ubiquitous communications protocol of the World Wide Web, HTTP (the Hyper Text Transfer Protocol), and are accessed through the familiar Universal Resource Identifiers (URIs) that begin with "http://".

The DMAC will endorse a suite of web services to serve as a shared communications toolbox connecting systems that are operated by regional, state, and Federal agencies, academic projects, and others. Data suppliers (including Primary Data Assembly and Quality Control sites) will be responsible for making data accessible through DMAC web services tools and standards. Data users will find that in many cases the software applications upon which they depend for product generation and scientific analysis will be "DMAC-ready" (possibly with some adaptation required), having been adapted to work directly with the DMAC web services. In this case, the applications will perform much as if the data existed on their local hard drive. At a minimum, compatibility between DMAC and user applications will be achieved using formatted files that are made readily available as products through DMAC web services.

DMAC will provide most of the necessary web services software components and will assist with the creation of additional software to meet special needs. The goal in doing so is to minimize the barriers to participation in the DMAC. The uniformity provided by the DMAC web services standards will permit all the components related to data transport to be interoperable at the machine level (i.e., data can be moved from one component of the system to another, retaining complete syntactic[7] and semantic[8] meaning without human interaction).

[7]Syntactic Meaning: refers to the syntax of a data set - the atomic data types in the data set (binary, ASCII, real, etc.), the dimensionality of data arrays (P is a 90 by 180 by 25 by 12 element array), the relationship between variables in the data set (lat is a map vector for the first dimension of P), etc.

[8]Semantic Meaning: refers to the semantics of the data contained in the data set - the meaning of variables (P represents phytoplankton abundance), the units used to express variables (multiply P by 8 to obtain number of specimens per cubic meter), special value flags (a value of –1 means missing data, 0 land,...), descriptions of the processing or instrumentation used to obtain the data values, etc.

Web services exist in the context of the World Wide Web (Web). Data transport on the Web involves protocols at multiple levels. The foundation of transport on the Internet is TCP/IP, which handles the routing of "packets" of information between source and destination hosts. Layered upon TCP/IP are a variety of protocols, for example, FTP, HTTP[9], and SMTP. These protocols are supported on a very wide range of computers and operating systems, and all of them will be used to move various types of data over the network as part of IOOS. There is, however, no uniform syntactic and semantic meaning that is guaranteed for data communicated via these transfers, and therefore no guarantee of immediate interoperability among computer applications. This function is the role of the DMAC web services standards.

Several solutions currently exist for the syntactic description and transport of binary data, however none is universally accepted. The most broadly tested and accepted of these solutions within oceanography is the OPeNDAP[10] data access protocol. OPeNDAP underlies the National Virtual Ocean Data System (NVODS[11]). OPeNDAP has been serving the marine community since 1995. OPeNDAP provides the very general approach to data management that is needed in support of research and modeling. OPeNDAP also supports server-side subsetting of data, which greatly reduces the volumes of data that need be transferred across the Internet in many cases. This capability is vital when considering the large volumes of data that will be produced in the near future by observing platforms and modeling activities. Tables 1 and 2 provide estimates of near-term data flow for the US IOOS using selected data streams as examples; the lists are not intended to include all observing system data types. The OPeNDAP protocol has been designated by the DMAC-SC as an initial "operational"[12] component for Data Transport of gridded data, and a "pilot"[13] component for the delivery of non-gridded data.

OPeNDAP

OPeNDAP uses a discipline-neutral approach to the encapsulation of data for transport. Discipline neutrality is seen as a key element of the data transport protocol for a system such as IOOS that covers such a broad range of data types and data users—for example, four-dimensional geospatial

[9]When using a Web browser most images and text are delivered via http.

[10]The Open Source Project for a Network Data Access Protocol (OPeNDAP), a non profit corporation formed to develop and maintain the middleware formerly known as the Distributed Ocean Data System (DODS).

[11]NVODS was created in response to a Broad Agency Announcement (BAA) issued by the National Oceanographic Partnership Program (NOPP) in 2000.

[12]"Operational" is stage four of a four-level classification scheme for the maturity of system components within IOOS: R&D, pilot, pre-operational, operational. See IOOS Implementation Plan Part I (www.ocean.us/documents/docs/ioos_plan_6.11.03.pdf).

[13]"Pilot" is stage two of a four-level classification scheme for the maturity of system components within IOOS: R&D, pilot, pre-operational, operational. See IOOS Implementation Plan Part I.

Table 1. Near-term Data Flow Estimate for U.S. IOOS[14] (NASA & NOAA sources excluding cabled observatories)

Data Source Class	Data Source Type	Annual Volume (MB)	Totals Annual Volume by Class (MB)
Direct Observation Systems: Buoys	Moored buoys - NDBC, TAO, MBARI, etc.	2,000	
	Drifting buoys - Surface, APEX, etc	100	2,100
Direct Observation Systems: Ships	NOS Charting/Resurvey	1,800,000	
	Other Ship Data - VOS MET, XBT, CALCOFI, etc.	12,000	1,812,000
Remote Sensing Systems	Surface currents-CODAR	13,000	
	Sea Surface Temperature - AVHRR, MODIS	500,000	
	Sea Surface Height – T/P, JASON1	120,000	
	Ocean Vector Winds - QuikSCAT, SeaWinds	130,000	
	Ocean color-MODIS, SeaWIFs	400,000	1,163,000
Total Near-Term Annual Data Flow			**2,977,100**

Table 2. Near-term Data Flow Estimate for U.S. IOOS: Cabled Observatories[14] (NASA & NOAA sources)

Data Source	Annual (GB)
MBARI/MARS	13,000
HUGO	5,000
LEO 15	5,000
Neptune (approval pending)	177,000
Total cabled Observations	**200,000**

[14]Science Applications International Corporation. October 18, 2002. "Consolidated Data Flow Estimates for the Integrated Ocean Observing System (IOOS)." Submitted to the National Ocean Service, National Oceanic and Atmospheric Administration, Department of Commerce. Note that this was a preliminary study. Further investigation is needed.

