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

I. Interoperable Data Discovery, Access, and Archive

Part II. Phased Implementation Plan for DMAC

May 10, 2004

Contents

Section 1. Functional Requirements

This document states high-level functional requirements for the Data Management and Communications (DMAC) Subsystem of the Integrated Ocean Observing System (IOOS). The intended use of this document is to supplement the white papers that describe the DMAC subsystems (Part III of this document) and to rephrase the notions into the formal terminology of a requirements specification. The use of the term "requirements" is consistent with that used by the International Council on Systems Engineering.[1] As described in the Handbook, this section also is intended "to establish a database of baseline system requirements derived from the source, to serve as a foundation for later refinement and/or revision by subsequent functions in the Systems Engineering process and for a non-ambiguous and traceable flow down of source requirements to the system segments."[2] This section includes requirements of the following types:

- Program requirements
- Mission requirements
- Customer specified constraints
- Functional requirements
- System requirements
- Interface, environmental, and non-functional requirements
- Unclear issues discovered in the requirements analysis process

This section is written at a level that primarily addresses mission and customer requirements, along with high-level functional and system requirements. It is intended that this section serve as an asset to be used in subsequent engineering efforts to articulate the detailed system and interface requirements that are required to develop a DMAC.

GENERAL REQUIREMENTS

1. IOOS DMAC Vision

1.1. The DMAC Subsystem of the IOOS will knit together the distributed components of IOOS into a nationwide whole that functions as a unified component within the international GOOS framework. The vision for the DMAC Subsystem is not limited to the ingesting and archiving of observations; it includes the data and communications components needed to move data among systems and users in a distributed environment. The DMAC Subsystem will be required to link observations collected from a broad range of platforms: buoys, drifters, autonomous vehicles, ships, aircraft, satellites, and cabled instruments on the sea floor. Observations

[1]Systems Engineering Handbook, International Council on Systems Engineering, 2000.
[2]Ibid.

may be point measurements, continuous measurements, or imagery and variables may be biological, geological, chemical, physical, or abstract. The many millions of individual measurements anticipated to be obtained daily by the sensor networks will be transmitted (in real-time, near-real-time, and delayed modes) directly to end users, as well as to the applications and data-assimilating models that process these measurements into maps, plots, forecasts, and other useful forms of information.

1.2. While the DMAC vision recognizes that data products, rather than raw data, are typically required by users, the development of most data products will be the responsibility of the Applications, Modeling, and Product Services Subsystem of the IOOS. The requirements of the DMAC with respect to product generation are as follows:

1.2.1. to ensure that the needs of product generators are met for timely delivery of quality-controlled data;

1.2.2. to provide accurate and thorough metadata accompanying the data;

1.2.3. to provide a uniform guaranteed minimum level of geo- and time- referenced graphical browse capability for all classes of data.

1.3. The guarantee of assured data discovery and minimal browsing capability depend upon descriptive metadata, ensuring that the data are readily intelligible to users.

2. DMAC Overall Functional Requirements

2.1. Data transport. The DMAC shall provide capability for the collection/transmission of data from sensor subsystems at entry points where the data become available using DMAC standards and protocols either on the Internet or a supplied IOOS backbone, to assembly centers, users, and archive centers in real time and delayed mode, for operational, research, and product generation applications.

2.2. Quality control. The DMAC shall provide a mechanism for assuring that data are of known, documented quality. QC operations are a partnership among data observation/collection components, processors, analysts, other users, and the DMAC.

2.3. Data assembly. The DMAC shall provide mechanisms for aggregation and buffering of data streams over useful spans of time and space. Data assembly allows users to more easily exploit real-time data, especially data from distributed sensor arrays.

2.4. Product generation. Products include data products such as assimilation-friendly, real-time measurements, model nowcasts and forecasts, GIS layers and climatological reference fields; graphical information products such as scientific plots and maps; and text information products such as written forecasts and numerical tables. The DMAC will provide a minimal level product-generation capability, only—the guarantee of a uniform, interactive, geo- and time-referenced browse capability suitable for quick evaluation of data by IOOS scientists. Most

product generation is the responsibility of the IOOS Modeling, Data Assimilation Subsystem and the value-added information product producers that will address the needs of specialized end-user groups.

2.5. Metadata management. The DMAC shall provide simple, clear guidelines and extensible standards for metadata; ensure that the linkages between data and metadata are maintained with great reliability; provide for communication of metadata among components of the system; provide training and tools to increase end users' and data providers' capacity in metadata generation and management.

2.6. Data archeology. The DMAC shall directly or indirectly facilitate activities to rescue, digitize, and provide access to legacy/historical data sets; retrieve data in danger of loss due to deteriorating media, out-of-date software, not in digital format, etc.

2.7. Data archival. The DMAC shall provide for the long-term archive and stewardship of IOOS data sets; conform to national archive standards, as well as IOOS standards and user requirements.

2.8. Data discovery. The DMAC shall provide a means for determining what data are available within the IOOS based upon queries that may be issued by users or by other machines. Data Discovery shall be seamlessly integrated with data and metadata access functions provided by the Data Transport and Metadata Management components, respectively.

2.9. Administrative functions. The DMAC shall provide oversight mechanisms to ensure the proper functioning and smooth evolution of IOOS. These include fault detection and correction, security, monitoring and evaluation of system performance, providing for system extensibility, establishing and publicizing policies for data availability, soliciting and responding to user feedback, and establishing and maintaining international linkages.

3. Participating Activities

3.1. The DMAC shall provide transport, access, and archival capabilities for TBD Regional Data Centers.

3.2. The DMAC shall provide transport, access, minimum browse, and archival capabilities for TBD Data Assembly Centers.

3.3. The DMAC shall provide transport, access, minimum browse, and archival capabilities for TBD Modeling Centers

3.4. The DMAC shall provide transport and access capabilities for TBD Archive Centers.

3.5. The DMAC shall provide transport, access, and minimum browse capabilities for TBD value-added product generators.

4. Infrastructure/Communications

4.1. Infrastructure

4.1.1. The DMAC shall leverage existing or deploy dedicated IOOS data servers at TBD locations, including up to all of the following: Regional Data Centers, Data Assembly Centers, Modeling Centers, and Archive Centers.

4.1.2. The DMAC shall leverage existing or provide aggregate storage as follows:

4.1.2.1. Regional Data Centers

4.1.2.1.1. Online - TBD

4.1.2.1.2. Near-line (e.g., online tape silo) - TBD

4.1.2.1.3. Offline - TBD

4.1.2.2. Data Assembly Centers

4.1.2.2.1. Online - TBD

4.1.2.2.2. Near-line (e.g., online tape silo) - TBD

4.1.2.2.3. Offline - TBD

4.1.2.3. Modeling Centers

4.1.2.3.1. Online - TBD

4.1.2.3.2. Near-line (e.g., online tape silo) - TBD

4.1.2.3.3. Offline - TBD

4.1.2.4. Archive Centers

4.1.2.4.1. Online - TBD

4.1.2.4.2. Near-line (e.g., online tape silo) - TBD

4.1.2.4.3. Offline - TBD

4.2. Communications

4.2.1. The DMAC shall leverage existing communications capabilities or provide dedicated broadband networks between/among the Regional Data Centers, Data Assembly Centers, Modeling Centers, and Archive Centers.

4.2.2. Data Providers for the IOOS will use existing Internet capacity to push data holdings to the Regional and National Backbone Data Centers.

5. Technology Infusion

5.1. The DMAC shall develop a plan to address technology infusion. The plan shall include mechanisms for member-provided technology infusion, as well as that which is centrally funded and maintained. The emphasis will be on integration, compatibility, and interoperability among all parties participating in the IOOS.

5.2. The plan shall include evolving mass storage technology.

5.2.1. The plan shall include strategies for storage media migration.

5.2.1.1. Current archive systems are based on magnetic tape cartridges, which typically have a three to five year life cycle, and are approaching a petabyte in size. These systems will grow and the rate of increase will accelerate. This growth can be accommodated in the Archive System, but will require increases in facilities infrastructure and support.

5.3. The plan shall consider new technologies in networks, computing systems, and evolutions in software.

5.4. The plan shall account for the following categories of technology changes.

5.4.1. Technology Upgrades – A change that incorporates the next generation product or product upgrade to an existing technology or component which improves overall system functionality.

5.4.2. Technology Refreshers – A change that incorporates a new product to avoid an ensuring end of life or product/COTS obsolescence, or to correct a problem identified via a customer.

5.4.3. Technology Insertion – A change that incorporates a new product or function capability which is a result of industry growth or advanced development.[3]

6. Other General System Requirements

6.1.1. The DMAC as a whole shall be extensible in terms of function, volume, capacity, and throughput.

6.1.2. The DMAC shall provide access to data in a manner that is (largely) transparent to the user.

6.1.3. The DMAC shall not adversely impact existing data access methods or systems of the data providers.

6.1.4. It is a goal that the DMAC will not require data repositories to reformat their holdings to tie into the system.

6.1.5. Interfaces to data repositories may reside at any location that has network connectivity with the application and the data repository.

6.1.6. The DMAC shall provide a backward-compatible, version-controlled software environment.

6.1.7. The DMAC shall provide for the generic treatment of data sources isolating the requesting client from specific representations, unique request semantics, and protocols.

6.1.8. The DMAC shall make data available in multiple forms including the data's native form.

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

6.1.9. The DMAC shall offer a cross-language and cross-platform data access mechanism that is independent of the data repository.

6.1.10. The DMAC shall enable the abstraction of encoding and transmission mechanisms and allow transparent distributed access to data using multiple protocols.

6.1.11. The DMAC shall provide access to all types of data: physical, chemical, biological, and geological.

6.1.12. The DMAC shall support system synchronization to permit multiple users access to the same database simultaneously.

DATA COMMUNICATIONS INFRASTRUCTURE AND ARCHIVAL

The DMAC Subsystem is envisioned to consist of a Data Communications Infrastructure (standards, protocols, and tools for Metadata, Data Discovery, Data Transport, and On-line Browse) and an Archival capability. Figure 1 shows the interfaces through which Data Discovery functionality is achieved within the DMAC Data Communications Infrastructure. Numbers in italics refer to the sections of the requirements that reflect the drawing portion.

1. Metadata/Data Discovery (MD) Requirements

(MD – Metadata; MMS – Metadata Management System; MC – Metadata Catalog)

1. Nature of Metadata

1.1. The IOOS MD shall be supplied using the guidelines established by the Federal Geographic Data Committee (FGDC) augmented by any applicable supplemental profiles.

1.2. The DMAC shall provide the capability to deliver metadata along with data delivery.

1.3. The MMS shall provide a mechanism to ensure that metadata found during data discovery are up to date, consistent, and understandable.

1.4. The MMS shall provide mechanisms for extensibility of the metadata.

1.5. The MD shall provide a framework for data versioning, data lineage tracking, and information citations.

1.6. The MD shall provide a framework for both semantic and syntactic metadata.

1.7. The MC shall provide a metadata query mechanism that supports access through a programming interface to any/all metadata fields.

1.8. The MMS shall support multiple standards that exist today and be able to extend beyond those to include expected future metadata standards.

1.8.1. Existing standards: FGDC; Biological Profile; Shoreline Profile, TBD

1.8.2. Possible future standards: TBD

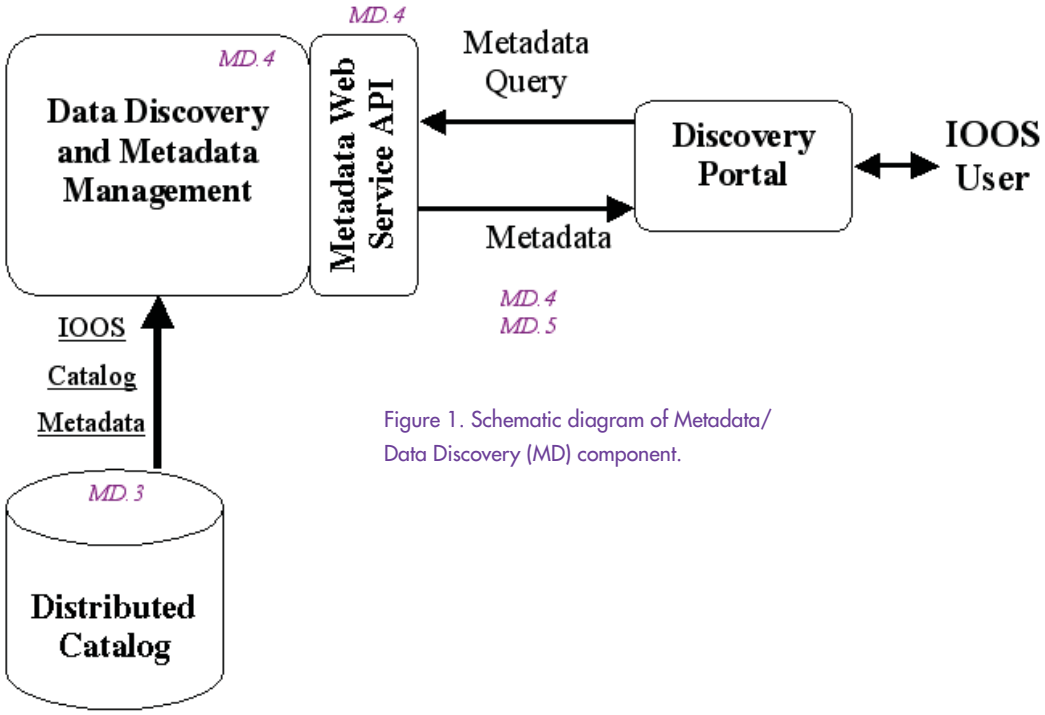


Figure 1. Schematic diagram of Metadata/ Data Discovery (MD) component.

2. Metadata Management System. The IOOS will include a master metadata management system.

2.1. The MMS shall be implemented as a distributed system that connects to all DMAC-compliant metadata holdings within the ocean community.

2.2. The MMS shall provide the capability for data providers to manage their metadata within a local system or through a centralized system via remote-access capabilities.

2.2.1. The MMS shall not require the data provider to maintain duplicate copies of metadata in two or more systems.

2.2.2. The MMS shall support a linkage between data discovery and data access that an application may utilize transparently to access both remote and local data via the DMAC Data Transport (DT).

2.3. The MMS shall include mechanisms to generate, validate and maintain metadata.

2.4. The MMS shall include a set of TBD controlled vocabularies for items such as keywords, entities and attributes, units, and other items to be determined.

2.5. The MMS shall provide support for parent/child metadata.

2.6. The MMS shall provide a mechanism for validation and approval of metadata.

2.7. The MMS shall include an automated metadata maintenance capability for checking URL links and any additional information within the metadata record that can be automated.

2.8. The MMS shall include mechanisms to facilitate the generation of metadata as close as possible to the collection and/or generation of the source data.
2.9. The MMS shall provide automated tools for versioning and configuration management of metadata.
2.10. The MMS shall provide a mechanism to access existing metadata servers to promote harvesting metadata.

