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

I. Interoperable Data Discovery, Access, and Archive

Executive Summary

May 10, 2004

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 needs of environmental protection, public health, industry, education, research, and recreation. IOOS will also be the U.S. contribution to the international Global Ocean Observing System (GOOS).

A coherent strategy that can integrate marine data streams across disciplines, institutions, time scales, and geographic regions will be 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 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 members of the IOOS network and the 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 establishment. The DMAC Plan, presented here, 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 Plan is the first in a series of documents that will address 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). It presents an overview of DMAC; provides a technical focus on interoperable data discovery, access, and archive; and presents a detailed development time line with costs. As one of several subsystems of the larger IOOS, DMAC will be developed, implemented, operated, and enhanced according to the planning and governance procedures described in the IOOS Implementation Plan (www.ocean.us).

TECHNICAL ANALYSIS OF THE DMAC SUBSYSTEM

IOOS will consist of three subsystems:

- OBSERVING SUBSYSTEM (remotely sensed and *in situ* measurements and their transmission from platforms);
- MODELING AND ANALYSIS SUBSYSTEM (evaluation and forecast of the state of the marine environment based upon measurements); and
- DATA MANAGEMENT AND COMMUNICATIONS SUBSYSTEM (DMAC) (the integrating component).

DMAC provides the links within and among the other IOOS subsystems. DMAC will also provide the means to connect IOOS to international ocean data management systems and to significant data management systems in other disciplines, such as atmospheric and terrestrial sciences. Figure 1 depicts the data flow relationships among the DMAC Subsystem and other IOOS components and partners.

IOOS Observing Subsystem elements are managed by regional, national, and international entities. The measurements made by these elements are highly heterogeneous, and 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. Primary Data Assembly and Quality Control (PDA&QC) processes 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 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.**

The DMAC Subsystem will include a data and communications infrastructure that consists of standards, protocols, facilities, software, and supporting hardware systems. DMAC will support: (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 (Data Transport); (4) the ability to quickly evaluate the character of the data through common Web browsers (Online Browse); and (5) secure, long-term data storage (Data Archive).