Biological Data Considerations

Management and stewardship of biological data present special challenges, which historically have often been neglected. Biological data management requires that special consideration be given to metadata (see Part III, Appendix 7 for a more detailed discussion). For example, the basic units for biological data are species. New species are continually being discovered and named, and names of recognized species are sometimes changed. The hierarchical tree of evolutionary relationships among species, and the associated hierarchical nomenclature, must continually be revised to incorporate new information. Biological data systems require name translators that provide currently recognized scientific names from synonymous scientific names and common names. The taxonomic authority for each major group of organisms maintains the accepted list of species, with oversight from the Global Biodiversity Information Facility (GBIF), Catalogue of Life, and organizations such as the Integrated Taxonomic Information System (ITIS)/ Species 2000, and the Ocean Biogeographic Information System (OBIS). Protocols for using DNA sequence data as a "bar code of life" have been proposed as an aid to taxonomic identification and evolutionary relationships.

grids, time series, vertical profiles, and species type and abundance. The protocol ensures that the structure, numeric values, and metadata attributes of the data are preserved between server and client. It does not, however, impose a particular geospatial data model. For example, OPeNDAP does not "understand" what a time series is, nor does it "know" the significance of "phytoplankton_abundance" as a variable name. When transmitting a simple time series, OPeNDAP merely knows that it is conveying a one-dimensional array of values with attributes such as units = "seconds" and title = "Phytoplankton Abundance" attached to it. Such an approach greatly lowers the barrier to initial participation by data suppliers, since nearly all data holdings can easily be encapsulated in this fashion and sent over the Internet. It also ensures the level of generality needed to provide semantically aware data transport for the very broad range of ocean data classes.

To achieve the desired level of interoperability, the mature DMAC will require that all data are delivered in a consistent geospatial data model (or family of models). In this phased DMAC Implementation Plan, the development of a rich and comprehensive data model occurs in parallel with the pilot deployment of OPeNDAP data servers and clients. The comprehensive data model(s) must harmonize with ongoing work in several communities, such as GIS and forecast and climate modeling. It must standardize controlled vocabularies, include the encoding of ocean biological data and taxonomies such as those demonstrated in OBIS[15], and describe a broad range of data structures, including for example, spectral and finite element models, arbitrary curvilinear coordinate systems, and multi-level hierarchies of *in situ* measurements. The parallel progress made by deploying OPeNDAP servers and clients, while simultaneously designing a comprehensive data model and community-wide metadata standards, is a key element of the phased DMAC Implementation Plan. This element will enable rapid progress both in capacity building and in broad community standards building. It is anticipated that the design of the data model may necessitate changes or additions to OPeNDAP.

On-line Browse

Effective management of the DMAC Subsystem requires the existence of a basic browsing and visualization capability that extends across the breadth of IOOS data. The browse capability must guarantee a minimum quality (to be determined) of geo- and time-referenced graphics and readable tables of values. It must be accessible through standard Web browsers. On-line Browse complements the Data Discovery and Data Transport elements of the DMAC Data Communications Infrastructure by providing a uniform means to inspect (visualize) and make comparisons among IOOS data subsets. The DMAC Subsystem must provide a seamless segue from Data Discovery to On-line Browse. The On-line Browse software may use DMAC Data Transport to access the data.

The On-line Browse capability is a form of information product (see Information Products and Applications). As such it must be an effective informational tool for its target user groups. The target users of the on-line browse capability are the marine data specialists across the IOOS community who have responsibilities for managing elements of IOOS. These users may be assumed to share a high level of technical training, but they represent diverse professions, including scientists, computer specialists, engineers, and technical managers. Although the on-line browse capability must be designed principally to address the needs of these users, it will also be accessible to the general public. It is expected that the on-line browse capability will prove useful to many other groups of users.

DMAC will provide at least one style of user interface through which users can browse IOOS data. In the mature phases of IOOS there may be multiple user interfaces, allowing the on-line browse capability to address a broader range of users. The on-line browse capability must be accessible through a computer-to-computer "web service" interface, enabling the browse products that are

[15]OBIS is an on-line, open-access, globally distributed network of systematic, ecological, and environmental information systems. Collectively, these systems operate as a dynamic, global digital atlas to communicate biological information about the ocean and serve as a platform for further study of biogeographical relationships in the marine environment (http://iobis.org/).

provided to be incorporated into a range of applications. The On-line Browse architecture must be designed to scale with the growth of IOOS, and must be highly configurable so that new and modified browse products can readily be added to meet evolving user needs.

Within the National Virtual Ocean Data System, the Live Access Server (LAS; http://www.ferret.noaa.gov/LAS) has been effectively used for several years in the delivery of on-line browse capabilities across a broad range of marine data types. The DMAC Steering Committee has designated the Live Access Server as an initial “pre-operational” component for On-line Browse.

Data Archive and Access

(See Appendix 3 for more detailed discussion)

The Data Archive System will receive and provide access to both real-time and delayed-mode data and metadata, serving the needs of real-time assessment and prediction, scientific research, and all others who require access to archived IOOS data. It will be a high priority for the Archive System to ensure that all valuable data are sent and that an exact copy is received. The Archive system will be designed to detect and correct failures using a combination of technological backup and expert oversight that will check the integrity of the data streams. In addition, during the phased implementation, a comprehensive process will be undertaken to ensure that all critical data streams and existing historical archives are inventoried and are scheduled to enter into the system.

The Data Archive System will consist of a designated set of existing and new facilities. Initially, existing centers will be the basis for the System; operating principles, requirements for additional facilities and cross coordination among facilities will all be defined during the early planning phase. The Archive System will include distributed archive centers, regional data centers, modeling centers, and data assembly centers (Figure 2). Although data may flow from observing subsystem components to any of the four types of centers, at least one copy of each observation will ultimately reside in an archive center. Data will be considered as officially in the Archive System if the following two conditions are met: (1) the data are held and access is provided by one of the Archive System centers, and (2) there are established procedures in place to preserve the data at an archive center. Through this approach, data will be under IOOS management early on in their life cycle, thereby maximizing the amount of securely archived and uniformly accessible data. It is probable and practical that more than one type of center may be physically collocated, for example, a data center may be an entity at a national archive center.