3. Metadata Catalog. The MMS shall include a metadata catalog.
3.1. The implementation of the metadata catalog is TBD, but it is a requirement that the collective holdings of metadata shall comprise a distributed catalog. The implementation shall provide for integration of all such distributed sub-catalogs.
3.2. The catalog shall provide a capability to generate metadata records from self-describing data sources in which metadata and data have been integrated.
3.3. The catalog contents shall include items that will be used for discovery.
    3.3.1. The catalog shall provide access control of metadata records, for maintenance and for viewing and searching on those records.
    3.3.2. The catalog shall allow a catalog search from public search engines.

4. Search/Query Mechanism. The IOOS shall include a search/query mechanism.
4.1. The search interface shall search the MC for records that meet user-defined criteria.
4.2. End users and data providers can search for specific data sets.
4.3. End users and data providers can browse metadata about IOOS data holdings.
4.4. Automated agents can search for data.
4.5. The MC shall include a stable, documented-defined application programmers' interface (API) and a defined access protocol.
4.6. The search system shall support TBD types of actual data searches along with metadata searches.
4.7. The search system shall provide the following types of searches:
    4.7.1. Full text and fielded searches
        4.7.1.1. Controlled vocabulary
        4.7.1.2. Free-text searches
            4.7.1.2.1. Single or multiple word searches
            4.7.1.2.2. Boolean operators on multiple words
            4.7.1.2.3. Thesauri to support text searches
    4.7.2. Geospatial search
    4.7.3. Temporal search
    4.7.4. Thematic search
    4.7.5. Parameter search

4.7.6. Taxonomic information
4.7.7. Browsing by thematic areas
4.7.8. Iterative/refinement searches

4.8. The system must be extensible to support other specific searches as required by the system, such as search by data quality or native format

4.9. Select - The Select functionality refers to those capabilities that allow an end user or data provider to examine data sets revealed from the data search and then choose sets of interest for downloading, on-line browse, or access via the DMAC Data Transport mechanism.
4.9.1. There shall be a user interface for allowing the user to select items for downloading. This will be referred to as the selection interface.
4.9.2. The selection interface shall display and accept selection requests for all data sets from the catalog software that meet the search criteria specified by the user in the search interface.
4.9.3. For each data set returned from a search of the catalog, the selection interface shall display the data set title and relevant metadata including spatial and temporal coverages and a method for viewing the metadata from that data set.
4.9.4. The selection interface shall provide a graphical means of viewing a thumbnail of each data set received from the catalog search.
4.9.5. The selection interface shall allow the user the ability to select from the items returned from the search and/or perform subsequent subsetting searches of the returned items.

4.10. Data set metadata shall be obtainable in multiple formats including both machine-readable XML and human-readable text.

5. Portal

5.1. Access to data and metadata shall be provided through the Internet via a portal. A portal is an Internet presence (e.g., web site) that redirects the user (possibly transparently) to a larger set of access points.

5.2. It is a system goal that graphical user interfaces (GUIs) shall be simple to use for a broad spectrum of users.

5.3. The IOOS portal will consist of an entry point (a Web "home page"), hierarchically lower level entries (other pages), and links to areas or functions within the IOOS.

5.4. There shall be simplified versions of the portal suitable for incorporation into non-IOOS Web sites for purposes of offering the capability to search IOOS data.

5.5. The portal shall provide mechanisms for accessing web-services enabled functions of the IOOS.

5.6. The portal shall conform to all Federal guidelines on Internet presence.

5.7. The portal shall be designed to be Section 508 compliant (see http://www.section508.gov/).

5.8. The portal shall provide all necessary policy statements and legal disclaimers.

5.9. The IOOS shall support web browsers Netscape and Microsoft Internet Explorer and others TBD.
5.10. The IOOS shall provide tools for remote content management of the portal structure.
5.11. The portal shall provide links to relevant information such as tools available for generation of the metadata required for this specific system.
5.12. The portal shall provide information on requirements for IOOS data providers.
5.13. The portal shall provide links to the supporting organizations.
5.14. The portal shall be easily modified to a new look and feel.
5.15. The portal shall provide FAQs.
5.16. The portal shall provide on-line documentation.
5.17. The IOOS shall provide a mechanism to solicit and receive user feedback concerning the operation of the system, data quality, portal content, and other issues.
5.17.1. User comments on data sets shall be accessible to IOOS staff for review.
5.17.2. The user feedback mechanism shall provide a "Help" function.
5.17.3. The user feedback mechanism shall provide a mechanism for Usage Tracking.

2. Data Transport (DT) Requirements

1. Overall requirement

1.1. The DT shall support machine-to-machine interoperability with semantic meaning, i.e., the DT shall incorporate some collection of methodologies that promote the scripted exchange of data between computers, with all computers involved in a transaction capable of determining both the syntax and the semantics of the exchanged data without human intervention.
1.2. The DT shall include an access method that is consistent with that which is referred to as "Web services" in the literature.
1.2.1. "A Web service is a software system identified by a Universal Resource Identifier[4], whose public interfaces and bindings are defined and described using XML[5]."
1.3. Other Web services requirements TBD.

[4]Uniform Resource Identifiers (URI): Generic Syntax, IETF RFC 2396, T. Berners-Lee, R. Fielding, L. Masinter, August 1998 (See http://www.ietf.org/rfc/rfc2396.txt)
[5]Web Services Glossary W3C Working Draft 14 November 2002 (See http://www.w3.org/TR/ws-gloss/)

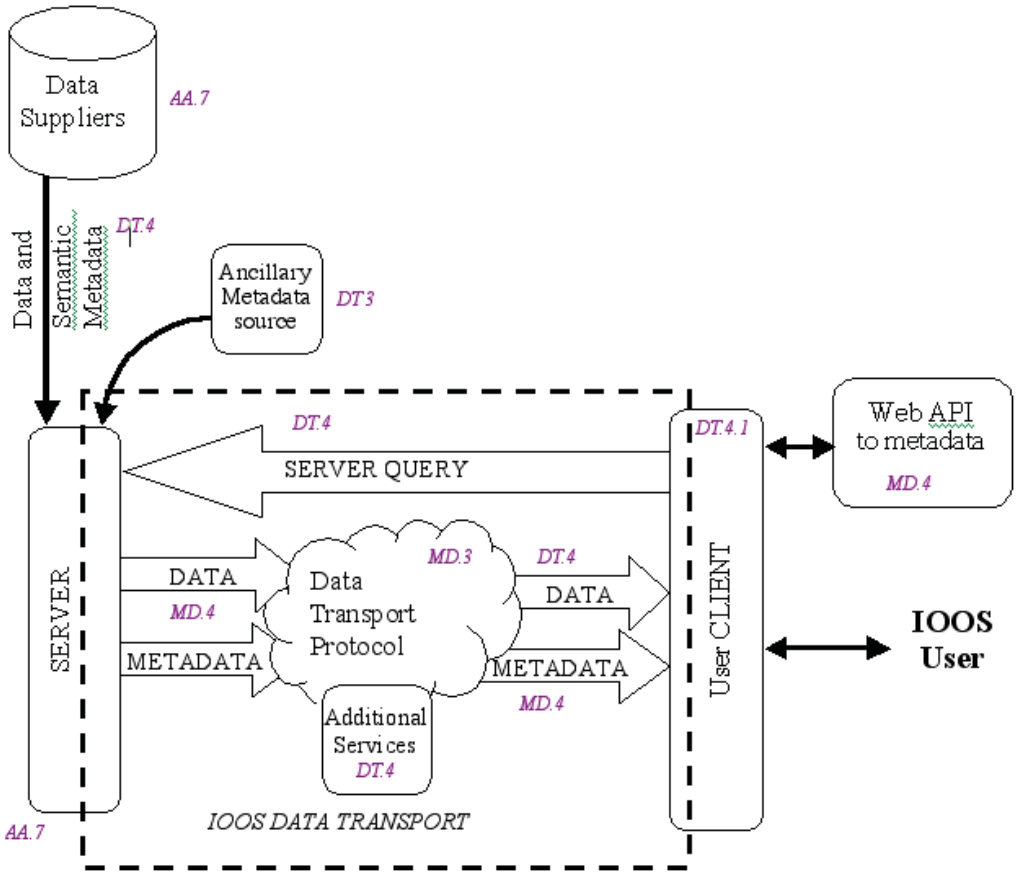


Figure 2. Schematic diagram of DMAC Data Transport (DT)

2. Representational requirements

2.1. The DT shall support metadata as described below:

2.1.1. Syntactic metadata are information about the data types and structures at the computer level, the syntax of the data. For example, variable D represents a floating point array measuring 20 by 40 elements.

2.1.2. Semantic metadata are information about the contents of the data set.

2.2. DT shall be able to transmit all relevant semantic metadata, that is translational use, descriptive use, and search metadata. They must be available in both human-readable and machine-readable forms.

2.3. DT shall be able to express the structure of the numeric data it will encounter in oceanographic data repositories.

2.4. DT shall be able to transmit the numerical data themselves without corruption or loss of precision.

2.5. The following simple and compound types shall be provided:
  2.5.1. Simple types:
    2.5.1.1. Integers (signed 16, 32, 64-bit; unsigned 8, 16, 32, 64-bit); floating point (32, 64-bit);
    2.5.1.2. Strings
    2.5.1.3. Pointers to types
  2.5.2. Compound
    2.5.2.1. Structures
    2.5.2.2. Arrays
    2.5.2.3. others TBD

2.6. The DT shall be capable of accessing data in a variety of formats.

2.7. The DT shall be capable of delivering data of a given data type in a structurally consistent form across all data sets in the system.

2.8. The DT shall provide the metadata needed to transform the data to a consistent semantic form, or it must be capable of delivering the data in a consistent semantic form.

2.9. The DT shall provide access to metadata in a variety of forms, including the standard FGDC forms of the metadata, to take advantage of the metadata developed by different communities. The DT shall be capable of providing access to metadata from a site other than that of the data server.

2.10. The DT shall be capable of providing access to metadata from a site other than that of the data server together with the data. The DT shall be capable of binding these metadata to a data request where appropriate.

3. Modular approach

3.1. The DT shall be constructed in a modular fashion that allows capabilities to be utilized over alternative transport protocols to HTTP.

3.2. **Format** –The DT shall be capable of moving data from a site in which they may be stored in one format to a client application that may require them in another format.
  3.2.1. Transport between sites will be implemented via an intermediate format, referred to as the DT syntactic data model.
    3.2.1.1. The data model shall be discipline-neutral.

3.3. **Structure** - The system shall provide the capability of delivering data to clients in a structurally consistent form where appropriate. In this context, structure means the way that the data are organized, for example, grid, array, etc.
  3.3.1. The structure layer protocol will define the organization of like data objects in a data set.
  3.3.2. Operations and the associated modules in the structure layer that can be performed in a discipline-neutral fashion shall be logically separated from those that require a semantic understanding of the data.

3.4. **Semantic** – The DT shall provide a semantic data model, defined as the semantics implicit in the structural transformations that the system provides and the semantic information transported in the data access protocol.

3.4.1. The core semantic data model shall include the set of translational use metadata.

4. Functional requirements

4.1. The DT shall be capable of providing direct access to data via a variety of client programs, communicating directly with the program without the need to create data files.

4.1.1. The mechanism for user access to IOOS data can be either through a DMAC-enabled browser, or through user-supplied application software implementing DMAC access routines. In either case, DT shall provide the requisite software capability.

4.2. The DT shall support access to real-time data as well as access to retrospective (non-real-time) data.

4.2.1. The DT shall provide a push data delivery service.

4.2.2. The DT shall support "informed pull" of data.

4.3. The DT shall provide for online acquisition of data into legacy applications and new applications packages through the syntactic data model.

4.3.1. The DT shall allow users to obtain data subsets as formatted files (formats TBD) and human-readable ASCII numeric values via a standard Internet browser, possibly implemented via plug-ins.

4.3.2. The DT shall be designed and developed to accommodate the following considerations.

4.3.2.1. Data will be heterogeneous in type and storage format.

4.3.2.2. Data storage will be distributed.

4.3.2.3. Data will often, but not always, reside with the data collector.

4.3.2.4. The system to be developed will be a client-server system.

4.4. The DT shall provide a mechanism for subsetting data sets for retrieval, by parameter, by area, by time window, and by other criteria TBD.

4.4.1. When subsetting data the DT shall provide appropriate metadata.

4.5. The DT shall provide mechanisms for aggregating data.

4.5.1. Data of the same type and from same provider.

4.5.2. Data from different sources that do not or cannot share a single parent metadata record (e.g., observational data from different sources/systems).

4.5.3. When multiple data sets are aggregated, the DT shall provide a mechanism for providing appropriate aggregate metadata.

4.6. The DT shall display, for each data set it contains, the approximate size of the data set selected.

4.7. The DT shall support data restructuring, i.e., any process that ingests a data set described by one data model and maps that data set into another data set described by another data model.

4.8. The DT shall support data manipulation, including (but not limited to) such manipulations as
  4.8.1. Re-projection—for example, Platte-Care to Mercator
  4.8.2. Re-gridding—for example, same projection, different resolution
  4.8.3. Averaging
  4.8.4. Summing
  4.8.5. Scaling of values such that they are delivered in a consistent system of units—for example, # specimens/m^3, m/s, °C
  4.8.6. Conversion of time to different representations
  4.8.7. Conversion of latitude and longitude
  4.8.8. Conversion of depth
  4.8.9. Conversion of missing values
4.9. The DT shall support access-restricted, secure transmission of data.
4.10. The DT shall support fault detection and localization within the DMAC.
4.11. The DT shall support the gathering of performance and usage metrics within the DMAC.

5. Design Constraints

5.1. The DT shall be designed to work cooperatively with other systems. For example, if a repository already uses a system that depends upon a particular data storage format, that site should not be forced to abandon its system in order to adopt IOOS.
5.2. The DT shall be designed to operate with minimum reliance on proprietary software.
  5.2.1. The DT specifications shall be fully and openly accessible to the public.
    5.2.1.1. There is a stated preference for software licensed under the General Purpose License.
5.3. The scheme that the DT adopts for generating syntactic and semantic data models shall be flexible and extensible so that any IOOS server can find a way to express its archive's storage format in an IOOS data model.
5.4. The DT shall support interoperability between Geographic Information Systems and Scientific Information Systems.
5.5. The DT shall support interoperability with other systems developed within other disciplines.

3. Data Archiving and Access (AA) Requirements

1. Vision

1.1. IOOS Data Archiving and Access (AA) will be a distributed system of interconnected archive and data centers that functions collaboratively to receive and preserve the data, and provide easy and efficient access to the data. Search and discovery of data and products will be easy and will directly support the seven IOOS goals.

1.2. Archive collections range greatly in size, complexity, and importance to public and scientific needs. Currently, diverse data service paradigms are used to support access to the archives. IOOS data transport methods, metadata standards, and data discovery interfaces shall be implemented in the Archive System. The result will be a system that provides more uniform access across multiple archive centers and that can handle all collections consistently. The data discovery component will allow access by both humans and machines.

