U.S. IOOS Program – Operations Division
Implementing Data Management and Communications (DMAC) in FY2012-13

The DMAC subsystem is the primary mechanism for data integration required for the U.S. IOOS to function effectively. DMAC will be developed with the intent of integrating all projected data sources and with flexibility to integrate future data sources. Which data sources are integrated will be determined consistent with user requirements, policy, and standards. The DMAC subsystem represents the primary direct equipment/material responsibility of the U.S. IOOS Program Office for development, deployment, and sustainment.

U.S. Integrated Ocean Observing System:
A Blueprint for Full Capability – Version 1.0 – November 2010

Introduction

This document describes eight DMAC activities being conducted in FY2012-13 by the U.S. IOOS Program Office (the program) to improve overall access to and use of ocean observation data and is based on a detailed, multi-year plan for DMAC execution developed by the Office[1]:

1. Refining existing web services for access/delivery of ocean observation data;
2. Building an enhanced U.S. IOOS data portal and catalog;
3. Sustaining our technical collaboration with NDBC and CO-OPS;
4. Improving use of controlled vocabularies;
5. Launching QARTOD (Quality Assurance for Real Time Ocean Data) for U.S. IOOS – a sustainable quality assurance/quality control process;
6. Implementing DMAC for biological observations in U.S. IOOS regions;
7. Improving access to animal telemetry observations by establishing data standards; and
8. Establishing baseline standards for glider operations and data.

Background

U.S. IOOS partners collect and distribute many different types of ocean information including observational data such as in-situ and remotely sensed measurements, model outputs, and derived products such as forecasts of ocean conditions and educational tools. This information is sometimes made available in near real-time, or can be historic or archived records subject to significant post processing. Examples of *in situ* ocean measurements include those gathered from fixed moorings, gauges, gliders, expendable bathythermographs (XBT), Conductivity-Temperature-Depth profilers (CTD), and animal borne sensors (ABS). How these various data are formatted, documented, exposed and accessed can vary widely which presents barriers to their wide use. The fundamental objective for DMAC is to establish a standards based framework for discovering, distributing and using ocean

[1] http://www.ioos.gov/library/dmac_implementation_v1_0.pdf

observation data. Within this framework, individual standards are adopted, adapted, or developed to address specific objectives of the system

For the past five years the U.S. IOOS Program's 11 Regions have established data assembly centers and data portals that enable access to ocean observations and derived products from a wide range of data providers including federal, state, local and tribal governments, academia, NGOs and industry.

From 2007-2010, the U.S. IOOS Program Office conducted a 36 month pilot project – the Data Integration Framework (DIF) – to systematically evaluate existing and emerging computer tools for improving the interoperability of ocean observation data. A key feature of this work was the adoption of the Open Geospatial Consortium's (OGC) Sensor Observation Service (SOS), an emerging, cutting-edge web service specification for access to ocean data, particularly *in situ* observations. Through the DIF, prototype SOS services were successfully established and tested.

The DIF effort demonstrated that a standards-based web service such as SOS was feasible to implement and many valuable lessons were learned, including important modifications to enable broader adoption across all IOOS Regions and other key data providers. Therefore, the major focus in the next 24 months will be to improve the existing web service tools and standards and have them adopted by all 11 IOOS Regions.

In addition, the program is actively increasing the data content available via these services. Specifically, we will focus on adding biological data over the next 24 months by developing data standards and access services for biological variables (e.g. fish species, fish abundance, zooplankton species and zooplankton abundance). Biological observations are highly heterogeneous and the variety of formats and sampling methods create significant challenges. The IOOS Biological Observations Project completed in 2011 addressed some of these challenges and developed DMAC standards that pertain to biological observations as well as implemented data access services, based on OGC standards and Environmental Research Division's Data Access Program (ERDDAP), to make these data accessible.

In the following sections, eight specific projects are described. These projects, conducted by the program, advance the implementation of DMAC across the enterprise, with some emphasis on interactions with the regional associations.

1. Refining existing web services to ensure enterprise-level interoperability

What are we doing?

- Writing an IOOS profile of the SWE Common data model encoding standard[2], to be used to serve *in situ* data from a Sensor Observation Service;

[2] The Sensor Web Enablement (SWE) Common Data Model Encoding Standard defines low level data models for exchanging sensor related data between nodes of the OGC® Sensor Web Enablement (SWE) framework (http://portal.opengeospatial.org/files/?artifact_id=41157). A profile is an implementation of a general standard that is made specific and tailored to the needs of a community of practice such as IOOS®.

- Actively engaging the regional developers in the process above to ensure that the regional requirements are accounted for and to leverage the considerable expertise that exists in the regional associations;
- Adopting and modifying a commercially developed open-source software solution, "SOS reference implementation" that can be implemented by IOOS data providers to serve *in situ* data. IOOS data providers will have the option of using this reference implementation, thus easing configuration management and future enhancements, or they can choose to develop their own backend services complying with the encoding standard such that queries of their data will yield results compatible with those using the SOS reference implementation;
- Providing support to National Data Buoy Center (NDBC), the Center for Operational Oceanographic Products and Services (CO-OPS), 11 IOOS Regional Associations and U.S. Army Corps of Engineers (USACE) with implementing the new IOOS encoding profile.