All centers in the Archive System may be responsible for acquiring and providing data, but (by definition) it is the archive centers that will have primary responsibility for preserving data for the long term (Figure 2). To qualify as an archive center, a data center must be able to manage multiple

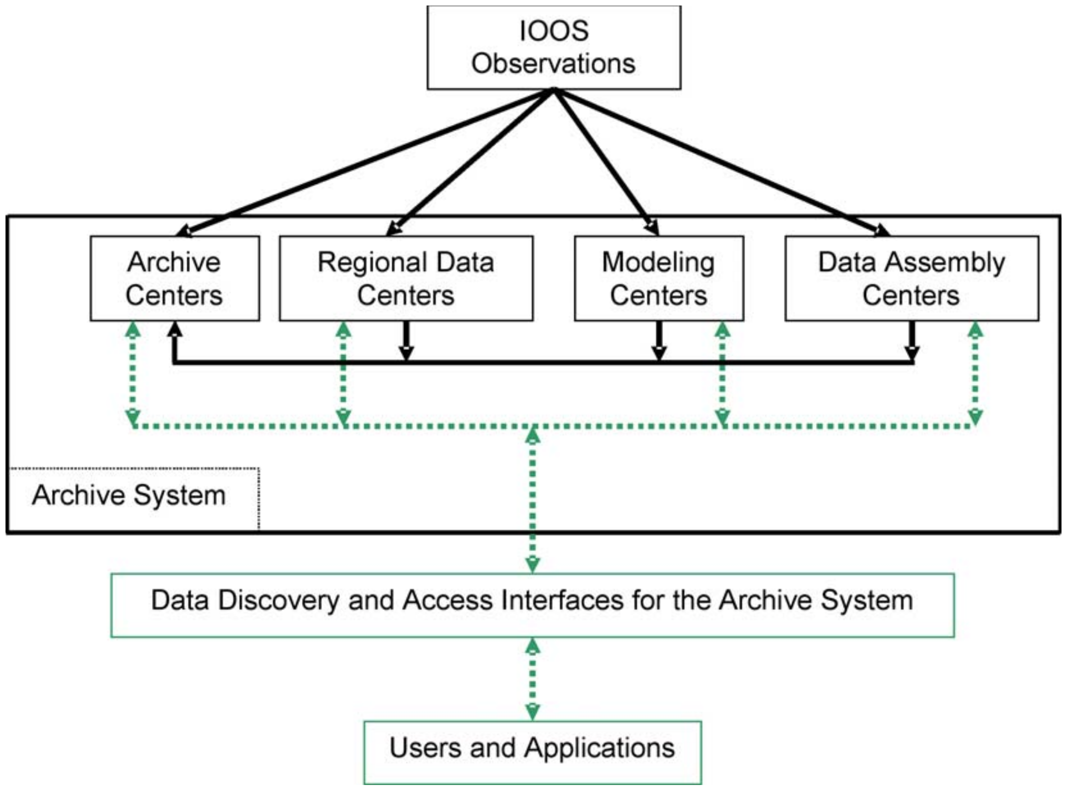

Figure 2. The Archive System represents an alternative view of those DMAC Subsystem elements that are involved with data archival. Primary archival (solid lines) and access (dashed lines) show data flow. Not shown are other data flows that are essential to IOOS but not directly pertinent to the Archive System.

copies of the data and metadata, create and verify the metadata, frequently check data integrity, and have plans to evolve systems and media through generations of technology. Data will be preserved according to data categories, which will be developed by an Archive Working Group during phased implementation of the DMAC, and according to U.S. National Archives and Records Administration (NARA) and other Federal guidelines.

Technologies and standards developed by metadata management and data transport activities will be an integral part of the Archive System. In the mature DMAC Subsystem, which is required to deliver data in a timely manner, it is anticipated that data and metadata will be received by the archive centers and redistributed through the DMAC transport mechanisms in standardized formats, eliminating many of the delays and difficulties of the non-standard and diverse methods that have burdened the systems in the past. Evolution to this state will be stepwise so that current data services are not interrupted and users can make a smooth transition. The result will be a system that provides uniform access across multiple centers, and provides data discovery and access by both humans and machines. Furthermore, all irreplaceable observational and research-quality data that are difficult to regenerate will be maintained and managed in perpetuity.

Unrestricted data access is a primary principle for all IOOS data sets, however, circumstances may arise where temporary restrictions are permitted. These instances are envisioned to be short term, where the burden of managing data set authorization and authentication can be offset by the reduced cost and increased efficiency of archiving the data at an early stage. The opportunities for limited restricted access, data security, and metadata and data discovery support offered by the Archive System are an asset, previously unavailable, and are intended to encourage broad participation from the scientific community.

The scientific community will add value to the data in the Archive System. The System will enable scientific endeavors that make comparisons of model and observed data, develop analyses and reanalyses data products, provide additional data quality control, and thereby quality checks on the observing systems. The Archive System will receive these additional data products, use the discoveries to augment data stewardship activities, and have mechanisms to inform the IOOS observation subsystem about data quality concerns.

Data Archive Policy

All facilities that participate as official archive centers in the Archive System will agree to adhere to archival guidelines that will be established in the phased implementation of the DMAC. A few key points that will be part of the guidelines are:

- Data distribution policies will follow the international recommendations of the IOC and WMO. Generally, the policy will call for full and open sharing of data and products. As a possible extension, the ability to provide restricted access for limited periods of time may be provided in certain cases;
- Data will be made accessible, to the greatest extent practical, on line and at no cost to the user. Data from off-line sources will similarly be available at no more than the cost of providing the service;
- Centers in the Archive System will make the data and metadata available using the DMAC transport protocols, metadata standards, and data discovery interfaces. The details of the transition from existing access systems to systems using the DMAC standards remain to be determined;
- The archive centers in the Archive System will have a data and metadata migration plan to accommodate media and system evolution and assure long-term preservation of irreplaceable data;

- All data collected and prepared under IOOS funding shall be submitted (or, in appropriate cases, notification of its availability shall be submitted) to the IOOS Archive System;
- As new versions (upgraded or changed) of a data set become available the versions will be distinguishable through standard metadata. Old versions can be deleted only under restrictive circumstances—when all relevant IOOS data policies and Federal regulations are met.