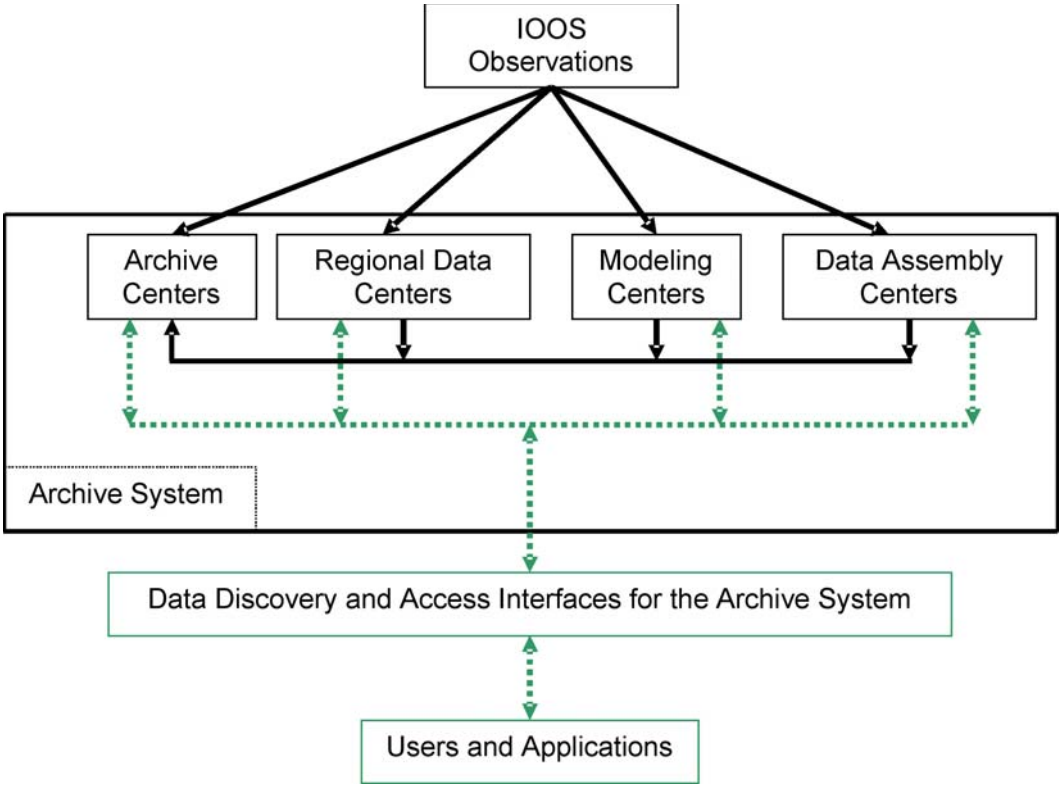


Figure 3. The Archive System represents an alternative view of those DMAC Subsystem elements that are involved with data archival. Primary archival (solid lines) and access (dashed lines) show data flow. Not shown are other data flows that are essential to IOOS but not directly pertinent to the Archive System.

1.3. As the amount of IOOS data steadily increases, the old and new systems of access must remain compatible in order to maintain the high levels of service and allow users to fully discover the archived data.

2. The Archive System

2.1. The Archive System shall use coordinated methods for data collection, quality control, archiving, and user access.

2.2. The system shall consist of a distributed network of archive centers, regional data centers, modeling centers, and data-assembly centers, all interconnected to provide efficient flow of data into the IOOS archive and easy access to its data and products (Figure 3, Data Archiving and Access Requirements).

2.3. Although data may flow from observing systems to any of the four types of centers, at least one copy of each observation desired by IOOS must ultimately reside in an IOOS archive center.

2.4. More than one type of center may be physically collocated, for example, a data assembly center may be an entity at a national archive center.

2.5. Archive centers

- 2.5.1. Archive centers shall acquire, preserve, and provide access to IOOS data in perpetuity.
- 2.5.2. Archive centers shall implement mechanisms to ensure integrity and completeness of the archives.
- 2.5.3. Essential functions include constant monitoring of data streams, accounting for all files and records, and frequent checks of accuracy.
- 2.5.4. Archive centers shall provide for the archival of metadata.
- 2.5.5. Archive centers shall have maintenance strategies that protect the data as storage media and systems change.
- 2.5.6. Data stewards at the archive centers shall maintain constancy in formats and software to prevent conditions that could make accessing the data more difficult, more costly, or impossible.

2.6. Regional data centers

- 2.6.1. Regional data centers shall acquire and provide access to IOOS data collected in a specific geographic region.
- 2.6.2. Regional centers may collect a variety of physical, biological, and chemical ocean data that are used to support scientific, public, and commercial interests in the area.
- 2.6.3. Regional data centers shall apply quality control measures to data and derive specialized products.
- 2.6.4. Regional data centers shall fulfill the long-term archive obligation if they meet the IOOS standards for data integrity and stewardship or if they systematically transfer the data to an archive center.

2.7. Modeling centers

2.7.1. Modeling centers shall procure and synthesize observational data to produce products such as analyses, predictions, or hindcasts that may span a wide range of spatial and temporal scales.

2.7.2. Modeling centers may provide access to their products, but their mission does not include long-term archiving.

2.7.3. Model products that are essential to IOOS goals shall be transferred and preserved at an appropriate archive center.

2.8. Data assembly centers

2.8.1. Data assembly centers shall obtain IOOS data and provide access to it.

2.8.2. Data assembly centers will typically specialize in certain types of data, and often provide quality control and data products in their area of expertise.

2.8.3. Data assembly centers may be permanent (e.g., NDBC) or exist only for limited periods (e.g., WOCE data assembly centers).

2.8.4. Data assembly centers do not provide long-term archiving, but often provide access.

2.8.5. Data assembly centers may gather distributed data and process data over a wide range of disciplines, with the assembled data and products then being submitted to archive centers for long-term storage and access.

3. Data Management

3.1. Although IOOS data may flow into the archive centers over several pathways (Figure 3, Data Archiving and Access Requirements), at least one copy of each set shall reside in a designated archive center.

3.2. Some categories of data will require that multiple copies be stored securely at separate locations under independent data management.

3.3. When data must be duplicated, a primary and secondary data steward shall be designated.

3.4. The primary data steward shall typically be an archive center and shall provide the highest level of access.

3.5. The secondary steward need not maintain full access, but shall maintain the data at the same level of integrity.

4. Access

4.1. Access services for IOOS users shall be provided from most centers in the Archive System.

4.2. Archive centers shall provide some real-time services, and enhance data discovery by using the IOOS metadata standards and data discovery techniques.

4.3. When regional, modeling, and data assembly centers provide access on schedules that meet the IOOS goals, duplication of this effort is not a requirement for the archive centers; however, the archive centers will ultimately receive the data, provide for their long-term preservation, and provide access to full archived data set.

5. Data Receipt

5.1. Modes of data receipt

5.1.1. Real-time-mode data arrive in real-time or near real-time, with the goal of being made available with minimum delay.

5.1.1.1. High-level quality control is not a requirement for real-time-mode data.

5.1.2. Delayed-mode data arrive later than real-time-mode data, and sometimes much later. They may be research collections that have been improved through further processing, or simply raw data collected under circumstances where prompt transmission was not feasible or needed.

5.1.3. The Archive System shall receive and archive sets of either type that address the seven IOOS goals.

5.1.4. All appropriate metadata should arrive with the data.

5.2. Integrity/Consistency

5.2.1. The Archive System shall implement mechanisms to ensure that all valuable data are sent and that an exact copy is received. The IOOS data transport system shall provide sufficient mechanisms to ensure accurate transfers over the networks.

5.2.2. Acceptable tools and procedures include:

5.2.2.1. Receipts and reconciliation reports for transfers over networks.

5.2.2.2. Skilled staff to review metrics (e.g., how much of the expected data were received and how much of the data set was made available).

5.2.2.3. Byte counts, inventories of data files, and checksums of records or files.

5.2.2.4. Test files that can be confirmed against archived data and used to verify local software.

5.2.2.5. Accuracy relative to other data sources (i.e., whether a set of data falls within acceptable ranges or compare acceptably with other data known to be correct).

5.2.3. The Archive System shall provide a failover mechanism for failed data transmissions.

5.2.4. The Archive System shall guard against unrecoverable data loss by making data integrity (or security) a primary objective.

5.2.4.1. Byte counts and checksums shall be calculated and used to verify that the data are uncorrupted when transmitted between data centers.

5.2.4.1.1. These quantities shall again be calculated after every internal process at the archive centers, and then re-calculated periodically on all archived data to protect against such problems as hard disk failures, media degeneration, incomplete file transfers, and malicious hacking.

5.2.5. Virus checks shall be performed on the data before archiving, then periodically on all data kept online.

5.3. The AA shall include guidelines to enable providers developing new data streams to select formats and metadata that can be easily integrated into IOOS. Specifications shall traverse the IOOS data-transport, metadata, and data-discovery components.

5.4. IOOS standards for metadata shall allow different versions of the same data and metadata to be traced by means of information on lineage and version.

5.4.1. The number of old versions of data to be preserved is TBD.

5.5. Data Formats

5.5.1. The AA shall process a broad range of data to be included in IOOS (physical, biological, chemical, geological, fisheries, socio-economic) encompassing many different native data formats.

5.5.1.1. Data providers shall use only established, fully documented formats, (TBD) which the data-transport methods handle. The data transport methods shall be robust and handle many established common formats.

5.5.1.2. It is not a requirement that each center be proficient in every format.

5.5.1.3. All metadata shall meet a common standard defined in Section 1, Metadata/Data Discovery Requirements.

5.5.1.4. If metadata do not meet the common standard, then the provider shall provide a mechanism that accepts the non-standard metadata as input and creates standard metadata as output.

5.5.1.4.1. Archive centers will consider accepting data in all formats.

5.5.1.4.1.1. Unique specialized formats (such as occasionally found in research or field data) are discouraged.

5.5.2. Proprietary formats (with undisclosed internal structure and typically with proprietary software) are unacceptable for long-term archiving and are prohibited.

5.5.3. Each center shall provide and maintain software for accessing each native format.

5.5.3.1. Centers shall maintain configuration management of the software in order to maintain currency with changing data formats.

5.5.3.2. This software shall also provide further documentation of data sets and changes in their lineage.

5.6. File-compression techniques used for transferring IOOS data shall use standard protocols with open documentation, such as GNU zip.

6. Data Preservation

6.1. All four component data centers of the AA will be responsible for acquiring and providing data, but only the archive centers will be primarily responsible for preserving data long term.

6.1.1. Long term is defined as much longer than the typical funding period of an oceanographic research project or the career of a principal investigator.

6.1.2. To qualify as an archive center, a data center shall be able to perform the following functions related to data preservation:

6.1.2.1. Create and manage multiple copies of the data and metadata;

6.1.2.2. Verify and generate metadata as well as preserve them with their associated data;

6.1.2.3. Frequently check data integrity;

6.1.2.4. Plan for evolution of technology.

6.2. Archive centers shall be able to create and manage one or more copies of all IOOS data and metadata, both online and offline, according to the specified IOOS data category and according to NARA and other Federal guidelines. Table 1 summarizes the four data categories and the number of archival copies required to meet the minimum IOOS Archive System standards.

6.3. Data Categories

6.3.1. Irreplaceable Data

6.3.1.1. The AA shall maintain two copies in separate archive centers in perpetuity.

6.3.1.2. The two copies of irreplaceable data shall be preserved in separate facilities under independent data management.

6.3.1.3. One facility will be designated as the primary archive center for a particular data set, and the other as the secondary archive center.

Table 1. IOOS Data Classes for Archiving and Access

Data Category	Data Description	Examples	Minimum Number of Archival Copies
Irreplaceable	Observational and research quality data that can not be reproduced or easily regenerated	Raw, ancillary satellite observations; Instrumental measurements; Biological samples; Model reanalyses; Complex merged data analyses	Two
Replaceable	Derived from irreplaceable data, can be regenerated through systematic processing	Calibrated satellite radiances; Simple composites or analyzed data	One
Perishable	Real or near real-time data; typically replaced by higher quality data.	Direct broadcast satellite data; Operational analyses; Quick-look analyses based on uncalibrated or incomplete data	One
Virtual	Data provided through on-demand processing	Subsets from GUI; Analyses from a Live Access Server	Two*

* Original generation algorithms and documentation only.

6.3.1.4. The primary and secondary archive centers storing irreplaceable data may operate as mirror sites, both offering the same level of access, or one as the exclusive access center and the other as a "deep" back-up center (e.g., a regional data center could serve as a secondary archive center).

6.3.2. Replaceable Data

6.3.2.1. The AA shall maintain one copy (residence time in the archive will vary with replacement cycle).

6.3.3. Perishable Data

6.3.3.1. The AA shall maintain one copy until higher quality data are available.

6.3.3.2. When decision-critical data products are derived from data in this class, and it is necessary to reproduce the data product, the perishable data may inherit an extended term for data preservation that is not obvious for the original data alone.

6.3.4. Virtual Data

6.3.4.1. No copies of the data are necessary, but an archive center and the virtual data provider should maintain separate copies of generation software and documentation.

6.4. Metadata. The AA shall maintain metadata as defined in Section 1, Metadata/Data Discovery Requirements, and will include the following types of metadata:

6.4.1. Use metadata (the semantic and syntactic information about a data set);

6.4.2. Discovery metadata (standard structured information describing a data set).

6.4.2.1. Data set lineage history (e.g., which irreplaceable data set was used to create this current data set);

6.4.2.2. Data category specification, which determines the storage requirements;

6.4.2.3. Release date, which is the date to remove temporary restricted access;

6.4.2.4. Version number and description of the version number;

6.4.2.5. Description of the file naming convention;

6.4.2.6. Unique IOOS-wide data set name or identification;

6.4.2.7. Mechanisms for correct publication citation and reference tracking.

6.4.3. Documentation metadata (bibliographic information about documentation associated with a data set).

6.4.4. Metadata shall be dynamic to accommodate through numerous incremental updates, modifications, corrections, and occasionally, full replacements.

6.4.5. Metadata shall be inclusive of sufficient information to provide an end-to-end lineage record, starting with the measurements or computation through the change and modification history and eventually to established scientific or public knowledge.

7. Data Provision and Access

7.1. The AA shall accommodate data access from any suitable component of the IOOS Archive System (Figure 3, Data Archiving and Access Requirements).

7.1.1. By querying the system with the DMAC data-discovery interface, users or applications can discover what data are available. The data may then be pulled automatically with the data transport methods, or by the user from a GUI that displays the various options.

7.2. The AA shall implement the protocol for transporting data defined in Section 1, Data Transport Requirements.

7.3. The AA shall provide for access services tailored to data sets as provided in Table 2.

7.4. The core protocols shall include FTP, HTTP, and the IOOS DT protocols. Most IOOS data sets will be available in at least one, and preferably two or three, of these protocols. As the IOOS standard transport protocol, OPeNDAP should be used whenever possible. The characteristics for each of these core services are:

7.4.1. FTP – Direct downloads of data files, unrestricted public access, and no application support.

7.4.2. HTTP – Direct downloads of data files, restricted or unrestricted access, and no application support.

7.4.3. OPeNDAP – Application-layer protocol that supports a number of data storage formats and allows a number of client applications to access data transparently.

