I. Interoperable Data Discovery, Access, and Archive

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

Executive Summary

March 2005

The National Office for Integrated and Sustained Ocean Observations
Ocean.US Publication No. 6

Executive Summary

Congress has directed the U.S. marine science communities to come together to plan, design, and implement a sustained Integrated Ocean Observing System (IOOS). IOOS is envisioned as a network of regional, national, and global systems that rapidly and systematically acquires and disseminates data and data products to serve the critical and expanding societal needs to:

- Improve predictions of climate change and weather and their effects on coastal communities and the nation;
- Improve the safety and efficiency of maritime operations;
- More effectively mitigate the effects of natural hazards;
- Improve national and homeland security;
- Reduce public health risks;
- More effectively protect and restore healthy coastal ecosystems; and
- Enable the sustained use of ocean and coastal resources.

Internationally, IOOS will be the U.S. contribution to the Global Ocean Observing System (GOOS) and the Global Earth Observation System of Systems (GEOSS).

A coherent strategy that enables the integration of marine data streams across disciplines, institutions, time scales, and geographic regions is central to the success of IOOS and other regional, national, and international ocean and coastal observing systems. The system that must be developed, while challenging, is within the scope of current information technology (IT). It can be developed by building upon existing capabilities through relatively straightforward software engineering. *The greatest challenge to enhancing marine data integration is one of coordination and cooperation among the members of IOOS and its user communities.*

Ocean.US, the IOOS national office, established the Data Management and Communications Steering Committee in the spring of 2002 to develop a detailed, phased implementation plan that will lead to an effective data management and communications (DMAC) component of IOOS, and to provide oversight during its evolution. The DMAC Plan has undergone multiple levels of review by technical and scientific experts, as well as by the broader marine environmental data supplier and user communities. It is divided into three main parts. Part I, intended for general readers, provides an overview of requirements, strategies for addressing them, and technological considerations. Part II, intended for technical readers, presents a detailed DMAC Implementation Plan in outline form. Part III, the Appendices, provides in-depth analysis of key technical topics.

This DMAC Plan is the first in a series of documents that addresses IOOS data management and communications requirements, and those of other observing systems such as the National Science Foundation's (NSF) Ocean Research Interactive Observatory Networks (ORION). This Plan pres-

ents an overview of DMAC; provides a technical focus on the issues of interoperable data discovery, access, and archive; and provides a development time line with estimated costs. As one of several Subsystems of IOOS, DMAC will be developed, implemented, operated, and enhanced according to the planning and governance procedures described in the IOOS Development Plan (www.ocean.us).

INTERNATIONAL COOPERATION

Producing global assessments and predictions of coastal ecosystem health and sea-level change, as well as addressing the other IOOS goals, requires that IOOS observations and data products be fully integrated with other national and international Earth observation efforts. Coordinated and sustained cooperation is already well established within the weather community, and the World Meteorological Organization's (WMO) World Weather Watch demonstrates the value of this international collaboration. Coordination is less well established in the ocean, ice, land, water, and climate observation communities. Nevertheless, much important work has been accomplished on the international front, and IOOS is well positioned to contribute to these efforts, especially in the area of data management and communications.

Many of the contributors to the IOOS DMAC Plan are involved with international efforts addressing global ocean and coastal observing needs. As a result, the DMAC Plan is being examined by the WMO as an early model for standards and protocol development. Effective coordination among the relevant international programs is essential to realizing a truly interoperable, global and national coastal and ocean observation framework. Expanded coordination and more formal programmatic linkages by IOOS with GEOSS, GOOS, and the WMO are needed. The DMAC Plan therefore recommends that steps be taken to address this need.

TECHNICAL ANALYSIS OF THE DMAC SUBSYSTEM

IOOS will consist of three subsystems:

- **Observing Subsystem:** remotely sensed and *in situ* measurements and their transmission from regional and national backbone platforms;
- **Modeling and Analysis Subsystem:** evaluation and forecast of the state of the marine environment based upon assimilated measurements; and
- **Data Management and Communications Subsystem (DMAC):** information technology infrastructure such as national backbone data systems, regional data centers, and archive centers connected by the Internet, and using shared standards and protocols.