MODELING AND ANALYSIS SUBSYSTEM

The IOOS Modeling and Analysis Subsystem has responsibility for the generation of numerical (digital) data products through: (i) computer modeling and the assimilation of marine observations, which provide estimations of the current state of the marine environment and forecasts its future state, and (ii) analysis of data collections, which incorporate late-arriving observations and apply further quality controls in order to ensure that the historical record of marine observations is as complete and accurate as possible. The activities of the Modeling and Analysis Subsystem are carried out at many distinct centers. The term "center" may refer to an organization that has a specific focus on numerical modeling or data analysis, or it may refer to an individual project embedded within a university, state government, or other organization. It is the responsibility of IOOS Governance and User Outreach mechanisms to ensure that the Modeling and Analysis Subsystem meets the needs of the full range marine stakeholders.

Planning for the Modeling and Analysis Subsystem and the particular numerical data products that IOOS must produce lies outside the scope of the DMAC Plan (this document). It is recognized, however, that these topics require careful planning within IOOS, and they will be addressed in future IOOS plans. The intent of the DMAC Plan is to provide a foundation for these community-building and planning activities.

INFORMATION PRODUCTS AND APPLICATIONS

The ultimate goal of DMAC is to provide information about the marine environment to end users in a manner which permits them to advance the seven goals of IOOS (Preface). Information can be provided to users through the generation of Information Products, or through ingestion of data into DMAC-ready applications. Information Products include text documents, such as printed assessments of fish stocks; verbal reports, such as wave height announcements on marine radio; maps and charts, including GIS layers; and graphics, animations, 3D visualizations and other media generated by computers to assist with the communication of information. The term "DMAC-ready Applications" refers to those applications that can directly access data and information through

DMAC standards and protocols. This designation will (in the mature DMAC) include common GIS applications, common scientific analysis and visualization applications, educational software, and common business tools, such as spreadsheets and word processors.

Because IOOS is a user-driven system, the IOOS User Outreach and Governance mechanisms must ensure that the users' needs for Information Products and DMAC-ready applications are continually assessed and accommodations made within IOOS to meet them. Information Products will be generated at all levels of IOOS: by Primary Data Assembly Centers/sites (e.g., the Argo GDACS); by the DMAC Communications Infrastructure (see On-line Browse above); by DMAC Modeling Centers and Archive Centers; and by the regional, international, and discipline-specific data management systems that interoperate with DMAC data. It is understood, however, that these products alone will not be sufficient to meet the needs of all user groups. There is a vital role for private sector IOOS partners in providing users with specialized value-added Information Products and DMAC-ready applications.

INTEROPERABILITY WITH OTHER DATA MANAGEMENT SYSTEMS

The DMAC Data Communications Infrastructure provides for interoperable communications between DMAC components and other data management systems containing data that are of interest to IOOS users. These entities include data management systems operated by disciplines lying outside of marine sciences (e.g., public health), data management systems operated by other nations or international bodies, and specialized data management systems that may be operated by Regional Associations within IOOS. Like the DMAC Subsystem these systems may be highly complex and involve multiple distributed partners. They may use data communications infrastructures that are distinct from the DMAC Standards and Protocols. To achieve interoperability with these entities, software "gateways" must be built that will translate between standards and protocols used within the DMAC Subsystem and those used in the other data systems. Similar considerations apply to commercial organizations that have adopted custom "enterprise" solutions[16], as well as frameworks adopted by communities that share specialized computing needs, such as the high performance computing community's DataGrid[17] and the GIS community's geospatial data systems. Thus, as IOOS matures, it should be viewed as a system of systems in which the DMAC Data Communications Infrastructure provides a uniform language for communications.

[16]"Enterprise" solutions are typically commercial interoperability frameworks that operate on secure network connections.
[17]http://www.globus.org/datagrid

Section 3. Management, Oversight, and Coordination

GOVERNANCE

The governance of the DMAC Subsystem should be designed to operate within the context of the IOOS governance mechanisms described in Parts I and II of the IOOS Implementation Plan.

The DMAC-SC recommends that Ocean.US establish a permanent DMAC Standing Committee (DMAC-StC), building upon the existing committee, to serve as a technical planning and oversight body, with some performance responsibilities. The new committee should be committed to the development and implementation of technical solutions (e.g., standards for data management and communications; system-level interoperability-enabling protocols, etc.) necessary to meet the requirements of the IOOS user communities and stakeholders. It would be the focal point for data management and communications within Ocean.US and across IOOS. Specific DMAC-StC activities should include: data and metadata standards development and implementation; development of priorities and investment strategies for IOOS DMAC implementation and integration across the IOOS Regional Associations and NOPP Agencies; coordination of IOOS DMAC activities with related national and international programs; and continued refurbishment and modernization of the DMAC Subsystem of IOOS.

Membership on the new committee should include technical experts drawn from the public and private sectors who represent the IOOS data provider, user, and development stakeholder communities, including NOPP Federal Agencies, IOOS Regional Associations (RAs) and Coastal Ocean Observing Systems, the international observing system community, the private sector, academia, industry, and NGOs.

The original DMAC-SC appointed by Ocean.US in the spring of 2002 will complete its work on the following activities in the spring of 2004:

- Finalize the DMAC Implementation Plan,
- Initiate and oversee interim pilot and demonstration projects, and
- Establish a permanent DMAC governance/management mechanism.

IOOS DATA POLICY

IOOS data policy is under development, and will be put into effect at an early stage of IOOS implementation. It will be consistent with:

- U.S. Federal data policies,
- International GOOS Design Principle 7, "GOOS contributors are responsible for full, open, and timely sharing and exchange of GOOS-relevant data and products for non-commercial activities," (IOC, 1998),
- The IOC/IODE Data Exchange Policy, adopted in 1993 (Meeting of the *Ad Hoc* Working Group on Oceanographic Data Exchange Policy IOC/INF-1144rev, 4 July 2000), and updates adopted at the 22nd session of the IOC Assembly in July 2003 - Resolution XXII-6 & 7,
- The WMO policy of free exchange of meteorological and related marine data (WMO Resolution 40, Publication WMO—No. 837).