7.5. The AA shall use the IOOS DT protocols to offer the following extended services:

7.5.1. Spatial subsetting – Extracting spatial sub regions from data sets for larger geographic areas.

Table 2.

Center	Data set	Core Services			Extended IOOS Services						
		FTP	HTTP	OPeNDAP	Spatial Subset	Parameter Subset	Temporal Subset	Temporal Aggregation	OpenGIS Map	Online Analysis	Online Ordering
Center 1	Data set 1		X	X	X						
	Data set 2		X	X	X	X		X	X	LAS	X
	Data set 3		X	X							X
Center 2	Data set 3	X	X	X						GrADS	
	Data set 4		X	X			X				

Conceptual matrix of data access services for different data sets at different component of the IOOS Archive system. Note that data set 3 is offered at two centers, but with different services.

7.5.2. Parameter subsetting – Extracting one or more variables from data sets containing many variables.

7.5.3. Temporal subsetting – Extracting short periods from data sets covering longer periods.

7.5.4. Temporal aggregation – Creating a longer time series from data files for shorter periods.

7.5.5. GIS products – Depicting data projected, interpolated, and rendered onto a map with GIS protocols.

7.5.6. On-line analysis – Analyzing online by using tools on the data server such as the Grid Analysis and Display System (GrADS) or the Live Access Server (LAS). The resulting data or graphics can then be downloaded.

7.6. Data sets that are stored offline shall be kept accessible and discoverable through the data-discovery interfaces.

7.6.1. This access to offline data may be initiated by online ordering. Online ordering is a mechanism by which data are ordered and then picked up or delivered later.

7.7. The AA shall accommodate maximum latency periods as defined in the metadata.

7.7.1. For IOOS access latency is defined as the time between the earliest primary observation (not counting ancillary data) in a data file and the availability of that file to users.

7.7.2. Latency requirements shall be assessed and suitably defined in the metadata.

7.8. The AA shall provide unrestricted access under normal circumstances

7.8.1. The AA shall restrict access under special circumstances including:

7.8.1.1. Proprietary embargo – Data are available only for sale from commercial companies.

7.8.1.2. National security – Data are available only for defense purposes.

7.8.1.3. Calibration and validation – Data are available only to the science team while they calibrate or validate instruments, data, or models.

7.8.1.4. Non-commercial use only – Data are available for government applications and academic research, but not for resale.

7.9. The AA shall provide user services and use metrics.

7.9.1. Online documentation and knowledgeable staff shall be available to provide assistance and advice on both access and content.

7.9.2. Additional background information will be available through references and citations in the metadata.

7.9.3. The AA shall provide a facility to collect broad use metrics to evaluate the system effectiveness and gain a sense of how to improve it. Metrics shall include the following as a minimum:

7.9.3.1. Number of "users" – The anonymous nature of much of the access prevents the true number of users from being collected. Unique Internet addresses are the closest proxy to this number that can be collected, and are useful for evaluating trends as well as access by well-constrained domains such as .gov, .mil, .edu and international domains.

7.9.3.2. Number of accesses – This is the number of files downloaded or otherwise accessed through the various services. Note that volume of data is not used here; a cornerstone of DMAC data access is to provide subsets, GIS maps, online analyses - in short, only the information required by the user. The data access metric shall also be broken down by data set and service method.

7.9.3.3. System performance statistics – This includes use of disks and computers as well as work performed (i.e., services executed and volume accessed).

7.9.3.4. In addition to numeric metrics, the AA shall provide for measurements of qualitative access.

7.9.3.4.1. Specifically, all archive systems shall have a means of soliciting and capturing user feedback on services and data sets.

Section 2. Phased Implementation Plan

METADATA/DATA DISCOVERY ACTIVITIES AND SCHEDULE

(see Figure 4)

1. Activity: Metadata: Determine IOOS Metadata Content and Format Standards

- **Description**: Determine the metadata contents and format for all IOOS metadata. Metadata will be FGDC CSDGM compliant but may require additional elements not in that standard.
- **DMAC Component:** Metadata
- **Milestone 1:** Compile IOOS metadata standards
- **Estimated Resources:**
- **Schedule:** Start beginning of Year 1 and continue for 3 years
- **Sequencing:**
- **Partnerships:** DMAC expert teams with additional input as needed.
  - **Task 1:** Convene an IOOS Metadata Standards Working Group with metadata representatives from all IOOS data disciplines to do a comprehensive assessment of metadata standards.
    - The make up of the W/G must:
      - represent the interests of all major ocean metadata holders;
      - be able to represent the broadest range of ocean data types: *in situ*, satellite-derived, biological data, sonar, various model output types, etc.;
      - liaisons to US environmental metadata standards activities should be identified (e.g., FGDC);
      - liaisons to international ocean metadata standards should be identified.
    - The W/G should consist of core standard group and a number of specializations to address special data types. The core group will have full responsibility for "format" issues as well as for content that is in common to all data types. The core group could be given responsibility to appoint specialist groups as it sees fit.
    - Also form expert subcommittees to address discipline specific issues. Identify need for extended elements to standard format. Evaluate developing a Standard Profile under the FGDC Content Standard.
    - After the initial work to produce the Interim Standards, some form of the committee will become a standing standards committee.
    - Consider whether to support standards other than FGDC CSDGM, e.g. Dublin Core, MARC21.
    - Category: Committee work

- **Task 2:** Develop Preliminary IOOS Metadata Standard
  - Category: Committee work
  - Deliverables: Initial IOOS Metadata Standards
- **Task 3:** Establish Liaison with metadata standards groups like FGDC.
  - Category: Committee work
- **Task 4:** Make interim list of keywords and a data dictionary
  - Category: Committee work
- **Task 5:** Study use of thesauri to enable machine-to-machine interoperability with semantic meaning.
  - Note: If thesauri are used, they will have to be maintained through a fair sized effort considering the variety of data and the complexity of language.
  - Category: Committee work
- **Task 6:** W/G incorporates input from expert subcommittees, results of studies, R&D, pilots, etc. into subsequent standards updates up to the release of the Interim Standards. Circulates draft interim standards for community comment. Specification will include guidelines on Data Quality issues along with lineage issues.
  - Category: Committee work
- **Task 7:** Develop policy for granularity of metadata.
  - Category: Committee work
  - Deliverables: IOOS Metadata Standards, keywords and data dictionary plus any updates or interim releases as needed
- **Task 8:** Develop guidance on metadata for subsetting and aggregation.
  - Category: Committee work
  - Description: This assumes that related metadata would be delivered along with transported data. How should the metadata change when the transported data are not identical to the source data? This must have strong representation from the Data Transport team and a joint sub-group should be considered. This is an activity overseen by the core standards group.
  - Task 8-1: Determine metadata modification for subsetted data.
    - Includes temporal, spatial, and parameter subsets
  - Task 8-2: Determine metadata modification for aggregated data.
    - Case 1: Data are from the same provider and same data type.
    - Case 2: Data are from different sources, may or may not be same data type.
  - Task 8-3: Determine metadata modification requirements for products or merged data.
    - Assumption is data products will have own unique metadata.
    - Key question: Should source data metadata be also delivered with product?
  - Deliverables: Documented guidance on metadata for subsetting and aggregation
  - Partnerships: Data Transport, Data Products

2. Activity: Develop Tools and Procedures to Support Metadata Providers

- **Description:** Develop or acquire procedures, practices and tools to aid developers and IOOS in designing, producing, verifying, and maintaining metadata.
- **Milestone 1:** Develop and maintain metadata.
  - **Task 1:** Select or develop a master metadata management system.
  - **Task 2:** Develop/acquire tools for metadata generation, validation, maintenance.
  - **Task 3:** Provide developer support to users in addition to tools and User Guide Training, support networks, consulting and help desk.
  - **Task 4:** Plan regular reviews of exiting metadata by data providers. Update or add information as data set circumstances change. Update perishable information such as contact info. This is assumed to fit within the standard metadata framework, otherwise make change recommendations to Standards Committee.
  - **Task 5:** Develop User Guide for Metadata.
- **Deliverables:** User Guide for Metadata
- **Estimated Resources:**
- **Schedule:** Year 2
- **Sequencing:**
- **Partnerships:** User Support

- **Milestone 2:** Develop tools to modify metadata as appropriate to data accessed.
- **Deliverables:** Tools to modify metadata appropriately
- **Estimated Resources:**
- **Schedule:** Years 3-4
- **Sequencing:** Following release of guidance on how to modify metadata for each of the conditions
- **Partnerships:** User Support

3. Activity: Discovery: Select or Develop and Maintain Catalog and Search Capability

- **Description:** The catalog(s) are the information source used for data discovery. Search capability is the prime purpose of the catalog.
- **Milestone 0:** (for DMAC Steering Committee): Designate an initial list of pre-operational Metadata Cataloging Centers. Initial list presumably to include NASA/GCMD and NOAA/NCDDC. Additionally, this committee should look at the OBIS system for inclusion of biological information.
- **Milestone 1:** Convene a Catalog Architecture Working Group.

- Membership should be of people expert in metadata searching and distributed metadata management. It must have broad representation from the data supplier community as well as DMAC teams.
- **Task 1:** Design Catalog Architecture.
- **Sequencing:** Can be concurrent with Activity 1.

• **Milestone 2:** Determine search/browse capabilities needed.
- **Task 1:** Determine the level of search/browse needed and hence the composition of the catalog.
  - Minimum metadata vs. all metadata or in-between
  - If full text search required, then full metadata record will be required.
  - Search candidates or features:
    - Spatial Search
    - Temporal Search
    - Thematic Search – text searching esp. important
    - Taxonomic Search – biological data
    - Parameter Search
    - Additional Search Parameters
    - Browse Option
    - Results listing and Search refinement
  - Category: Committee work
- **Task 2:** Get user feedback on search/browse needs.
- **Task 3:** Choose and schedule search capabilities to be implemented.

• **Milestone 3:** Determine metadata loading and update procedures.
- **Task 1:** Write Catalog Management Plan.
  - How is catalog managed and maintained?
  - How are catalogs kept up-to-date as data source metadata changes?
    - Option 1: Catalog harvests metadata from data source.
    - Option 2: Catalog metadata maintained by data source.
    - Option 3: Allow both (committee recommendation)
- **Deliverables:** Design recommendations

• **Milestone 4:** Plan catalog security (cross-discipline with all DMAC)
- Plan security tools, procedures, and practices to protect all participating systems from inappropriate access, intentional, or accidental.
- **Task 1:** Determine catalog security needs with DMAC.
  - Category: Committee work
- **Task 2:** Determine catalog access needs.
  - Category: Committee work
- **Task 3:** Write Catalog Security Plan.

    - Category: Committee work
  - **Deliverables:** Catalog Security Plan
  - **Partnerships:** Archive team, Data Transport team
- **Milestone 5:** Build initial capability (pilots are needed, TBD).
  - Multiple pilots and pre-operational tasks are acceptable if they can co-exist with the catalog structure including the OBIS development efforts for distributed biological data search.
  - **Task 1:** Build catalog.
  - **Task 2:** Populate Catalog.
    - Populate DMAC catalog based on current archive centers.
    - Accept or Harvest Metadata from archive centers.
  - **Task 3:** Build initial user interface (may be web portal).
  - **Schedule:** Beginning Year 2 and continuing
  - **Sequencing:** Will need to be repeated as archive centers join.
  - **Partnerships:** All archive centers currently part of IOOS.
- **Milestone 6:** Transition to pre-operational and operational systems.
  - **Task 1:** Transition to Pre-Operational.
  - **Task 2:** Transition to Operational.

4. Activity: Discovery: Develop Discovery Interface for Archive System

5. Activity: Discovery: Design Discovery Portal

- **Description:** Decide if a data portal is desirable and recommend functionality especially in search capabilities. Initiate pilot task(s) followed by a pre-operational task based on lessons learned and user feedback.
- **Milestone 1:**
  - **Task 1:** Design overall architecture.
    - Single vs. many, governance?
    - Category: Committee work
  - **Task 2:** Search content and scope
    - Subscriptions
    - Event association with parameters
    - Broad vs. narrow search
    - Note: These depend on and feed back to metadata standards and design. Functionality will be added in stages, probably from simple to more complex.
    - Category: Committee work
  - **Task 3:** Solicit and incorporate user feedback.
    - Category: IOOS Standards Process
  - **Task 4:** Determine access given to public search engines, for example, Google.

    - Category: Committee recommendations => DMAC policy (governance)
  - **Task 5:** Pilot Data Portal - NOAA NCDDC
    - Provide for a catalog and data access portal at NOAA's National Coastal Data Development Center and include discovery of NDBC data (hub of 70 moored buoys and 60 C-MAN shore sites – transporting hourly observations).
    - Category: Contract(s)
  - **Task 6:** Pre-operational Data Portal
    - Category: Contract(s)
  - **Task 7:** Operational Data Portal
    - Category: Contract(s)
- **Deliverables:** Design recommendations
- **Estimated Resources:**
- **Schedule:** Start right away using existing data portals for study.
- **Sequencing:**
- **Partnerships:** User Support, user community, Data Transport, portal experts

6. Activity: Discovery: Study Alternate Discovery Approaches

- **Description:** Study alternate discovery approaches like Web Services and Semantic Web to address feasibility of this type of approach. This study could reveal methodologies to translate among multiple ontologies and allow the user to search among multiple controlled keywords and thesauri.
- **Milestone 1:** Study an implementation language neutral approach.
  - **Task 1:** Study, report alternatives like web service.
- **Milestone 2:** Study feasibility of using semantic web.
  - Description: Determine feasibility of using semantic web and ontologies in connection with IOOS metadata and for interoperability with other data catalogs.
- **Milestone 3:** Convene a working group to study semantic web feasibility
  - **Task 1:** How is this broken up if at all?
    - If initial decision is "No," shouldn't it be revisited in a few years as the technology matures?
    - If decision is "Yes," initiate a R&D or pre-operational task?
    - Category: Committee work
- **Deliverables:** Feasibility Report
- **Estimated Resources:**
- **Schedule:** Year 1
- **Sequencing:**
- **Partnerships:** Interim Standards Committee, User Support team, semantic web experts.
- **Milestone 2:** R&D or Pilot study of alternate approaches in DMAC
  - **Task 1:** R&D or Pilot study of alternate approaches

  - **Category:** Committee work
- **Deliverables:** Functioning demonstration for review.
- **Estimated Resources:**
- **Schedule:** Year 2
- **Sequencing:**
- **Partnerships:** Interim Standards Committee, User Support team, semantic web experts.