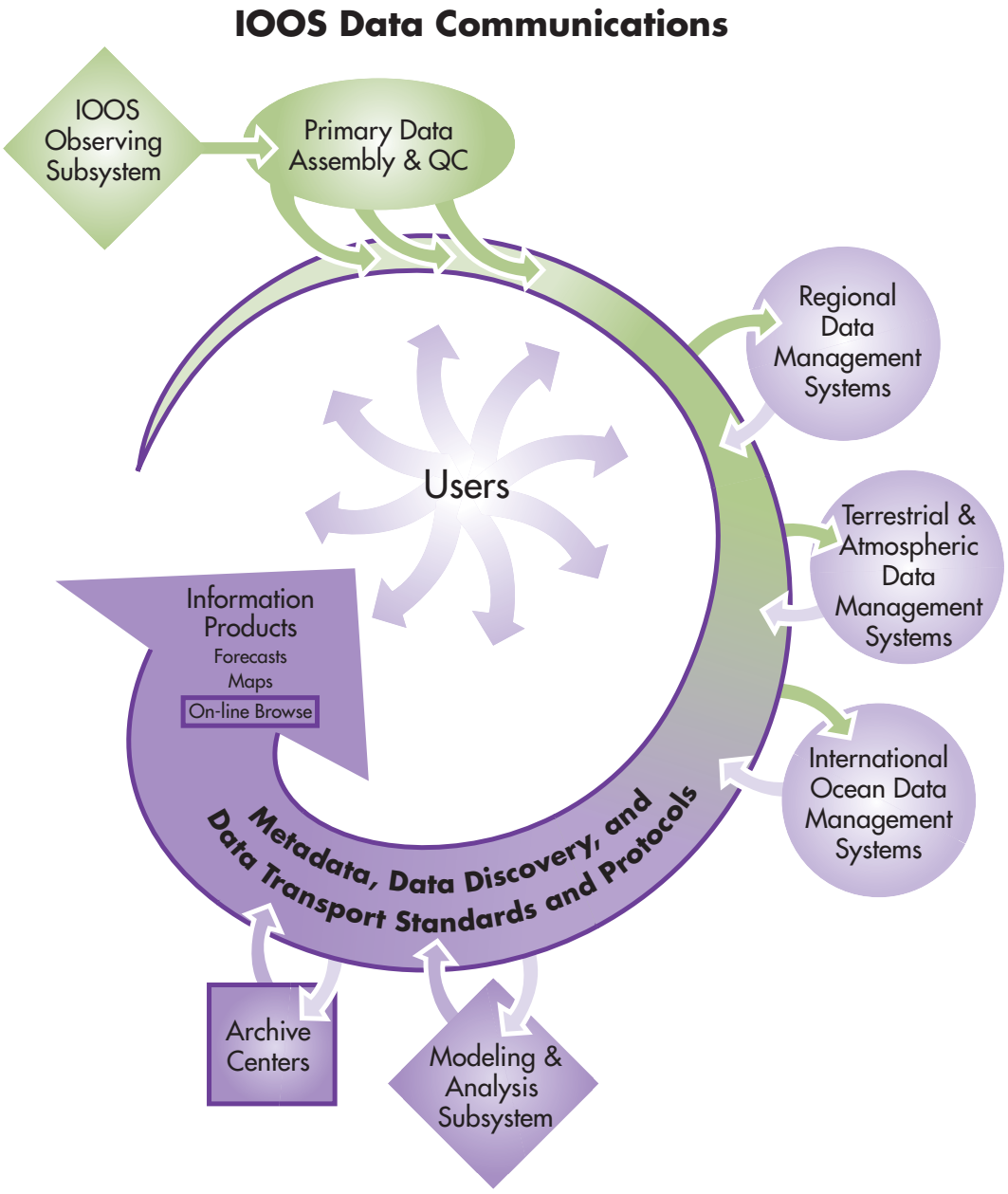

Figure 1. Solid outlines indicate the elements of the IOOS Data Communications framework, which are detailed in the DMAC Implementation Plan. The arrows flowing outward from users indicate the feedback and control mechanisms through which users ultimately direct the functioning of all parts of the system.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

DMAC Metadata will be based upon Federal Geographic Data Committee (FGDC)[1] standards. Discipline-specific FGDC profiles will be developed, adopted, and/or harmonized by interdisciplinary working groups drawn from the marine data community. DMAC Data Discovery capabilities will complement and extend the publicly accessible search capabilities that are available today through Web search engines such as Google®. The Data Discovery architecture will be determined by a working group that includes representatives from existing metadata management facilities and other metadata experts.

DMAC Data Transport will be built upon a suite of "web services." A web service is a standardized protocol encoded in eXtensible Markup Language (XML) and transported by the Hyper Text Transfer Protocol (HTTP—the "Web"), through which one computer can request data and/or computations from another. The web services approach has been selected because it has only minimal impacts on the data management choices made by contributors of IOOS data, and is broadly adaptable to existing and new (client) applications. DMAC web services will connect data management systems operated independently by Regional Associations, state and Federal government entities, academic projects, commercial and international partners, and within other disciplines.

The DMAC Plan designates the OPeNDAP data access protocol, the web service that underlies the National Virtual Ocean Data System (NVODS), as an initial "operational"[2] component for transport of gridded data; and as a "pilot" [3] component for the delivery of non-gridded data. OPeNDAP is a discipline-neutral transport protocol that conveys data, metadata, and structure without regard to the scientific interpretation of the data. A community-based, interdisciplinary working group will be convened to develop a consistent, geospatial semantic data model that will allow the scientific meaning of the data to be fully captured.

Effective management of IOOS requires a basic browsing and visualization capability that extends across the full breadth of IOOS data. The browsing capability will provide geo- and time-referenced graphics and readable tables suitable for the evaluation of IOOS data through standard Web browsers. The On-line Browse capability of DMAC will use the Data Transport web services for access to IOOS data. The DMAC Plan designates the Live Access Server (LAS), which provides browsing capabilities with NVODS, as an initial "pre-operational"[4] component for On-line Browse.

---

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

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

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

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

The Data Archive component of DMAC will be assembled from existing and new marine data archive facilities. To be recognized as an official partner in IOOS Data Archive, these facilities must each enter into formal agreements stipulating that they will perform archive and access functions using DMAC standards and protocols and conform to IOOS Data Policy. A community-based interdisciplinary working group of archive specialists and advisors will 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 permit the generation of numerical (digital) data products through computer modeling and analysis of 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 an essential partner in IOOS for the production and distribution of specialized, value-added information products and applications that address the needs of particular user groups.

MANAGEMENT

The DMAC Plan recommends that Ocean.US establish a permanent DMAC Standing Committee to provide oversight and coordination of long-term planning, development of data policies, system engineering activities, user outreach, and the evolution of DMAC standards. The membership of the DMAC Standing Committee should balance technical and management expertise across the disciplines encompassed by IOOS with appropriate representation from Federal, state, and local government agencies, academia, user communities, the private sector, the National Federation of Regional Associations, and the international community.

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 it is recommended 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 cost estimates provided in the DMAC Plan reflect only those expenses associated with development and implementation of the policies, standards, protocols, and tools comprising the DMAC Subsystem. They 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. 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, 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 (to Year 10) total an additional $85 M. All cost estimates provided in the DMAC Plan include an inflation factor of 2.2% per year. 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 (and other ocean observing systems) measurements, modeling, and usage at that time. Thus, the DMAC Plan includes tasks totaling $2.1 M during FY04-06 that are deemed to be very high priorities for immediate implementation to prepare for FY07 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 and data model standards; (b) development of missing data transport components; and (c) community building and partnerships.
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 a permanent DMAC Standing Committee.
4. Data suppliers who wish to begin participation in IOOS immediately should adopt the initial DMAC Plan recommendations: (a) creating FGDC compliant metadata; (b) enabling data discovery by sharing metadata with designated IOOS metadata facilities; (c) making data available through the OPeNDAP data access protocol and/or participating in enterprise GIS systems; (d) making data browsable on-line through the Live Access Server; and (e) ensuring that designated IOOS archive centers have plans in place for long-term archiving of the contributed data.