All data collected and prepared under IOOS funding shall be subject to the IOOS data policy. Generally, the policy will call for full and open sharing of non-proprietary data and metadata, products, model code, and related information. It will also call for adherence to data, metadata, and data products standards promulgated by IOOS. Specific requirements for each of the DMAC Subsystem elements are discussed in the Plan sections on Metadata, Data Discovery, Data Transport, On-Line Browse, and Data Archive and Access. Archive facilities that participate as official IOOS Archive Centers will further agree to adhere to archival guidelines that will be established in the Archive phased implementation of the DMAC.

IOOS will not compete with the private sector because it will not distribute commercial data, products, or services produced by commercial enterprises.

IOOS/DMAC STANDARDS PROCESS

The marine sciences community has made significant progress toward data integration during the past two decades. Community-wide programs such as GLOBEC, OBIS, LOICZ, WOCE, JGOFS, and, recently, Argo have done much to establish new traditions in marine data management. These new traditions recognize the importance of data standards and the value of shared data access within individual disciplines. The community is now facing a new generation of standards-related problems: to be effective, operational oceanography will require integration of data and product streams from many distinct disciplines. Marine data standards that have been narrowly focused, though documented and well supported, are often not sufficiently interoperable to address this requirement. With the development of GOOS, similar standards issues are also being addressed by JCOMM, the IODE, and the OITP.

Defining data standards is a slow and expensive process. Typically a group of technical experts must meet repeatedly over a period of years to develop and agree upon a data standard of modest scope. Thus the DMAC development cannot wait upon a systematic redesign of marine data standards, alone, to achieve the required level of interoperability. Rather, the focus of this Plan is on the use of protocols and translators that can achieve an acceptable level of interoperability building upon standards that exist today. This approach is discussed in greater detail in Part II and in the Data Transport Appendix.

Adopting this approach represents a compromise. The level of interoperability that can be achieved among differing standards is often limited by mismatches in the information content of the standards, or differences in the semantic data models that underlie them. In the long term, achieving the desired level of data interoperability will require that the community develop and utilize fewer standards that are of greater breadth.

In parallel with building the interoperable Data Communications Infrastructure, this phased DMAC Implementation Plan recommends that DMAC begin work to foster an improved standards process. The DMAC standards process must be open so that it represents community consensus. It must be highly visible so that the standards are broadly used, and it must carry official stature so that the standards will be respected and used appropriately. It must also build on existing standards and standards processes whenever possible. To be fully successful, IOOS must foster the adoption of community standards that encompass quality control, scientific analysis, data-set versioning, metadata, products and services, data discovery, network data transport, file formats, and data archival.

USER OUTREACH

(see Appendix 4 for complete User Outreach Team Report)

The recognition and incorporation of users' needs is essential to the success of IOOS and all other components of GOOS. This effort extends well beyond the boundaries of the DMAC Subsystem. Indeed, from its inception IOOS is envisioned as an "end-to-end" system tailored to address the needs of the "end user." True end users are generally not information technology specialists, but professionals who rely on information that has been developed from data by other professionals. Examples of end users include the commercial fishery manager, the oil spill response team leader, the U.S. Coast Guard watch officer, and the harbor master.

User Outreach refers to the structures within IOOS governance that will recognize and address the needs of end users on an on-going basis. The goals of User Outreach are to: (1) identify the needs of users, (2) influence IOOS to meet those needs, (3) assess how well those needs are being met, and (4) report the results of 1 and 3 to all parts of IOOS. User Outreach also seeks to establish and enhance the societal relevance of IOOS through informing the concerned public about existing products and services.

There are at least three classes of users: (1) the end users mentioned above, (2) users who process (assimilate, analyze…) data and provide information in the form of forecasts, ocean state estimations and hindcasts, and (3) users who provide specialized software and other services. Geographic and institutional structures add further complexity to identifying the needs of the various user groups. The needs of users are generally structured according to the seven phenomena of interest to IOOS: marine operations, natural hazards, national defense, public health, climate change, healthy ecosystems, and sustainable marine resources. For an in-depth description of user needs from the viewpoint of a national panel of experts, please see Appendix 4 (the complete User Outreach Team Report).

User needs are translated into software requirements and then into software product design through the spiral design cycle described below in the System Engineering Approach. On-going input from users will be solicited at each stage of the design cycle. This input will shape the standards-generating and system implementation tasks that are outlined in Part II of the Plan.

SYSTEM ENGINEERING APPROACH

(see Appendices 5 and 6 for more detailed discussions)

This document describes a wide variety of requirements addressing the needs of a diverse group of stakeholders. Because of the resulting complexity, the success of the DMAC Subsystem requires a formalized system engineering process. A brief description of three system engineering process models is presented in Part III (Appendix 5), as well as recommendations for the approach that should be used for the DMAC development and integration.

Based on a review of the subsystem requirements and a comparison of the features of several common models, it is recommended that the Spiral Model for systems engineering be selected for DMAC implementation. The Spiral Model accommodates a "task-oriented," highly structured approach, while allowing rapid prototyping and risk-analysis to be performed at juncture points of the project. In the Spiral Model, selected requirements are chosen for development to an operational level. Then, more requirements are added, and the development process is repeated through this "spiral" until all requirements are accomplished. The phases can be executed using a waterfall-like process, i.e., with requirements specification (or updates), analysis and design, system development, and verification performed for each phase. Each phase (sometimes referred to as an effectivity), would then represent a complete end-to-end execution of a subset of the requirements. A phased approach adapted to fit the DMAC purpose is presented in Appendix 5.

To ensure that the DMAC Subsystem meets the program goals, it is critical that the technology stay current and operational through adherence to a concrete plan for maintenance and refreshment. Appendix 6 discusses the life cycle maintenance and refreshment of the technology components of the DMAC.

Much of this document describes systems capabilities that are reliant on technology, whether hardware or software. A Price Systems LLC[18] paper defines technology refreshment as "the periodic replacement of commercial off-the-shelf (COTS) components; e.g., processors, displays, computer operating systems, commercially available software (CAS) within larger … systems to assure continued supportability of that system through an indefinite service life." We would add communications capabilities and storage media to this list. Systems are being acquired that are ever larger and more complex, constructed out of components designed and built by commercial third parties. If the DMAC is viewed as a "system of systems," this situation still applies; whether "component" in this context means a server or an application software suite, it still represents an item that must be evaluated and, if appropriate, refreshed periodically in order for the overall system to meet its evolving mission requirements.