7. Activity: Design and Implement Data Location Service

- **Description:** The end result of the Data Discovery process should segue seamlessly into the Data Transport (access) process—either the: (1) DMAC middleware connection, (2) on-line browse (visualization and subsetting), or (3) web file transfer (e.g., FTP). On-line, distributed data are dynamic in the sense that the point of access for data may move, and fine granularity information about the data sets may not be available in the catalog (e.g., the catalog cannot have the ability to perform GIS-style proximity queries about available data). The Data Location Service will be a standard machine-to-machine interface which enables the seamless segue from Data Discovery to Data Access.
- **Milestone 1:** Determine specifications for data location service.
  - **Task 1:** Determine specification for finding path to the requested data (e.g., the directory on the file server in the data archive).
    - Category: Contract(s) or RFP
  - **Task 2:** Determine specification for finding individual granules of requested data (e.g. the individual files which satisfy user request)
    - Category: Contract(s) or RFP
- **Milestone 2:** Data Location Service pilot projects
  - **Category:** Pilot projects with collaborating data suppliers and catalog services
- **Milestone 3:** Broad Deployment of Data Location capability
  - **Cross cut:** jointly with "population" of DMAC middleware solution
- **Deliverables:**
- **Estimated Resources:**
- **Schedule:** Year 2 or 3
- **Sequencing:**
- **Partnerships:** User Services

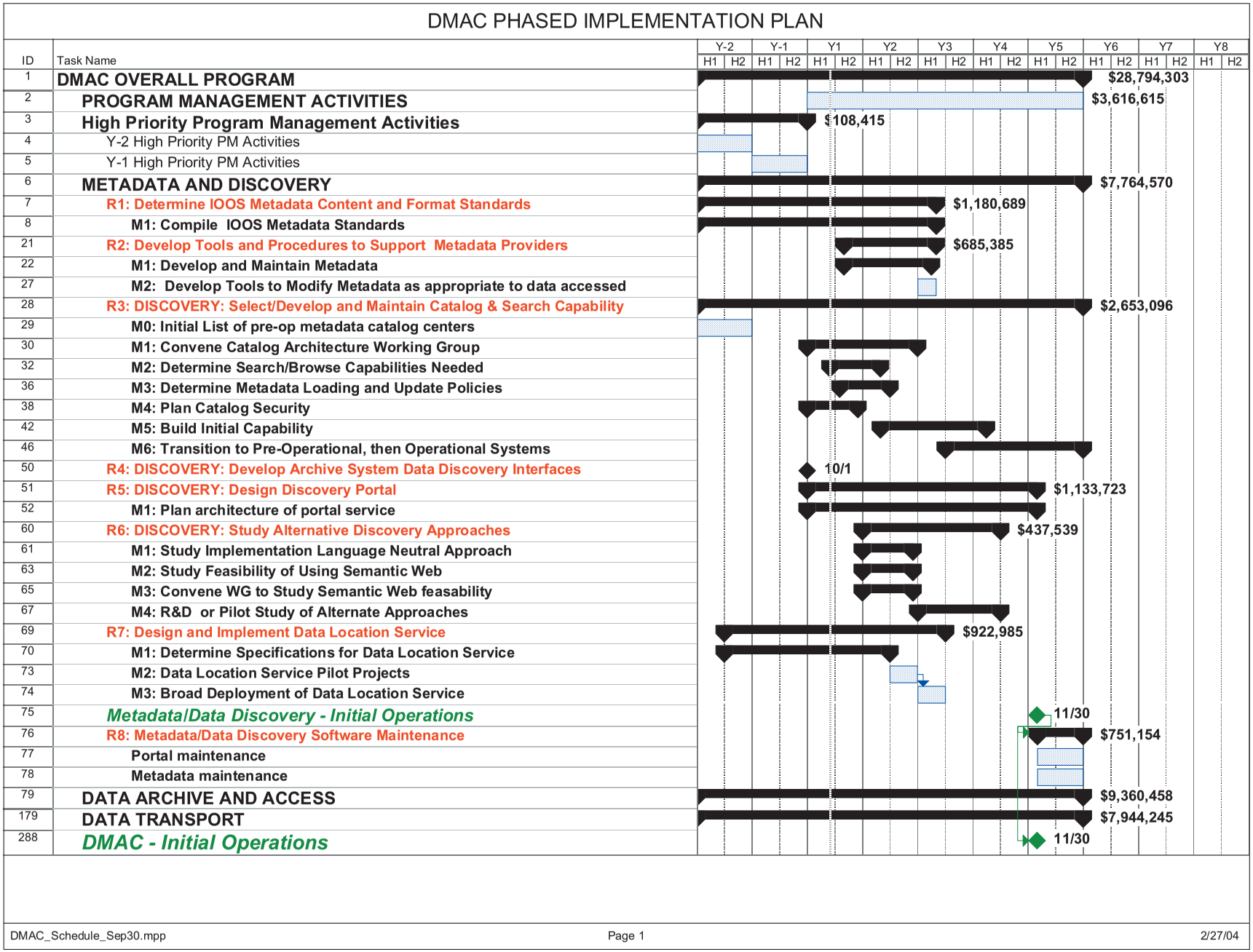


Figure 4. Metadata and Data Discovery Gantt Chart

Data Transport Activities and Schedule

(see Figure 5)

1. Activity: Develop Comprehensive IOOS Data Model

- **Description:** Accessed data will be moved into the data model for transport between server and client.
- **Comment:** There is no strict requirement for a single comprehensive model, but there should generally be only a single, correct representation for any given data class. If multiple representations of the same data class do exist, a reliable procedure to translate between them must be documented
- **Milestone 1:** Develop Comprehensive Data Model - Complete from the perspective of both syntactic and semantic elements – To provide the maximum flexibility in system design, the data model should be divided into a syntactic portion that is discipline neutral and a semantic portion that contains the discipline specific characteristics of the data.
  - **Task 1:** Develop a syntactic data model[6] for the system.
    - Approach: Adopt the OPeNDAP data model as the initial, provisional fast-track solution[7].
    - Special Considerations: The primary focus of this task should be the augmentation (if needed) of the OPeNDAP data model to accommodate data types within the OBIS and GIS data models.
    - Level of Effort: FTE[8] (TBD)
    - Duration: 6 months. The task outlined here is only for the initial development of the semantic data model. A subsequent task addresses the evolution of the model. This is true of Task 2 also.
    - Start: Immediately[9].
    - Category: Committee, OPeNDAP, OBIS, or the Data Model Working Group.
  - **Task 2:** Develop a semantic data model for the system.
    - Approach: Consider existing semantic data models such as that being promoted by OGIS, the DEI data model and others that may already exist. It is imperative that the semantic data model be kept as simple as possible to ensure the maximum compliance within the ocean community.

[6]The resulting data model may in fact consist of several data models. If so, they will collectively be referred to as the data model in this work plan.

[7]The OPeNDAP data model has been developed explicitly for this purpose, has already been vetted within the ocean community as part of the NOPP funded NVODS effort and is now in operational use by NVODS (as well as other non-oceanographic communities).

[8]The levels of effort identified in this work plan are supported FTEs (full time equivalent years). This support could be provided by subcontract to one of the groups indicated or provide funding to the Data Model Working Group for this purpose. Addition community participation is anticipated on many of these efforts through committee work. The level of effort for such committee work is not included here.

[9]Immediately means as soon as practical. These are absolutely essential components of the system, components on which much of the rest of the data management system rests.

- Level of Effort: FTE (TBD)
- Duration: A candidate semantic data model should be developed within 6 months of project initiation. There will likely be additions/modifications to the semantic data model over the next two years as it becomes more heavily exercised and as a result of some of the pilot efforts discussed below.
- Start: Immediately.
- Category: Subcontract, Metadata Standards[10], OBIS, OGIS, OPeNDAP. The group that puts this model and the various components together must include the data representation issues community very broadly. It should include representatives from each of the oceanographic sub-disciplines (biology, physics, chemistry, and geology), from GIS community (OGIS, ESRI, EaSY, etc.), from the ocean modeling community (GCM, coastal and finite element), and from the data collection communities—satellite (projections, swath), and *in situ* (hydrographic, moorings, floats).

- **Task 3:** Develop a controlled vocabulary for system contents.
  - Approach: Adopt as a starting point the controlled vocabulary developed by the Marine XML consortium.
  - Level of Effort: FTE (TBD)
  - Duration: 6 months.
  - Start: Immediately.
  - Category: Subcontract, Metadata Standards, OBIS, OPeNDAP, and national and international metadata standards. The group that assembles the controlled vocabulary must include broad representation from the community of ocean data users: biologists, chemists, geologists, physicists, community planners, etc.
- **Task 4:** Synthesize the work of Tasks 1 and 2 into a complete data model.
  - Special Considerations:
    - This task involves the assembly of the syntactic and semantic data models into a complete data model; i.e., linkages between the data model components in the two must be established. This work may require the addition of data types to the syntactic data model.
    - Pilot implementations designed to exercise the data model are discussed under Milestone 2.
  - Level of Effort: FTE (TBD)
  - Duration: 9 months.
  - Start: 6 months.
  - Category: Subcontract, OPeNDAP, OBIS.

[10]Metadata Standards refers to the IOOS Metadata Standards group.

- **Task 5:** Publish draft data model—follow IOOS Standards Process for Review.
  - Subtasks:
    - Devise a plan to obtain community feedback.
    - Publish and circulate the draft data model.
    - Obtain feedback from the community on the data model.
  - Level of Effort: FTE (TBD)
  - Start: 1 _ year
  - Duration: 6 months
  - Category: Data Model Working Group
- **Task 6:** Pilot implementations of data model
  - Brief pilot task using data model for non-gridded data, such as remote sensing "swath" data.
    - Restructuring and aggregation of this sort of "sequence" data are important due to the large amount of data in that format.
  - Pilots should include network transport utilizing the "fast-track" transport mechanism at a minimum
  - Category: Contract
- **Task 7:** Broad testing of data model by distinct ocean data communities
  - including data observed from biological/laboratory sampling, cruises, moorings, floats, satellites, … and produced by the broadest range of models
  - Category: Community participation activity
  - Sequencing: Must follow pilot testing

- **Deliverables:** Comprehensive IOOS Data Model Standard
- **Estimated Resources:**
- **Schedule:** Year 1 and 2
- **Sequencing:** IOOS Standards process must be completed before Task 6
- **Partnerships:** oceanographic data communities, OPeNDAP, national and international metadata standards

2. Activity: Deliver time-critical (real-time) data to data assembly and operational modeling sites

- **Description:** IOOS sites that have regular, repeated need of time critical observations may best be served by a subscription-based "data push" service.
- **Milestone 1:** Provide operational support for time-critical data.
  - **Fast-track note:** An effective implementation of DMAC real-time delivery is singularly important in commencing the integration of IOOS operational observations with modeling activities. If a suitable candidate is available it will significantly advance the IOOS toward implementing a fast-track solution to this component. The Plan must include procedures to evaluate the effectiveness of the solution adopted and either make adjustments to it or abandon it, if it proves unsuitable.

- **Task 0:** Characterize the need for real-time data.
  - Category: The mode in which real-time data should be received has not been clearly articulated. A careful examination of the issues related to this must be undertaken. Such an examination would include the number of sites that will want access to real time data, the number of originating data sites, the type of data that is required, etc.
- **Task 1:** Adopt Unidata IDD as initial, provisional, fast-track transport solution.
  - Category: Unidata IDD is IOOS-pre-operational for operational, real-time delivery of formatted files to IOOS modeling sites. It is IOOS-pilot for format conversions of the data.
- **Task 2:** DMAC evaluation and review of Unidata IDD, including data carry capacity, data integrity, and data delivery assurance.
  - Category: DMAC Governance Committee
- **Task 3:** Evaluate the state of real-time data access to current and potential modeling operational sites.
  - Work cooperatively to minimize unproductive duplication and maximize timely, reliable availability of quality-controlled, real-time observations for modeling.
  - Category: Committee work
- **Task 4:** Identify IOOS partner sites to serve as real-time data assembly and distribution centers
  - Adopt the US GODAE server (collocated with USN FNMOC) as a provisional, pre-operational IOOS real-time data assembly and distribution point (identification of other sites to follow).
  - Category: IOOS Governance
- **Task 5:** As needed, initiate complementary and/or alternative real-time delivery solutions as R&D activities, Pilots, or Pre-operational solutions.
  - Category: DMAC Governance Committee
- **Task 6:** Explore blended Push/Pull delivery in which data are pushed only to data assembly centers. All others use Pull delivery (typically middleware, FTP, or HTTP transfers).
  - Category: Contract

- **Deliverables:**
- **Estimated Resources:**
- **Schedule:**
- **Sequencing:**
- **Partnerships:** Modeling centers, data assembly centers

3. Activity: Develop DMAC Middleware

- **Description:** The middleware solution embodies four components: (1) the Ocean Data Access Protocol (ODAP) – the format-neutral method of requesting and receiving data and metadata over an Internet connection, (2) translating data from legacy data management systems (formatted files, RDBMS, etc.) into the ODAP, and (3) ingesting data from the ODAP into

legacy and new client applications. The system will be capable of restricting data delivery based on data volume to be delivered. These must be changeable at the discretion of the DT team or DMAC Governance.

- **Milestone 1:** Determine the breadth of data management solutions in use by IOOS data suppliers, which must be supported by the middleware.
  - **Task 1:** Survey the IOOS participants to determine the current data management solutions in usage, the file formats and data management systems that must be addressed by the middleware, and the particular data sets which depend upon each management approach. Evaluate the subsetting needs that attend each data management system.
    - Category: contract
  - **Task 2:** Prioritize the server-side requirements based on IOOS theme priorities and critical data streams.
    - Category: DMAC Governance Committee
  - **Task 3:** Initiate development of server-side solutions based upon priorities
    - Sequencing: must follow adoption of initial transport protocol
    - Category: contract
- **Milestone 2:** Determine the breadth of legacy and new client applications that should be supported. Similarly survey and prioritize requirements for delivery of formatted subsets to users. Priorities should reflect the seven IOOS themes
  - **Task 1:** Survey user groups (and potential ocean information product suppliers) to access application and formatted file needs.
    - Category: IOOS User Outreach Committee activity
  - **Task 2:** Prioritize the client-side requirements based on IOOS theme priorities and critical data streams.
    - Category: DMAC Governance Committee
  - **Task 3:** Initiate development of application solutions and downloadable formats based upon priorities.
    - Sequencing: must follow adoption of initial transport protocol
    - Category: contract
- **Milestone 3:** Determine the specification for the ODAP.
  - **Fast-track note:** An effective implementation of the DMAC middleware component is singularly important to the ability to begin the integration of IOOS, as it allows data suppliers and users to bypass traditional barriers of file format, size, and locality. If a suitable candidate is available, it will significantly advance the IOOS to implement a fast-track solution to this component. The Plan must include procedures to evaluate the effectiveness of the solution adopted and either make adjustments to it or abandon it, if it proves unsuitable.
  - **Task 1:** Adopt OPeNDAP as initial, provisional fast-track transport solution.