Figure 1 illustrates the flow of data from observation platforms to intermediate components (e.g., modeling centers and archive centers), to generators of information products, and finally to end users. The DMAC Subsystem is a framework for integration of large and small independent and heterogeneous data management and communications systems. Most of the planning and investment within the DMAC Subsystem for data management *per se* lies outside the scope of the DMAC Plan. The thousands of individual organizations that comprise IOOS will continue to manage their data in the manner they deem most appropriate to their individual missions, but through DMAC they can broaden the impact of their data, serve a larger community, and contribute to long-term data archives that will benefit generations to come.

IOOS Observing Subsystem elements are managed by regional, national, and international entities. Measurements made by these elements are highly heterogeneous. A wide range of data distribution and dissemination systems, including the WMO'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 (PDA&QC) occur. **The systems that convey data from sensors to primary data centers/sites lie outside the scope of the current DMAC Plan.** PDA&QC processes typically lie at the interface between the Observing Subsystem and the DMAC Subsystem. In general, some form of PDA&QC is required before ocean observations and measurements can be used.

The DMAC Subsystem will include a data and communications infrastructure that consists of a suite of components—standards, protocols, facilities, software, and supporting hardware systems. The design and planning for the DMAC framework will emphasize continual, smooth evolution. The components upon which the architecture is built will, themselves, be an evolving collection. New components will be introduced; recognized components will be advanced; obsolete components will be removed. A significant level of duplication of function among components will be tolerated as a necessary consequence of a continuously evolving system. The DMAC Standards Process will define the manner by which the level of maturity of components will be designated: R&D, pilot, pre-operational, and operational. At the outset, no formal DMAC standards process exists. Yet, there is an imperative to provide immediate guidance to would-be data providers. To address this need, the DMAC Plan includes preliminary recommendations for (1) the maturity designations of certain named components that are viewed as essential to the initial architecture and (2) a roadmap leading to rapid designation of other initial components by community-based working groups.

The DMAC Plan provides a roadmap to achieve the following functionality for the DMAC Subsystem: (1) IOOS-wide descriptions of data sets (**Metadata**); (2) the ability to search for and find data sets, products, and data manipulation capabilities of interest (**Data Discovery**); (3) the ability to access measurements and data products from computer applications across the Internet