For the DMAC Subsystem to remain current, the technology might be refreshed during the system life for a number of reasons, including the following:

- The existing system component has malfunctioned and either cannot be repaired, or the repair costs exceed the replacement costs,
- The existing system component has reached its life expectancy,
- The surrounding technical infrastructure has evolved and is incompatible with the existing component under consideration,
- Evolving requirements have made an alternative technology more cost-effective,
- Newer technology has come to market that provides more capability for the same or lower total cost of ownership.

[18]Technology Refreshment - A Management/Acquisition Perspective" (2001), available at http://www.pricesystems.com/downloads/pdf/technology%20refresh.pdf

The DMAC Technology Refreshment Plan (TRP) (discussed in Part II) will to the greatest extent practical draw upon established protocols such as the Navy's Technology Assessment and Management Methodology (TeAM)[19] or the Technology Refreshment Cost Estimating and Planning Model[20].

DMAC COST MODEL

The provision of marine data over the Internet is experiencing a period of rapid growth in both the volume and breadth of services offered. The aggregate cost of these activities is likely to increase over time irrespective of the existence of IOOS. The cost estimates for the DMAC Subsystem that are presented in Tables 3 and 4 reflect only those expenses that will be incurred over and above this background of growth. Expressed another way, the cost estimates in the DMAC Plan represent only those expenses that explicitly address the tasks of data integration achieved through the development of and participation in the standards, protocols, tools, and policies of the DMAC Subsystem.

Table 4 shows the cost model for the DMAC Subsystem for a ten-year period. The Plan calls for the initiation of the full DMAC Subsystem over a five-year period at a cost of $82 M in new funding. The initiation costs include the development of core standards, protocols, and tools ($28 M); costs of hardware, software, networking capacity, archival center expansion and systems integration labor ($37 M); and a budget for focused pilot projects to usher in and test the new technologies ($17 M). Out-year recurring costs over the following five years (Years 6 through 10) total an additional $85 M. Substantial new funding for IOOS is not anticipated until fiscal year 2007 (FY07), yet a minimally functioning DMAC Subsystem must already be in place to support the initial growth in IOOS measurements, modeling, and usage at that time. The DMAC Plan includes tasks and associated costs totaling $2.1 M during FY04 , 05, and 06 that are deemed to be very high priorities for immediate implementation in order to prepare for FY07 demands on the Subsystem.

The assumptions driving the cost model, and additional details about the model components are presented below:

- General Assumptions
  - The plan assumes that the program cost reflects new efforts above and beyond existing program elements. It further assumes the existence of data collection and processing components. Therefore, program costs reflect new services, hardware, software, and infrastructure costs, and do not include costs that are already budgeted for ocean observing systems.

[19]Technology Assessment and Management Methodology – An Approach to Legacy System Sustainment Dynamics" (1998), available at http://smaplab.ri.uah.edu/dmsms98/papers/samuelson.pdf

[20]Technology Refreshment Cost Estimating and Planning Model: User s Guide (2001), available at http://www.its.berkeley.edu/nextor/pubs/RR-00-5.pdf

Table 3. DMAC Overall Program (thousands of $)

	Y-2	Y-1	Y1	Y2	Y3	Y4	Y5
Pilot Projects							
Data Discovery			$1,000	$1,022	$783	$534	$0
Acess/Infrastructure			$500	$511	$522	$0	$0
Data Transport			$1,000	$511	$522	$534	$545
Archive			$500	$409	$418	$801	$1,091
Information Assurance			$500	$204	$104	$534	$545
Innovative Architectures			$300	$307	$522	$1,601	$818
Total Pilots			**$3,800**	**$2,964**	**$2,871**	**$4,004**	**$2,999**
Program Initiation Labor							
Program Management Activities	$36	$72	$726	$742	$752	$769	$792
Metadata and Data Discovery	$335	$271	$2,480	$2,412	$1036	$463	$975
Data Archive and Access	$235	$335	$1,612	$3,076	$1,799	$1,497	$1,139
Data Transport	$450	$348	$2,234	$2,980	$699	$740	$693
Total	**$1,056**	**$1,026**	**$7,052**	**$9,210**	**$4,286**	**$3,469**	**$3,599**
Program Initial Fixed/Maintenance Costs (Inflation-adjusted costs shown)							
Communication/Infrasturcture			$1,460	$1,594	$1,734	$747	$764
Servers at Centers			$2,400	$2,821	$3,259	$1,153	$1,178
Archive Capacity						$4,163	$2,836
Engineering/Integration			$3,000	$3,475	$3,342	$1,921	$1,746
Total Initial Fixed/Maintenance			**$6,860**	**$7,890**	**$8,335**	**$7,984**	**$6,524**

Table 4. Total Program Costs (thousands of $)

Program Initiation Costs							Outyear Recurring Costs				
Preparatory		Core Program									
Y-2	Y-1	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
$1,056	$1,026	$17,712	$20,064	$15,492	$15,457	$13,122	$17,645	$18,033	$18,430	$15,265	$15,600

Grand Total, Preparatory	**$2,082**
Grand Total, Initiation	**$81,847**
Grand Total, 10 Years	**$166,819**

- The plan assumes that funds will be appropriated to Executive Agencies as a part of the normal budget process, and that the funds will be earmarked for DMAC. The exact mechanism for appropriation and allocation to DMAC is outside the scope of this plan (for details, refer to the IOOS Implementation Plan, Parts I and II).
- It is assumed that Agencies will apply earmarked DMAC funds to the execution of the activities in this plan, under the coordination of a DMAC Program Management Office.
- The cost model is designed to support the Data Management and Communications subsystem of the IOOS only; costs for sensors, data ingest, modeling, and end user capabilities are not included. Specifically, the DMAC does not include the costs of the observing subsystems nor the modeling and analysis subsystem, although those components are part of the IOOS.
- The cost model is transparent with respect to the origin of the observation type; e.g. remotely sensed data would be within scope, but only from the point at which the data enter the IOOS, not including data providers' sites.
- The overall costs of the plan were not derived by comparison to like systems; indeed, it is not believed that a comparable system of the magnitude of IOOS exists today. Rather, the cost estimates were derived by a combination of level of effort analysis and estimates of the magnitude of system capacity required. It is fully anticipated that the planned Systems Engineering Tasks will provide more refined estimates of acquisition and integration costs.