    - Category: OPeNDAP is IOOS-Operational for gridded data; IOOS-Pilot for all other classes of marine data
  - **Task 2:** Publish draft OPeNDAP specification document. Request comments.
    - Category: contract
  - **Task 3:** DMAC evaluation and review for OPeNDAP
    - Category: DMAC Governance Committee
  - **Task 4:** As needed initiate complimentary and/or alternative ODAP solutions as R&D activities, Pilots, or Pre-operational solutions
    - Category: DMAC Governance Committee
  - **Task 5:** Develop an ancillary information framework allowing OPeNDAP servers to convert the native structure and attributes of a data set into the standard DMAC data model
    - Category: Contract
  - **Task 6:** Develop detailed requirements and software specifications, followed by design and implementation of the aggregation servers required for *in situ* data collections.
    - Category: Contract
  - **Task 7:** Identify (or adopt) procedures for developing consistent semantic use metadata for all IOOS data sets.
    - Category: Contract
- **Milestone 4:** Implement Server-side Middleware Tools
  - **Task 1:** Survey IOOS participants to determine current data management solutions
  - **Task 2:** Prioritize server-side requirements
  - **Task 3:** Initiate development of server-side solutions based upon priorities
- **Milestone 5:** Adapt or develop Client Software for Initial Transport Protocol
  - **Task 1:** Survey and recommend priority for applications to be supported by DMAC
  - **Task 2:** Initiate development of application and format solutions
- **Deliverables:** Functioning IOOS middleware component
- **Estimated Resources:**
- **Schedule:**
- **Sequencing:**
- **Partnerships:** Archive and Access, Metadata and Discovery, OPeNDAP, OpenGIS, OBIS, etc.

4. Activity: Make data available using IOOS middleware solution

- **Description:** Work with suppliers of data to make data available through the DMAC middleware solution.
- **Milestone 1:** Ensure that IOOS data suppliers make data available through middleware.
  - **Task 1:** Train middleware installers/troubleshooters/trainers.
    - Category: Contract

  - **Task 2:** Install middleware adaptors and other software as needed at suppliers' sites and train local personnel in configuration and management of the software.
    - Category: Contract
- **Deliverables:**
- **Estimated Resources:**
- **Schedule:** Start Year 1. Continuing.
- **Sequencing:**
- **Partnerships:** IOOS community, DMAC teams, governance, representatives from data providers in the disciplines.

5. Activity: Data Manipulation Services

- **Description:** Add optional functionality which may be so commonly required that great efficiencies and additional levels of integration are achieved through adding them as core DMAC services.
- **Milestone 1:** Prioritize and implement Data Manipulation Services.
  - **Task 1:** Prioritize Data Manipulation Services, including aggregation, regridding, and simple transforms such as averages and extrema.
    - Category: Governance Committee
  - **Task 2:** Develop specifications and implement services
    - Category: Contract
- **Deliverables:** Services descriptions and schedule for implementation
- **Estimated Resources:**
- **Schedule:** Year 2 and 3
- **Sequencing:** After initial DT work is done
- **Partnerships:**

6. Activity: Develop Metrics and Implement Performance Monitoring

- **Description:** Metrics and performance monitoring are necessary during development to monitor efficiencies, track user activity and inform further work. They are necessary in operations for reporting, monitoring for problems and to direct further improvements. Monitoring and metrics will evolve with time and should be reviewed periodically.
- **Milestone 1:** Determine specifications for Metrics and Performance Monitoring.
  - **Task 1:** Determine metrics to be used in DMAC and requirements for performance monitoring
    - Category: Committee work
  - **Task 2:** Implement performance monitoring in DMAC systems
    - Category: Contract/Pilot
  - **Task 3:** Pre-operational
    - Category: Pre-operations

  - **Task 3:** Operational
    - Category: Ongoing operations
- **Deliverables:**
- **Estimated Resources:**
- **Schedule:** Year 2, activity will continue
- **Sequencing:**
- **Partnerships:**

7. Activity: Implement Middleware Security (Cross-discipline effort with all DMAC)

- **Description:** Implement Security tools, procedures and practices to protect all participating systems from inappropriate access, intentional or accidental. Data providers must be able to configure compute and network resource limits. Selected data streams may have restricted access.
- **Milestone 1:** Create middleware Security Plan
  - **Task 1:** Determine Data Transport Security needs with DMAC
    - Category: Committee work
- **Milestone 2:** Develop and deploy middleware Security Plan
    - Category: Contract
- **Deliverables:** Middleware Security Plan
- **Estimated Resources:**
- **Schedule:** Year 3
- **Sequencing:**
- **Partnerships:** Archive team, Metadata and Discovery team

8. Activity: Provide guaranteed geo-temporal-referenced browse for all IOOS data

- **Description:** Ensure that all IOOS data are viewable through common web browsers (e.g., Netscape®, Internet Explorer®) in the form of intelligible "working graphics" on demand and human-readable numeric tables, and that they may be subsetted and downloaded in a range of common file formats. This will provide a necessary overview tool for IOOS scientists and a common entry point for all users who wish to explore IOOS.
- **Milestone 1:** Guaranteed minimum geo-temporally referenced graphics and numeric listings for viewing, and formatted file subsets for downloading.
  - **Fast-track note:** An implementation of a guaranteed minimal uniform DMAC data visualization (graphics) and access mechanism is vital to the integration of IOOS. It is fundamental to the ability to make quick evaluations of the effectiveness of all aspects of the DMAC and the quality of the data. If a suitable candidate is available, it will significantly advance the IOOS to implement a fast-track solution to this component. The Plan must include procedures to evaluate the effectiveness of the solution adopted and either make adjustments to it or abandon it, if it proves unsuitable.

- **Task 1:** Adopt the NVODS Live Access Server (LAS) as initial, provisional, fast-track guaranteed minimum browse solution.
  - Category: LAS is IOOS-Pre-operational for gridded data sets and IOOS-Pilot for others.
- **Task 2:** DMAC evaluation and review for OPeNDAP
  - Category: DMAC Governance Committee
- **Task 3:** As needed, initiate complementary and/or alternative real-time delivery solutions as R&D activities, Pilots, or Pre-operational solutions.
  - Category: DMAC Governance Committee
- **Task 4:** Determine the suite of visualization styles and file formats and data comparison capabilities that should be guaranteed by IOOS.
  - Category: Governance committee
- **Task 5:** Implement additional visualization styles and file formats and data comparison capabilities as determined.
  - Category: Contract

- **Deliverables:** Guaranteed minimum visualization and download capabilities
- **Estimated Resources:**
- **Schedule:** Year 1-4
- **Sequencing:**
- **Partnerships:** Archive and Access, Metadata & Discovery, GIS experts, User Support

9. Activity: OPeNDAP-OBIS Integration

- **Description:** OBIS is a globally distributed network of systematic, ecological, and environmental information systems. Data held in associated archives will be seamlessly integrated with those data accessible via the OPeNDAP.
- **Milestone 1:**
  - **Task 1:** Develop design for integration of the systems.
    - Category: Contract
  - **Task 2:** Implement design
    - Category: Contract
- **Deliverables:** Documented software
- **Estimated Resources:** 0.5 FTE
- **Schedule:** Year 1
- **Sequencing:** Immediately
- **Partnerships:** OPeNDAP and OBIS project

10. Activity: Aggregation of unstructured (a.k.a. vector, point, sequence, or profile) data

- **Description:** Design and implement a server or suite of servers capable of aggregating unstructured data. This server(s) will be capable of aggregating data across sites as well as within a site.

- **Milestone 1:** Design and implement a server for mooring data – fixed location variable in depth and time.
  - **Task 1:** Develop a consistent data model for mooring data.
    - Category: Contract
  - **Task 2:** Convene workshop of mooring data providers to evaluate the model.
    - Category: Contract
  - **Task 3:** Implement test and document design.
    - Category: Contract
- **Milestone 2:** Design and implement a server for hydrographic data – fixed location and time variable in depth.
  - **Task 1:** Develop a consistent data model for hydrographic data.
    - Category: Contract
  - **Task 1:** Convene workshop of hydrographic data providers to evaluate the model.
    - Category: Contract
  - **Task 1:** Implement test and document design.
    - Category: Contract
- **Milestone 3:** Design and implement a server for underway data – variable in space, depth and time (ship and drifter data).
  - **Task 1:** Develop a consistent data model for underway data.
    - Category: Contract
  - **Task 2:** Convene workshop of underway data providers to evaluate the model.
    - Category: Contract
  - **Task 3:** Implement test and document design.
    - Category: Contract
- **Milestone 4:** Design and implement a general server for unstructured data.
  - **Task 1:** Evaluate feasibility of a general aggregation server for unstructured data.
    - Category: DMAC
  - **Task 2:** If appropriate, design a general aggregation server for unstructured data.
    - Category: Contract
  - **Task 3:** Implement test and document design.
    - Category: Contract
- **Deliverables:** Documented software
- **Estimated Resources:** 2.0 FTE plus 3 workshops of ~20 attendees.
- **Schedule:** Year 1 and 2
- **Sequencing:** Immediately
- **Partnerships:** Data archivists for unstructured data, software developer

11. Activity: Develop a generic OPeNDAP server for unsupported data formats

- **Description:** A significant fraction of data to be made available by IOOS participants will not be in standard data formats, formats for which OPeNDAP servers already exist. In this project, a configurable server will be developed that may be used with a wide range of data formats.
- **Milestone 1:** Develop a generic OPeNDAP server for unsupported data formats.
  - **Task 1:** Design the server.
    - Category: Contract
  - **Task 2** Implement, test, and document design.
    - Category: Contract
- **Deliverables:** Documented server
- **Estimated Resources:** 2.0 FTE
- **Schedule:** Year 1
- **Sequencing:** Immediate
- **Partnerships:**

12. Activity: OPeNDAP-GIS client and GIS-OPeNDAP server

- **Description:** A significant fraction of IOOS users are expected to be GIS users while much of the data generated as part of IOOS will not generally be accessible from GISs. In this project, OPeNDAP client(s) will be developed for one or more commonly used GIS systems and OPeNDAP servers will be developed for commonly used GIS data formats.
- **Milestone 1:** Develop an OPeNDAP server for commonly used GIS data formats such as GeoTIFF.
  - **Task 1:** Design OPeNDAP server for GeoTIFF.
    - Category: Contract
  - **Task 2:** Implement, test, and document design.
    - Category: Contract
- **Milestone 2:** Develop a GIS OPeNDAP client.
  - **Task 1:** Delineate the issues, which GISs should be targeted and the level of support (access) that is appropriate.
    - Category: DMAC
  - **Task 2:** Design OPeNDAP GIS client
    - Category: Contract
  - **Task 3:** Implement, test and document OPeNDAP GIS client
    - Category: Contract
- **Deliverables:** Documented software
- **Estimated Resources:** 4.0 FTE
- **Schedule:** Year 1 and 2
- **Sequencing:** Immediate
- **Partnerships:** GIS, DMAC

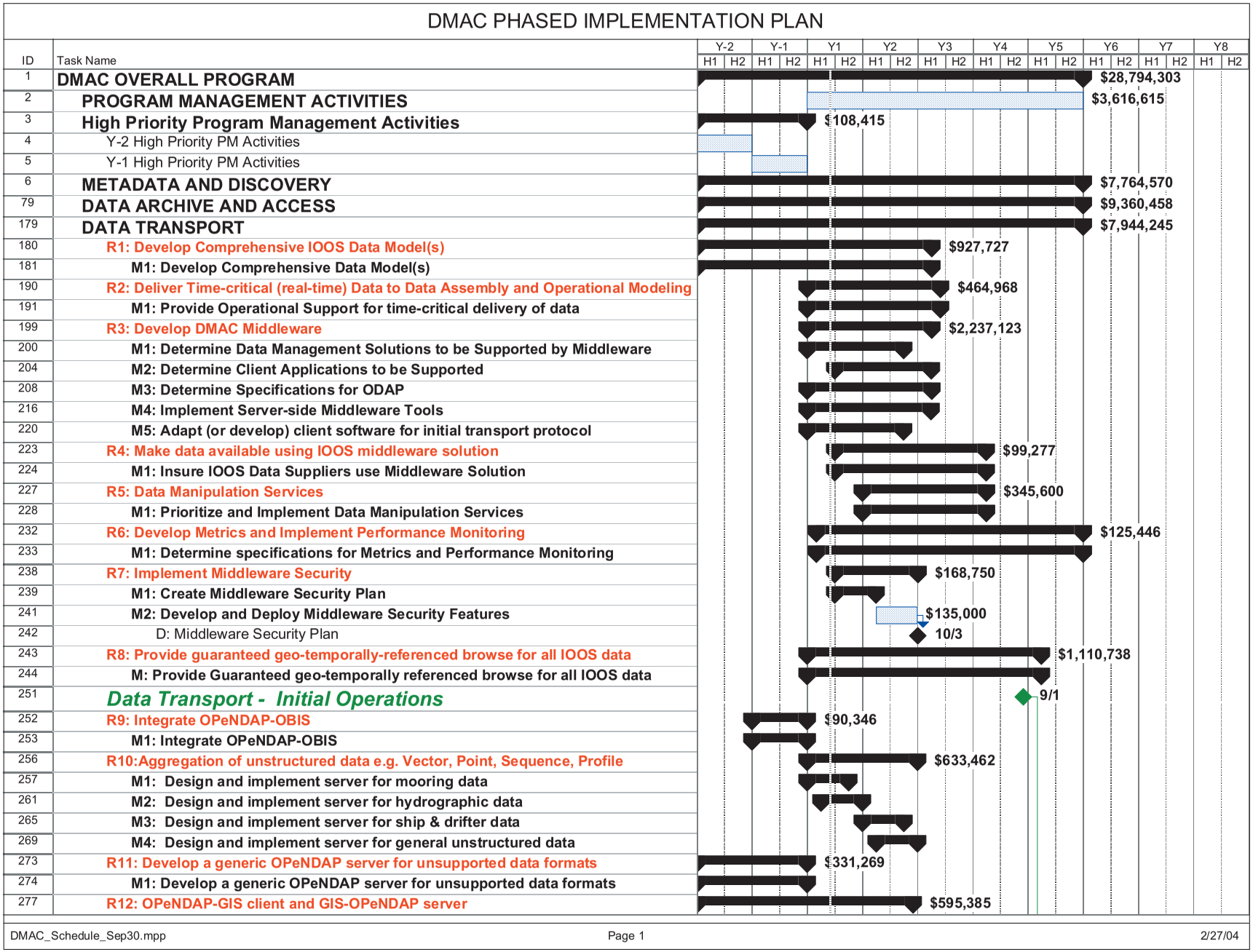
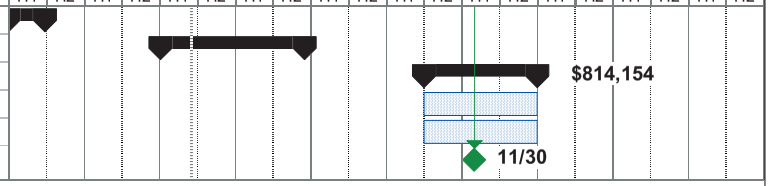


Figure 5. Data Transport Gantt Chart

DMAC PHASED IMPLEMENTATION PLAN

ID	Task Name	Y-2 H1	Y-2 H2	Y-1 H1	Y-1 H2	Y1 H1	Y1 H2	Y2 H1	Y2 H2	Y3 H1	Y3 H2	Y4 H1	Y4 H2	Y5 H1	Y5 H2	Y6 H1	Y6 H2	Y7 H1	Y7 H2	Y8 H1	Y8 H2
278	**M1: Develop an OPeNDAP server for common GIS data formats, e.g. GeoTIFF**																				
281	**M2: Develop GIS OPeNDAP Client**																				
285	**Data Transport Maintenance**																				
286	**Data Model Maintenance**																				
287	**Middleware Maintenance**																				
288	***DMAC - Initial Operations***																				



DATA ARCHIVE AND ACCESS ACTIVITIES AND SCHEDULE

(see Figure 6)

1. Activity: Current archive and access assessment

- **Description:** A comprehensive assessment of the current archive holdings and access methods is needed. A tabulation of data set name, content (temporal and spatial coverage), variables, format, storage location (online, offline, hardcopy), volume, resident center, and available access methods will form the starting benchmark for the IOOS. It will also uncover gaps in either the archive or access that need to be addressed and will point to archiving efforts where center-to-center collaborations would be beneficial.
- **Milestone 1:** Publish a current archive and access assessment report covering all U.S. centers holding IOOS-relevant data sets.
  - **Task 1:** Participate in metadata working group to define core metadata standards. Ensure that core development team has archive and scientific representatives from all IOOS data disciplines.
  - **Task 2:** Establish the set of data set descriptive parameters to be standard in the Archive System.
    - Description: To be effective, the set of descriptive parameters must be uniform across all participating data centers. There should be a two-way information exchange with the DMAC metadata development effort during this process. There are possibilities that this work could form the basis for the IOOS Discovery and Documentation metadata.
  - **Task 3:** Merge, tabulate, and evaluate the current status for all data sets and publish the findings.
  - **Category:** Tasks are committee work.
  - **Deliverables:** A published report on the starting benchmark for IOOS data archiving and access.
- **Estimated Resources:** Funding to cover costs for two or three meetings of eight to 10 people and publication.
- **Schedule:** During Year 1.
- **Sequencing:** Done prior to allocating or mapping of new data streams onto the existing set of data centers.
- **Partnerships:** Includes all U.S. IOOS data centers and is placed in the context of parallelisms, overlaps, and collaborations with GOOS as they apply.