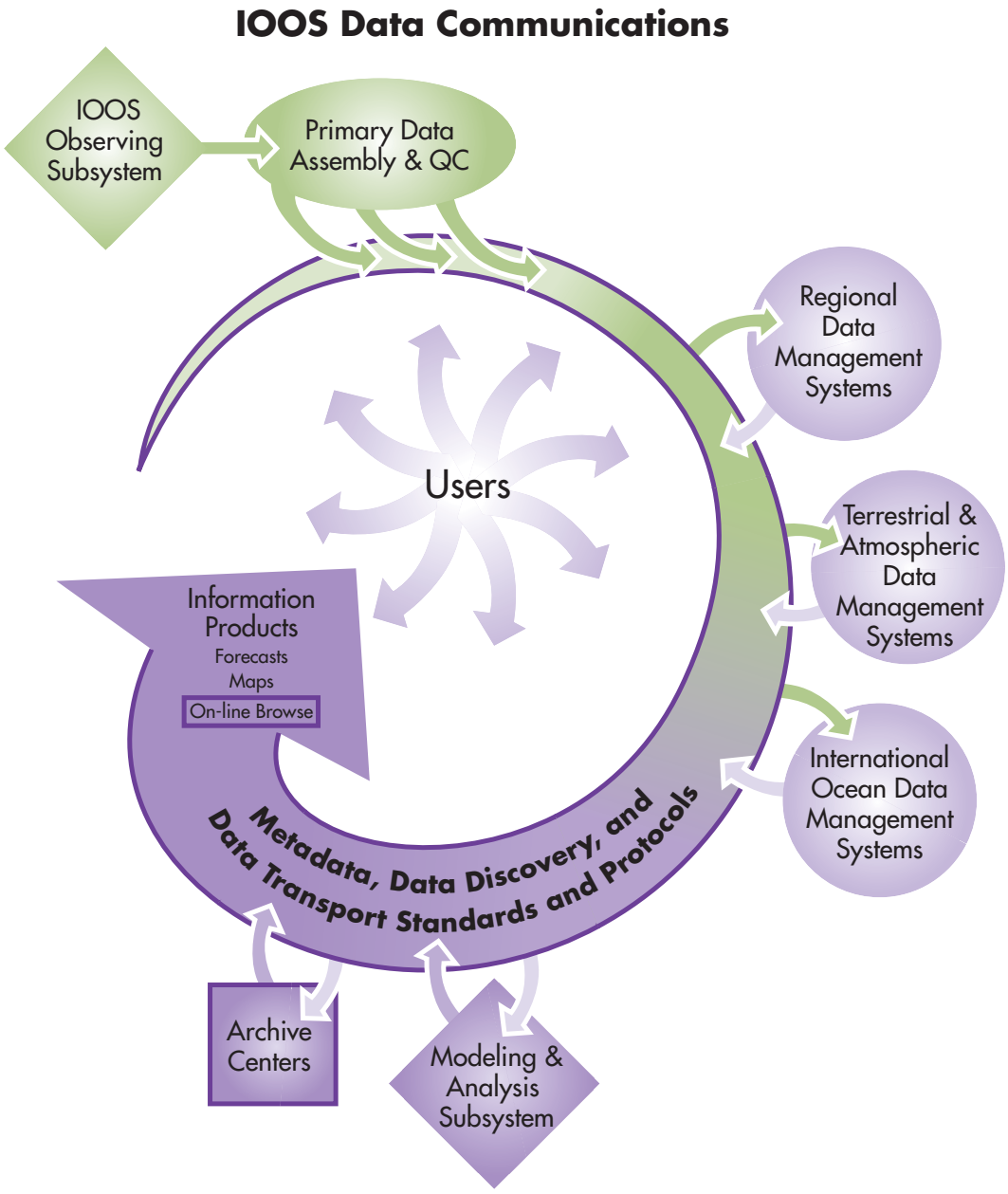

Figure 1. Solid outlines indicate the elements of the IOOS Data Communications framework, which are detailed in the DMAC Plan. The arrows flowing outward from users indicate the feedback and control mechanisms through which users ultimately direct the functioning of all parts of the system. Note that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

(**Data Transport**); (4) the ability to quickly evaluate the character of the data through commonly-available web browsers (**Uniform On-line Browse**); and (5) secure, long-term data storage (**Data Archive**).

DMAC Metadata will be based upon standard vocabularies and content specifications. The metadata specifications will support the publication of Federal Geographic Data Committee (FGDC)-compliant records. Agreement on vocabulary and content does not yet exist. The DMAC Plan recommends establishing an interdisciplinary Metadata Working Group to provide recommendations on a common vocabulary and initial metadata content for IOOS data. DMAC metadata will support data discovery capabilities that complement and extend the publicly accessible search capabilities that are available today through web search engines such as Google®. The DMAC Plan recommends that the **Data Discovery** architecture be determined by a working group that includes representatives from existing metadata management facilities and other metadata experts.

Underlying **DMAC Data Transport** is the unifying vision of DMAC web services. Through web services all types of client applications—for example, tools for end-users, modelers, and planners; and value-added marine information web sites—can access data from the broad range of IOOS data suppliers (servers). The methods by which client applications access web services remain uniform despite the servers being layered upon various legacy data management systems, developed with diverse programming languages, and run under different operating systems. DMAC web services will provide the means to connect IOOS to data management systems operated by international marine data partners and by partners in other disciplines such as meteorology.