- *Pilot Projects* - The goals of the Pilot Projects are to provide proof-of-concept demonstrations and advanced technology development and integration. Given these goals, it should be noted that the cost estimation for Pilot Projects is much less precise than that for the labor costs of the three subsystems. Pilots are intended to move components from the developmental to the operational stage.
  - Pilot Projects include both design and demonstration, depending on the level of maturity of the underlying technology. The program assumes an approximate division of 60% for design and 40% for demonstration.
- *Labor by Subsystem* – This category reflects the labor costs to develop system capabilities and reach a Full Operating Capability by Year 5.
  - "Program Integration Labor" consists of the total cost of labor for the development and integration activities of the three segments: Metadata and Data Discovery (including On-line Browse), Data Archive and Access, and Data Transport. Labor costs are based on a standard contractor rate of $180,000 per annum, except for staffing of the AA, at $100,000 per annum. The exact distribution of the labor costs can be determined from the Phased Implementation Plan, Part II of this document. The DMAC Subsystem has been conceived to minimally impact the freedom of individual data providers to make independent data management choices. Yet providers of data—Federal, regional, commercial, etc.—will nonetheless incur significant expenses creating metadata that conforms to DMAC (FGDC) standards and configuring the Data Transport software components that will make their data accessible to others. These costs are included under "Program Integration Labor."

- *Program Initial Fixed/Maintenance Costs* – This category refers primarily to non-labor costs of initiating the program in the first five years. These costs include the following:
  - Communications infrastructure including communications hardware at 30 sites (or equivalent). These sites contribute to essential DMAC infrastructure (e.g., archive centers and primary data assembly centers). The majority of sites that provide data to IOOS will not fall into this category.
  - Communications lease for the entire infrastructure.
  - Servers at a total of 30 sites, including hardware and software, and hardware maintenance after the year of installation. It is understood that participation in DMAC may in some cases require additional dedicated hardware to address issues of security and elevated performance demands. Notably the DMAC Subsystem will require one or more primary data assembly centers as fundamental infrastructure.
  - Server hardware and software, including media at five archive center sites.
  - Professional services of a Systems Integrator to (1) coordinate and manage the total hardware, software, and infrastructure definition, design, procurement, installation, integration, and maintenance; and (2) oversee *Capacity Building,* the effort in providing labor and services to data providers to enable them to reach and maintain the level at which they can participate. This effort might include technical training, system administration or help with preparing metadata, for example.
- *Outyear Recurring Costs* – These costs reflect a hardware recapitalization beginning in Year 6. “Maintenance of Custom SW (software)” reflects software maintenance and upgrades (i.e., custom programming) for the totality of deliverables at Initial Operational Capability. Systems Engineering/Integration will be provided for the life cycle maintenance of the hardware/software systems, including integration of the custom software. Archive operations consist of two staff per center per year at a cost of $100,000 per staff member. *Outyear Recurring Costs* also include
  - *Capacity Building* – (as defined under *Program Initial Fixed/Maintenance Costs)*; and
  - *Outreach* – referring to the effort to acquire new data providers, to educate the pool of possible data providers on the IOOS process and requirements, and to perform similar services as *Capacity Building*, on a limited basis.

Inflation costs are calculated based on the Real Discount Rates, published by the Office of Management and Budget. Published rates for five, seven, and nine years are 1.9%, 2.2%, and 2.5%, respectively. For this model, a rate of 2.2% was used.

Section 4. Immediate Priorities for Implementation

Implementation of IOOS has already begun. Once the IOOS Plan has been approved, participation will accelerate. Developments will be initiated at local, regional, and national levels. To support these activities, DMAC must quickly achieve a useful minimum level of functionality. The DMAC outcomes listed below are required to reach that point. They include (i) community-building and planning, (ii) standards-generating, (iii) the development of critical (currently missing) technology components, and (iv) initiating a technical oversight function by a professional systems engineering service. The activities that accompany these outcomes are described in detail in Part II, Section 2 of this phased Implementation Plan. The activities will be overseen by the DMAC Standing Committee, which will also have initial responsibilities for coordination of data management and communications among national backbone components, regional observing systems, and international collaborators.

METADATA/DATA DISCOVERY/DATA LOCATION

Outcome 1: The development of (i) initial metadata standards that will guide IOOS/DMAC data providers in the creation of metadata and (ii) agreement upon the initial organizations that will participate in metadata management for IOOS:

Task 1: Convene a community-based working group of metadata and archive experts to agree upon and document: (i) interim metadata standards, and (ii) an initial list of IOOS member organizations that will provide metadata catalog management services.

Outcome 2: The development of initial data discovery services needed for IOOS data users to identify data sets of interest.

Task 1: Convene a community-based working group of data discovery experts to agree upon the Data Discovery architecture.

Task 2: Implement a testbed based upon existing data discovery services leading to the development of an interim distributed metadata search capability.

Outcome 3: Agreement upon a technical solution to create bi-directional linkages between metadata-based DMAC Data Discovery services and (i) DMAC Data Transport and (ii) On-line Browse (and other information products).

Task 1: Conduct an applied R&D activity that will explore existing technologies capable of linking metadata-based searches to points of data and product access. This effort must address the dynamic (changeable) nature of the access points and the problems of multiple (replicated) copies of data sets that are available at different points.

DATA TRANSPORT

Outcome 4: The development of an initial semantic data model (for a restricted class of marine data) that (i) is capable of demonstrating machine-to-machine interoperability with semantic meaning and (ii) will form the foundation for further semantic data modeling leading (in the mature DMAC) to a comprehensive marine data model. This work will be conducted in close coordination with the Metadata Working Group activities.

Task 1: Convene a community-based working group of data management experts from a broad range of marine disciplines to (i) develop and document a (restricted) interim semantic data model and (ii) initiate development of a comprehensive semantic data model.

Outcome 5: The existence of three critical infrastructure components that will be needed to support Data Transport for vital classes of data: biological, GIS, and "generic" (non-standards conformant).

Task 1: Conduct a software development activity to develop a linkage from biological data accessed through the OBIS system to the DMAC Data Transport component.