2. Activity: Determine data set priorities for all IOOS data disciplines

- **Description:** The available and forthcoming data sets need to be ranked according to IOOS users' needs. Data set ranking is determined in conference with IOOS scientific representation, the DMAC User Outreach representatives, archive experts, and with reference to documentation on the most important variables as determined at the IOOS workshops. These lists set the priority for development efforts in the Archive System, and will lead to projects for improving archiving practices and access to both real-time and historical data sets.
- **Milestone 1:** Develop criteria for ranking data sets. Establish separate priority lists for each IOOS data discipline.
  - **Task 1:** Participate in the Transport "Data Population" group to identify and set priorities for ensuring accessibility of specific IOOS data sets.
  - **Task 2:** With an expert team and scientific representation prioritize extant archives according to IOOS needs.
  - **Category:** Tasks are committee work
  - **Deliverables:** A priority list for IOOS data sets.
- **Milestone 2:** Map the unfulfilled IOOS archiving needs onto the set of participating centers in the Archive System (see also the activity, Recruit Centers for the IOOS Archive System).
  - **Task 1:** Through a working group develop a plan to ensure all unarchived IOOS critical data become part of the Archive System.
  - **Deliverables:** A plan that maps IOOS data onto the Archive System.
- **Milestone 3:** List the products that are unavailable, but could be developed.
  - **Task 1:** Form a ranked list of data sets that could be developed. During this process a two-way sharing of information with the User Outreach component of the DMAC is required.
  - **Task 2:** In conjunction with the mapping exercise, assess the infrastructure capabilities at the centers that are to absorb additional archiving work. Document new infrastructure needs of IOOS support.
  - **Category:** Tasks are committee work.
  - **Deliverables:** A list of required products and estimated costs (by organization).
- **Estimated Resources:** Funding to cover costs for meetings.
- **Schedule:** Beginning 4th quarter during Year 1.
- **Sequencing:** Done immediately after (or possibly overlapping) the activity Current archive and access assessment.
- **Partnerships:** Inside IOOS (Archiving and Access (A&A) representatives, and Facilities Outreach members); Outside IOOS (scientific representation).

3. Activity: Determine IOOS data set categorization

- **Description:** Categorize IOOS data sets as irreplaceable, replaceable, perishable, and virtual. The data set storage strategy and retention period are determined by the categorization. In addition, adherence to Federal regulations at some centers in the Archive System is mandatory.
- **Milestone 1:** Assign data set category to all IOOS data sets.
  - **Task 1:** Convene a working group with experience in data archiving, metadata development, user outreach, and Federal regulations.
  - **Task 2:** Establish a process to review the categories set out in this document and that results in category assignments for all IOOS data sets.
  - **Task 3:** Assure irreplaceable data security.
    - It is planned that all irreplaceable data have two copies stored at separate locations and under independent data management. Irreplaceable data preserved below this standard are to be clearly documented, and archive centers need to immediately seek collaborations and support that resolve any deficiency.
  - **Category:** Tasks are committee work.
- **Deliverables:** A published report that categorizes all IOOS data sets.
- **Estimated Resources:** Funding to cover costs for meetings and publication.
- **Schedule:** Beginning 4th quarter in year one.
- **Sequencing:** Done immediately after the activity: Current archive and access assessment and possibly in conjunction with the activity. Establish data set priorities for all IOOS disciplines.
- **Partnerships:** Inside IOOS (A&A expert team, User Outreach Team, IOOS scientific representation, Metadata and Data Discovery Team)

4. Activity: Recruit centers for the IOOS Archive System and form partnerships

- **Description:** Effectiveness of IOOS will be achieved only by broad participation of the U.S. centers in the Archive System. Recruitment strategies need to be developed. Integration and cooperation with international programs are also critical. Global sharing of data will yield the maximum benefit to all programs, so the international contacts must be identified and actively engaged. IOOS must also be sensitive to the commercial 'value-added' data providers. This group will have objectives that overlap IOOS goals. IOOS should nurture partnerships based on open understanding and collaborations with this business sector.
- **Milestone 1:** Establish a set of guidelines for IOOS Archive System centers.
  - **Task 1:** Bring together the relevant IOOS governance and data policies into a set of guidelines so that interested centers can quickly know:
    - What is required to become part of the IOOS Archive System.
    - What are the benefits, for example, data sharing, backup, and archive.
    - What funding potential might exist for their centers.

The IOOS data policy document may not be in final form at this time, but a draft version could be used during the initial organization work.

- **Category:** DMAC policy (governance)
- **Deliverables:** Guideline document for Archive System centers
- **Estimated Resources:** Funding to cover meeting costs, publications, and potentially travel for a facilities management outreach liaison.
- **Schedule:** Beginning 3rd quarter Year 1 and onward
- **Sequencing:** Done after implementation is approved and data policies are in draft form.
- **Partnerships:** Inside IOOS (led by Facilities Management Outreach Team with assistance from A&A expert team, User Outreach Team, and IOOS governance)

• **Milestone 2:** Build international partnerships.

- **Task 1:** Establish and devise a way to maintain a list of international centers and programs that could be collaborating partners for IOOS.
- **Task 2:** Identify contacts, share understanding, and promote cross program partnerships and support.
- **Category:** Planning and outreach
- **Deliverables:** Document identifying relevant international partnerships and contacts
- **Estimated Resources:** Funding to cover meeting costs, publications, and potentially travel for outreach liaison to international meetings and program offices
- **Schedule:** Year 2 and onward
- **Sequencing:** Done in parallel with IOOS developments following initial U.S. organization efforts.
- **Partnerships:** To be defined

• **Milestone 3:** Evaluate and plan for commercial overlaps.

- **Task 1:** Survey the commercial data business and identify overlaps with the IOOS goals. Assess and suggest way to integrate the business efforts with IOOS development so as to best serve the public needs.
- **Category:** Planning and outreach
- **Deliverables:** Document identifying relevant commercial interest overlaps.
- **Estimated Resources:** Funding to cover meeting costs and publications.
- **Schedule:** Year 2 and onward
- **Sequencing:** Done in parallel with IOOS developments following initial U.S. organization efforts.
- **Partnerships:** Inside IOOS (IOOS governance and access experts); Outside IOOS (commercial interests)

5. Activity: Develop archive critical metadata

- **Description:** To aid and ensure systematic (human and machine) access and data management across IOOS some archive specific metadata are critical. As a limited and brief example, some important elements are:
  - Unique data set identification code;
  - Expiration date;
  - Data set lineage and version history;
  - Points of access and available access methods;
  - Data set citation and references;
  - Data set latency specification.
- **Milestone 1:** Develop a comprehensive list of archive critical metadata and organize a plan that will lead to a system-wide metadata standard that is easy to implement and maintain.
  - **Task 1:** Convene a working group with experienced representation for data archiving and metadata development to prepare the list.
  - **Task 2:** Interact with the DMAC Metadata and Discovery Data and Data Transport development teams and working groups to ensure archive needs are accommodated.
  - **Category:** Tasks are committee work.
  - **Deliverables:** Archive specific metadata will appear in IOOS metadata standards.
  - **Estimated Resources:** Funding to cover committee work and collaboration with the DMAC Metadata and Discovery Data team.
  - **Schedule:** Beginning in Year 1 and continuing.
  - **Sequencing:** Done prior to completion of IOOS metadata standards work.
  - **Partnerships:** Inside IOOS (A&A expert team, and Metadata and Data Discovery Team, Data Transport Team).

6. Activity: Define IOOS archive and access data policy

- **Description:** The policies for contributing data and using data from IOOS need to be formally documented. A few key policy issues from the archive and access perspective are:
  - Full and open data sharing as per IOC and WMO policy and at no cost or minimum cost for reproduction. The conditions and authoritative protocol for allowing restricted access need to be discussed and defined, if required.
  - Data collected or prepared with IOOS funding must be placed in the Archive System. If possible, actions to be taken for non-compliance should be articulated.
  - Agreement that the four data categorizations are suitable to determine IOOS data preservation requirements.
  - Full cooperation with GOOS.

- **Milestone 1:** Develop a draft IOOS data policy.
  - **Task 1:** Form a committee with experienced representation for data archiving and IOOS management to draft the policy.
- **Milestone 2:** Receive community comment on the draft policy.
  - **Task 1:** Circulate the policy to interested parties and make it widely known that IOOS has a data policy.
- **Milestone 3:** Create a final draft.
  - **Task 1:** Resolve problems raised by the IOOS user community and ensure the data policies can be applied satisfactorily within the limits of standing Federal regulations.
- **Category:** Milestones and all Tasks are committee and IOOS management work
- **Deliverables:** IOOS data policy
- **Estimated Resources:** Funding to cover committee work.
- **Schedule:** Year 1
- **Sequencing:** Done before or at the same time as the activities of Develop archive critical metadata and Current archive and access assessment. A draft of the data policies is necessary for the data center recruitment and partnership requirement.
- **Partnerships:** Inside IOOS (A&A expert team, Metadata and Data Discovery Team, and IOOS management); Outside IOOS (done with consultation to GOOS).

7. Activity: Establish IOOS data stream developers guidelines

- **Description:** A document defining the IOOS data stream guidelines must be available for data providers. It will include at least:
  - IOOS archive and access data policy;
  - IOOS metadata and data discovery standards and recommended ways to easily develop the metadata;
  - IOOS recommended formats;
  - IOOS data transport standards and recommended ways to implement them and get support.
- **Milestone 1:** Publish a guideline document that is updated as progress and evolution in the DMAC take place.
  - **Category:** Milestone is committee and IOOS management work.
  - **Deliverables:** IOOS data stream developers' guidelines
- **Category:** Archive committee work
- **Estimated Resources:** Funding to cover committee work
- **Schedule:** Year 2
- **Sequencing:** Done after the data policy, data transport, and metadata and discovery data standards are in beta release form.
- **Partnerships:** Inside IOOS (A&A expert team, Metadata and Data Discovery Team, and IOOS management).

8. Activity: Develop Archive System data discovery interfaces

- **Description:** The IOOS Archive System will be distributed. Data discovery (both machine and human) must work across all centers, all data sets, and all methods for access.
- **Milestone 1:** Define the human data discovery interface
  - **Task 1:** Have archive component representatives work with the Metadata and Data Discovery component during the interface development. Some necessary interface features are:
    - It is dynamic, formed based upon user query;
    - It shows which data centers hold the data;
    - It shows the data set titles and unique IOOS data set identification;
    - It shows available core services (OPeNDAP, HTTP, FTP);
    - It shows available extended services (subsetting, aggregation, OpenGIS Map, online analysis, online ordering, etc.);
    - It provides users with links to all services that are available.
  - **Category:** R&D
  - **Task 2:** Dynamically harvest metadata from the Archive System and build IOOS metadata databases to serve the discovery interfaces.
  - **Category:** R&D
  - **Task 3-n:** Other tasks to be defined as standards and methods are developed by the Metadata and Data Discovery and Data Transport components. Following R&D, pilot projects will be necessary.
- **Milestone 2:** Define the machine data discovery interface.
  - **Task 1:** The machine, or application, interface will be specified by collaborations between the Data Transport, and Metadata and Data Discovery components. They are not described here.
- **Deliverables:** Data discovery interfaces for the Archive System.
- **Estimated Resources:** To be determined by other experts.
- **Schedule:** Beginning 1st quarter Year 2 and continuing.
- **Sequencing:** Done following the development of metadata standards and data transport methods.
- **Partnerships:** Inside IOOS (all components); Outside IOOS (other organizations that are attempting to do the same thing).

9. Activity: Receive and provide more data in real time

- **Description:** To meet the IOOS goals the Archive System must receive and provide more data to users in real time. Many IOOS goals have time critical schedules requiring prompt access to observed data and data products. Note: The provision of real-time data will come primarily from the modeling centers, regional centers, and data assembly centers in the Archive System. More limited real-time access will be the norm at the archive centers.

- **Milestone 1:** Put in place pilot projects and pre-operational real-time data systems based on DMAC data transport mechanism and existing data delivery infrastructure.
  - **Task 1:** Based on the activities of "current archive and access assessment" and "determine data set priorities for all IOOS disciplines," select several real-time data streams that are most important.
  - **Task 2:** Design pilot projects within the Archive System that must include the components for real-time data receipt and immediate public access.
  - **Task 3:** Depending on the advances in DMAC Data Transport, Metadata and Data Discovery, and Archive System collaborations, the following could also be components within the projects:
    - Data receipt and delivery through DMAC data transport methods;
    - Metadata records and catalogs in the DMAC standard;
    - Archive System data transfer to an archive center if initially received at a regional, assembly, or modeling center;
    - Value added data development (QC checks, and data merging, provision for server side subsetting), and product development (data analysis, maps, and data formatted for GIS ingest).
    - Archive System backup at a second archive center for irreplaceable data
  - **Category:** Tasks 1-3 are pilot study leading to pre-operational systems
  - **Deliverables:** Real-time test systems
  - **Estimated Resources:** Staff and facilities infrastructure commensurate with data source volume, complexity, and access service.
  - **Schedule:** During Year 2.
  - **Sequencing:** Done following the activities of "current archive and access assessment" and "establish data set priorities for all IOOS disciplines."
  - **Partnerships:** Internal to the IOOS DMAC and probably in collaboration with developments in IOOS measurement component.
  - **Task 4:** Pilot project to serve near-real-time GTSPP data.