Both the Open-source Project for a Network Data Access Protocol (OPeNDAP) and the Open Geospatial Consortium (OGC) support web services of relevance to DMAC Data Transport. OPeNDAP, the web service that underlies the National Virtual Ocean Data System (NVODS), is a discipline-neutral transport protocol that conveys data, metadata, and structure without regard to the scientific interpretation of the data. The DMAC Plan recommends the designation of OPeNDAP as an initial "operational" component for transport of gridded data, and recommends that a "pilot" activity be undertaken to explore the delivery of non-gridded data using OPeNDAP (See IOOS Development Plan at www.ocean.us for definitions of system component maturity). The DMAC Plan further recommends that two OGC web services, the Web Feature Service (WFS) and the Web Coverage Service (WCS) be examined for incorporation into the DMAC data transport suite.

The DMAC Plan anticipates that many IOOS data providers will host metadata-enabled, open source, or commercial on-line browse tools for end users. In addition, the DMAC Subsystem must provide a system-wide view of IOOS data—the ability to visualize and assess all IOOS data in a

uniform manner. The **Uniform On-line Browse** capability of DMAC will use the Data Transport web services for access to IOOS data. The DMAC Plan recommends the designation of the Live Access Server (LAS), which provides browsing capabilities with NVODS, as an initial "pre-operational" component for system-wide Uniform On-line Browse. The DMAC Plan further recommends that OGC-compatible GIS web servers likewise be examined as candidates for DMAC Uniform On-line Browse clients.

The **DMAC Data Archive** component will be assembled from existing and new marine data archive facilities. The DMAC Plan recommends that, to be recognized as an official partner in the IOOS Data Archive enterprise, a facility must enter into a formal agreement(s) stipulating that they perform archive and access functions using DMAC standards and protocols and conform to IOOS Data Policy. The DMAC Plan further recommends that a community-based, interdisciplinary working group of archive specialists and advisors initiate an orderly strategy to determine DMAC Data Archive policies and procedures, and to ensure that designated archive facilities exist for all IOOS data.

The **IOOS Modeling and Analysis Subsystem** will provide numerical (digital) data products through computer modeling and analysis of real-time and historical data collections. Planning for the numerical data products that IOOS must produce lies outside the scope of the DMAC Plan. It will be handled elsewhere within the IOOS framework.

Information products, such as text and verbal forecasts, maps, and scientific plots, will be generated throughout the IOOS network. It is understood that the private sector will be a primary producer and distributor of value-added information products within IOOS, particularly to meet the specialized needs of targeted user groups.

IT SECURITY

IOOS is being deployed in a distributed, heterogeneous information technology (IT) environment with web services as the eventual target architecture. The IOOS therefore faces a number of security challenges that include:

- Participants joining the IOOS enterprise are accustomed to operating under diverse security guidelines and cultures that may not conform to required federal IT security practices.
- Agreement on and compliance with a common security policy must be reached across multiple heterogeneous systems.
- "Desktop-level administrators" must be able to understand and implement IOOS security policies and measures, and will often be in environments where IT management resources are limited.

- Many legacy applications will be incorporated into IOOS that will be web-enabled, but may not have been originally designed for exposure to the public Internet or for use in a structured IT security environment.

A thorough treatment of this topic is not possible at this time due to time and resource constraints. Therefore, it is recommended that a community-based working group on IOOS IT Security be established to develop an IOOS security policy, and to provide more specific guidelines for IOOS participants on implementation.

GOVERNANCE

The DMAC Subsystem operates within the context of the overall IOOS governance mechanisms described in the IOOS Development Plan. At the First Annual IOOS Implementation Conference (August/September 2004), the Ocean.US Executive Committee (EXCOM) agencies and the emerging Regional Associations (RAs) endorsed a DMAC governance strategy to ensure that the development and implementation of the DMAC Subsystem is coordinated closely with, and leverages upon, related activities in the federal agencies and other national, regional, and international Earth observing systems. This strategy includes: a community-based **DMAC Steering Team** (DMAC-ST) to coordinate and oversee the evolution of DMAC standards and best practices; **Expert Teams and Working Groups** to support the DMAC-ST; and a federal-government-only **Implementation Oversight Working Group** (IOWG) to coordinate DMAC implementation within the federal agencies. It is also recommended that the National Federation of Regional Associations (NFRA) establish a DMAC subcommittee to oversee and facilitate coordination, communications, and data and technology exchange at the regional level. IOOS stakeholders will be urged to participate in the DMAC planning and assessment activities to ensure that current and future community needs and priorities are addressed.