Task 2: Conduct a software development activity to develop a "generic" server that would access such data as ASCII tab-delimited files.

Task 3: Conduct a software development activity to develop a linkage from data accessible through the DMAC Data Transport component into common GIS applications.

DATA ARCHIVE

Outcome 6: The existence of community-wide agreements that will establish a framework for cooperation among IOOS archive centers.

Task 1: Convene a community-based working group of archive center representatives to agree upon an initial list of IOOS partner organizations that will provide permanent archive services.

Task 2: Archive working group to agree upon an initial framework to inventory and assess the state of marine data archival in order to ensure that all irreplaceable marine data are associated with a responsible archive center.

Outcome 7: A demonstrated capability of IOOS Archive Centers to provide Data Discovery and Data Transport services using DMAC standards and protocols.

Task 1: Continue development of existing pilot projects at NODC that use DMAC standards and protocols for Data Transport and Metadata to deliver near-real-time and real-time data sets (Global Temperature-Salinity Pilot Project and Shipboard Environmental Acquisition System).

Task 2: Modernize access to data currently received in real-time at Archive Centers by deploying pilots that rapidly deliver data to users employing DMAC standards and protocols.

SYSTEM ENGINEERING APPROACH

Outcome 8: The development of well-organized documentation of the DMAC Subsystem and its initial participants needed by IOOS participants and planners.

Task 1: Engage the services of a software engineer to provide these services to IOOS.

CONCRETE GUIDANCE TO DATA PROVIDERS

(Technical guidance for data managers)

At the time of this writing (March 2004) the DMAC Subsystem is at an early stage of implementation, yet it is possible to recommend some concrete actions that may be taken by data and metadata managers and product producers that will streamline their future compatibility with IOOS. Integration of data into IOOS implies that would-be users will quickly and efficiently be able to (i) discover data through a comprehensive search; (ii) browse and visualize data through standard Web browsers; and (iii) access data from many common computer applications. It is understood that any recommendations made at this time are subject to change as the DMAC Subsystem evolves.

There are two classes of solutions for sharing data that will ensure consistency with the emerging DMAC standards, protocols, and tools:

1. Providers of certain types of data may delegate responsibility for managing these data to another entity. For example, a data provider may be able to enter into an arrangement with the NOAA National Data Buoy Center (NDBC[21]) to perform quality-control and distribute mooring data or with the U.S. GODAE Server[22] to distribute operational model outputs.

2. Providers of all types of data can make choices to manage the data in a manner that is consistent with the emerging DMAC standards, protocols, and tools. The following approaches to managing data and metadata will help ensure compatibility with emerging DMAC:

   - **Metadata management and data discovery**
     It is recommended that all data providers:
     – Create metadata that are compliant with Federal Geographic Data Committee (FGDC[23]) standards for both current and legacy data holdings and inventories.
     – Submit metadata to the NASA Global Change Master Directory (GCMD[24]) and/or the NOAA Coastal Data Development Center (NCDDC[25]) so that data sets may be easily found through an open data discovery process.

[21]contact MarineObs@noaa.gov for details.

[22]contact the U.S. Global Ocean Data Assimilation Experiment (GODAE) Server at http://www.usgodae.fnmoc.navy.mil

[23]Information on FGDC is available at http://www.fgdc.gov/metadata/metadata.html. Contact either the NOAA Coastal Data Development Center (http://www.ncddc.noaa.gov/Metadata) or the NOAA Coastal Services Center (http://www.csc.noaa.gov) for direct assistance with creating FGDC metadata

[24]see http://gcmd.gsfc.nasa.gov

[25]see http://www.ncddc.noaa.gov/Metadata

– Participate in the DMAC Metadata Working Group[26] (see Part II, Section 2, Metadata/Data Discovery Activity 1) to ensure that the special characteristics of their data will be thoroughly considered during the formulation of DMAC metadata standards.

- **Data Transport**
  Depending upon the nature of the data to be provided, it is recommended that data providers of:
  – *Gridded data* install servers providing access to their data through OPeNDAP[27] data access protocol. (OPeNDAP is an *operational*[28] component of IOOS for access to gridded data. OPeNDAP servers are available for download without licensing costs)
  – *Complex data collections in a relational data base (SQL)* make data accessible to DMAC by installing an OPeNDAP relational data base server. (OPeNDAP is a *pilot* component of IOOS for access to unstructured data collections). Full operational support for relational data bases will be developed early in the evolution of DMAC.
  – *Geographic Information System (GIS) data collections (for groups that are already participating in enterprise GIS networks)* continue pursuing these efforts. Gateways that provide translation from the GIS network protocols to the OPeNDAP protocol will be developed by IOOS.
  – *Large collections of files that comprise a single (logical) data set* install an OPeNDAP "aggregation server" or participate in a DMAC data aggregation pilot project.

  It is recommended that all data providers:
  – Participate in the DMAC Transport (Semantic Data Model) Working Group[28] (see Part II, Section 2, Data Transport, Activity 1) to ensure that the special characteristics of their data (if any) will be thoroughly considered during the formulation of DMAC data transport standards

- **Uniform on-line browse to all IOOS data**
  It is recommended that all data providers either:
  – Install a Live Access Server (LAS[29]) and notify DMAC[26] of its existence, so that it can be integrated into the community-wide data browsing environment. (LAS is a *pre-operational* component of IOOS for on-line browsing and visualization of data. LAS is available for download without licensing costs);

  or

[26] contact DMAC@ocean.us.net
[27] http://www.unidata.ucar.edu/packages/dods
[28] http://www.ocean.us/documents/docs/part_1_for_web_comment_022403.pdf
[29] http://www.ferret.noaa.gov/LAS/

– Notify the National Virtual Ocean Data System (NVODS[27]) that their data set has become available through OPeNDAP and request that it be added to the community-wide data browsing capability[28].

- **Archive**

  It is recommended that all data providers:

  – Review their current data holdings to ensure that irreplaceable data are archived at a responsible entity.

  – Contact the archive entity[29] that is responsible for their classes of data and make arrangements for archiving the data.

[27]http://www.nvods.org

[28]http://www.ferret.noaa.gov/nvods

[29]http://www.nsf.gov/pubsys/ods/getpub.cfm?ods_key=nsf94126