    As a pilot project, near real-time profile data, which are harvested from the Global Telecommunications System (GTS) and delivered three times per week from the Marine Environmental Data Service (MEDS) of Canada to the U.S. National Oceanographic Data Center (NODC) as part of the Global Temperature Salinity Profile Program (GTSPP), will be posted by NODC on a DODS server (a precursor to the IOOS data transport protocol) for use by operational oceanographic data customers. For each data set delivered, NODC will create metadata, archive the data set according to its data category requirement, subject the data set to preliminary quality control, and then post the data set on a DODS server. After these data

have been available via the DODS server for several months, customer feedback will be evaluated to determine subsequent improvements in this data system and whether the metadata are adequate for effective use of the data.

This pilot system will provide practical information for the DMAC developments in data transport, metadata, and data discovery. It would also be well positioned to transition to the IOOS standards as they become available.

- **Category:** Pilot
- **Deliverables: (**1) Near-real-time profile data available on a DODS server, (2) compilation report of customer feedback, and (3) data system improvement plan.
- **Estimated Resources:** Since this is an extension of an existing base funded project at NODC, no additional resources are requested for this pilot project. Transition to IOOS functionality may require additional support.
- **Schedule: (**1) March 2003, (2) August 2003, and (3) October 2003.
- **Sequencing:** Independent of other tasks.
- **Partnerships:** Internal to the IOOS DMAC, and in collaboration with MEDS and near real-time profile data customers such as the Naval Oceanographic Office.

10. Activity: Establish a protocol to report and resolve data and data flow problems

- **Description:** Inevitably, there will be problems with the data flows and data set integrity. These problems will have wide impact. Irregularities and changes will affect the data providers, the Archive System, the metadata, and most importantly, the data users.
- **Milestone 1:** Establish a protocol for reporting and resolving problems.
  - **Task 1:** Establish a method to post or broadcast problems to users as soon as possible after they are discovered. Use the same strategy to publicize when corrections to the real-time data stream are completed.
  - **Task 2:** Establish a method to publicize, in delayed mode, the analysis of problems and identify the impacted data in the Archive System.
  - **Task 3:** Establish a method to publicize, in delayed mode, substantive corrections that have been applied to the archive data.
  - **Category:** Committee work
- **Milestone 2:** Determine effective ways to solve chronic problems.
  - **Task 1:** Investigate if IOOS management through Facilities Management Outreach can assist in resolving chronic problems that are not addressed in a reasonable time period.
  - **Category:** Committee work
- **Milestone 3:** Implement a DMAC system for data problems.
  - **Task 1:** Pilot study leading to pre-operational then operational reporting system.

- **Deliverables:** A system and protocol through which problems can be reported, resolutions can be sought, and data providers and users can be informed.
- **Estimated Resources:** Staff and facilities infrastructure commensurate with the pilot project and services provided during the pre-operational and operational phases.
- **Schedule:** Beginning 4th quarter of Year 1.
- **Sequencing:** Done after the IOOS Archive System activity of Current archive and access assessment and when the IOOS data streams are coming online in the DMAC.
- **Partnerships:** Internal to the DMAC, and in collaboration with IOOS management and Facilities Management Outreach.

11. Activity: Broaden the base for user services

- **Description:** The IOOS is required to provide new services to many new users. In particular, decision- and policy-makers need rapid access to suitable ocean information. Plans are needed to determine which services are missing and most critical, and how to provide those services from the Archive System.
- **Milestone 1:** Accommodate new user services in the Archive System.
  - **Task 1:** Participate in the plans developed by the User Outreach component and determine how to improve extant and expand user services.
  - **Category:** Committee work (with IOOS User Outreach), R&D, and pilot.
- **Deliverables:** New services to meet the needs for the broad IOOS user base.
- **Estimated Resources:** Commensurate with the data products and systems needed for development.
- **Schedule:** Beginning 4th quarter in Year 2.
- **Sequencing:** Done following the gathering of advice from the User Outreach component and after completion of the activities, "current archive and access assessment" and "establish data set priorities for all IOOS disciplines."
- **Partnerships:** Inside IOOS (Archive and Access, User Outreach).

12. Activity: Verify data security requirement for irreplaceable data sets

- **Description:** It is intended that irreplaceable data be stored in two separate locations, preferably under independent data management.
- **Milestone 1:** Run live tests to verify the security of irreplaceable IOOS data.
  - **Task 1:** Devise a DMAC pilot project to randomly check primary and secondary archive copies between centers for data in the irreplaceable category.
  - Category: Pilot project
  - **Task 2:** Execute data and inventory checks, and cross system archive file recovery.
  - **Category:** Pilot project

- **Milestone 2:** Establish regular verification procedures.
  - **Task 1:** Document and form cross-agency agreements to perform archive checks.
  - **Category:** Pre-operational and operational
  - **Deliverables:** Systematic methods to verify irreplaceable data security.
  - **Estimated Resources:** Commensurate with the pilot projects proposed.
  - **Schedule:** During Year 3.
  - **Sequencing:** Done following the activity to Determine IOOS data set categorization, and after data transport capabilities are available.
  - **Partnerships:** Inside IOOS (A&A expert team, IOOS management, and Facilities Outreach)

13. Activity: Establish procedures to document the Archive System metrics

- **Description:** From the onset, measurement metrics for the Archive System are required. Tracking of incoming and outgoing data must be thorough, complete, and measured on equivalent scales system wide.
- **Milestone 1:** Compile a test annual report for the DMAC Archive System.
  - **Task 1:** As a pilot project, collect annual metric data from all Archive System centers. Work to develop an annual report. Through the process identify problems and needs to achieve accurate reporting. Use these results to refine the metric requirements for the system. General elements for consideration, review and refinement are:
    - media receipt and delivery;
    - network receipt and delivery;
    - DMAC data transport receipt and delivery;
    - error discovery and correction.
  - **Category:** Pilot project
- **Milestone 2:** Include the Archive System metric requirements in the guideline document for IOOS data stream developers.
  - **Category:** Committee work
- **Deliverables:** DMAC system wide data movement metric report.
- **Estimated Resources:** Funding for meetings, and pilot project work.
- **Schedule:** Beginning 2nd quarter in Year 2.
- **Sequencing:** Done after the data transport mechanism is established, data sets are categorized, and suitable metadata are available.
- **Partnerships:** Internal to IOOS

14. Activity: Improved efficiency for archive growth and access

- **Description:** A concerted effort to implement DMAC Data Transport methods and Metadata and Data Discovery standards is required and will result in improved efficiency within the Archive System.
- **Milestone 1:** Implement DMAC Data Transport receipt methods in parallel with extant methods.
  - **Task 1:** Pilot and Pre-operational, critically compare data received for six months to a year. Assure all aspects of data integrity, track system performance, and problems.
- **Milestone 2:** Implement DMAC Data Transport and Data Discovery in parallel with extant delivery systems.
  - **Task 1:** Pilot and Pre-operational, offer data access through extant interfaces and procedures, and through DMAC Data Discovery interfaces and Data Transport protocols. Track all user metrics, problems, and success. The pilot and pre-operational systems should run for at least one year each.
- **Category:** Pilot and pre-operational
- **Deliverables:** Archive System tests of DMAC functionality.
- **Estimated Resources:** Commensurate with Pilot and Pre-operational plans.
- **Sequencing:** Following beta tests on DMAC Transport and development and testing of Metadata and Data Discovery standards. Probably Year 2 and onward.

15. Activity: Procedure to resolve data retention issues

- **Description:** Data in archive systems are commonly resubmitted and replaced. The number of old versions of data to be preserved remains an open question. Managers of data centers need a formal procedure to help them resolve this difficult issue.
- **Milestone 1:** Reach a consensus agreement on how to resolve data retention issues and questions for data sets with multiple versions.
  - **Task 1:** Form a diverse expert team with representatives from archive management and the scientific user community.
    - Discuss and draft a set of procedures on data retention that possibly include:
      - An annual scientific review panel to consider data retention questions;
      - Input from IOOS management;
      - An adequately long public review period before any action is taken.
  - **Category:** Committee work (with IOOS policy implications)
  - **Task 2:** Develop schedule for removal of data sets identified for removal.
    - This should require approval from science, governances and parent agency.
    - This affects Data Discovery and Data Transport as they will need to make alterations to remove reference and support for that data.

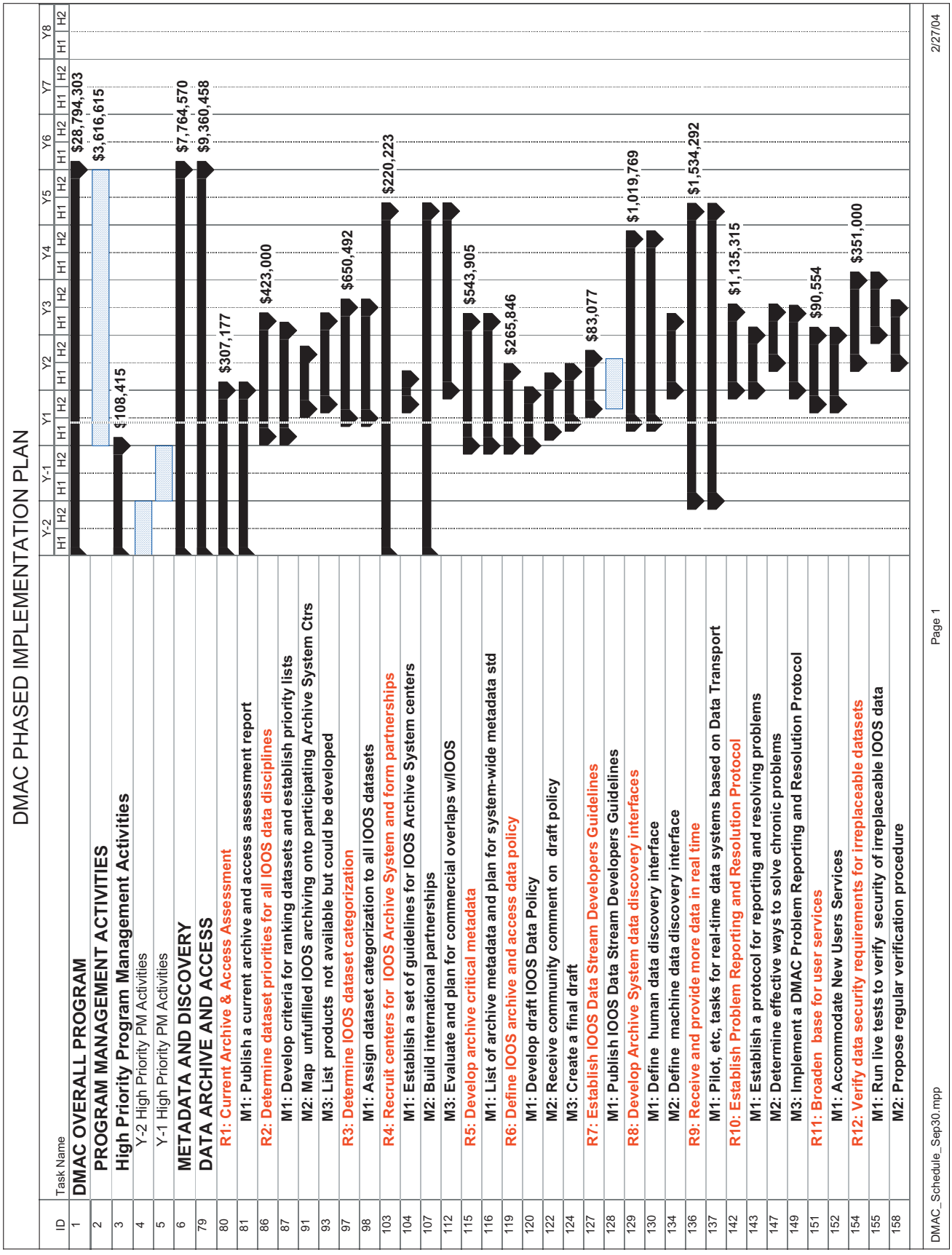
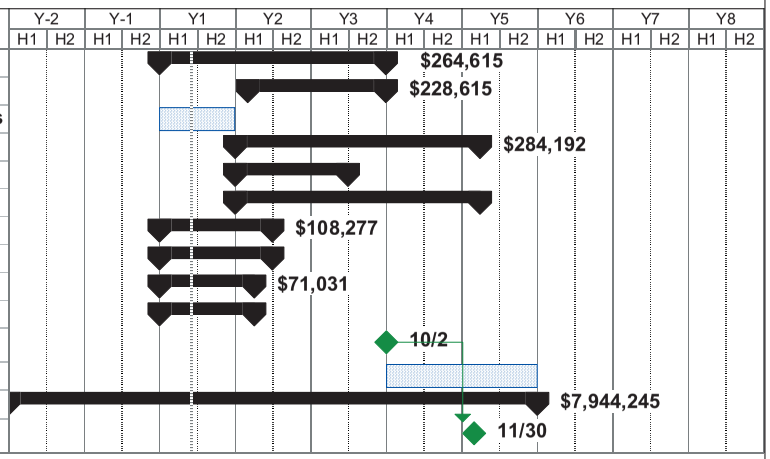


Figure 6. Data Archive and Access Gantt Char

DMAC PHASED IMPLEMENTATION PLAN

ID	Task Name
160	R13: Establish procedures to document the Archive System metrics
161	M1: Compile a test annual report for the DMAC Archive System
163	M2: Write Archive System Metric requirement for IOOS data developer guidelines
164	R14: Improved efficiency, archive growth, and access
165	M1: Implement DMAC Transport receipt methods
167	M2: Implement DMAC Data Discovery and Transport delivery methods
169	R15: Procedure to resolve data retention issues
170	M1: Consensus resolving data retention issues
173	R16: Write Plan for ARCHIVE&ACCESS Security
174	M1: Develop Archive & Access Security Plan
177	*Archive & Access - Initial Operations*
178	R17: Initial Archive Operations
179	DATA TRANSPORT
288	*DMAC - Initial Operations*



- **Deliverables:** Procedures to address data retention questions
- **Estimated Resources:** Funding for meetings.
- **Schedule:** During Year 1 or 2.
- **Sequencing:** Done after IOOS data archive and access policies have been established.
- **Partnerships:** Internal to IOOS, with scientific representative input.

16. Activity: Write plan for Archive and Access Security

- **Description:** The Archive System will publicly expose data and systems. To protect the data suppliers and the systems a security plan is required.
- **Milestone 1:** Develop an Archive System security plan.
- **Task 1:** Determine the level of security that is required for the DMAC and the data held for IOOS. Write a security plan.
- **Category:** Working Group activity
- **Deliverables:** Archive and access security plan
- **Sequencing:** In conjunction with the evolution and deployment of DMAC Data Transport and Metadata and Data Discovery standards in the Archive System.