The DMAC Plan recognizes that data interoperability is largely a reflection of the ability of the marine community to successfully agree upon and use standards. Therefore, the DMAC Plan recommends that Ocean.US convene (or participate in) a working group to investigate and recommend a process for the development of future community data and metadata standards. The process should include guidelines that maximize the compatibility of new standards with pre-existing ones, and uniform review procedures for standards.

The DMAC Subsystem plays an essential role in IOOS user outreach by providing Internet portals for user feedback, and mechanisms for automated collection and analysis of system performance metrics.

COSTS

The DMAC Plan provides first-order estimates of those expenses associated with development and implementation of the policies, standards, protocols, and tools comprising the DMAC Subsystem. These cost estimates do not reflect any formal review or endorsement by the participants or agencies supporting IOOS deployment. Further, it should be noted that these estimates do not include costs resulting from growth in data services that would occur irrespective of IOOS development; nor do they include the costs of sensors, data telemetry, modeling and applications, and most product-development activities. These estimates also do not include those costs associated with the implementation of DMAC standards within the regions, or with anticipated capitalization and maintenance/operating costs likely to be incurred.

The DMAC Plan calls for the initiation of the full DMAC Subsystem over a five-year period at a cost of $82 M. The initiation costs include the development of core standards, protocols, and tools ($28 M); costs of hardware, software, networking capacity, data archiving 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 (to Year 10) total an additional $85 M. All cost estimates provided in the DMAC Plan include an inflation factor of 2.2 percent per year. Substantial new funding for IOOS is not anticipated until fiscal year 2007 (FY 07), **yet a minimally functioning DMAC Subsystem must already be in place to support the initial growth in IOOS** (and other ocean observing systems) measurements, modeling, and usage at that time. Thus, the DMAC Plan includes tasks totaling $2.1 M during FY 05 to FY 06 that are deemed to be very high priorities for immediate implementation to prepare for FY 07 demands on the Subsystem.

HIGH-PRIORITY RECOMMENDATIONS

Implementation of IOOS and other regional, national, and international ocean and coastal observing systems has already begun, and will continue to accelerate over the coming years. To support these activities, the DMAC Subsystem must quickly achieve a useful minimum level of functionality. The DMAC Plan recommends the following steps as high priorities for implementation:

1. Initiate working groups and/or applied R&D activities to address: (a) development of metadata, including vocabulary, content, and discovery components; (b) assessment and selection/development of missing data transport components; and (c) community building and partnerships as outlined in Part II of the DMAC Plan.

2. Engage software engineering services to initiate development of well-organized documentation, centralized coordination of assistance to IOOS data suppliers and product generators, and software life-cycle planning for the critical components of the DMAC Subsystem.

3. Ocean.US should establish (1) a permanent DMAC Steering Team with representation from the full IOOS community responsible for technical planning and recommending standards and (2) an all- federal group responsible for allocation of resources to address DMAC priorities and implementation.

4. Data providers should examine and where possible implement the Part I, "Concrete Guidance to Data Providers." This section offers guidance to help coordinate the implementation of initial DMAC functionality, while ensuring compatibility with future IOOS standards. Highlights of that guidance include: (a) create FGDC-compliant metadata; (b) enable data discovery by sharing metadata with designated IOOS metadata facilities; (c) make gridded data accessible through the OPeNDAP data access protocol; (d) implement on-line browse solutions (using Live Access Server, GIS web servers, or others); and (e) ensure that designated IOOS archive centers have plans in place for long-term archiving of the contributed data.