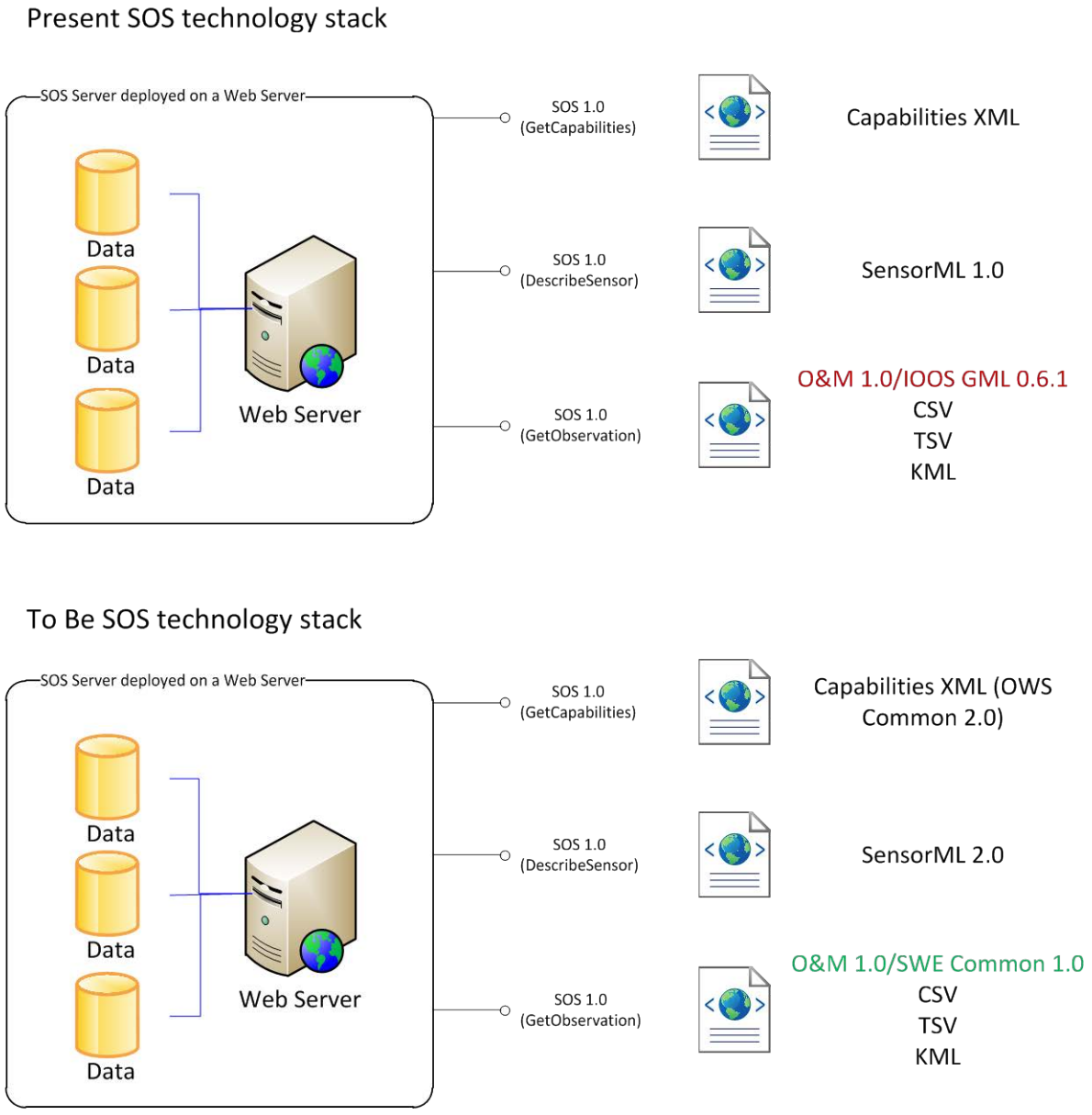

Figure 1: Present (upper) and desired (lower) implementation of the SOS technology and standards implementation configuration. Currently, many regions have some flavor or dialect of SOS deployed, but interoperability is greatly enhanced when there is a homogeneous set of services and data formats provided by those services. The reference implementation effort and the SWE Common migration will result in upgrading and homogenizing several of the standards that are relevant to an SOS deployment. The most important enhancement will be a move from the IOOS DIF GML observation format to the new SWE Common 1.0 based observation format. Additionally, the GetCapabilities and DescribeSensor output formats will be upgraded to the newest version. There are several different code families or lineages that implement all of the above functions. Standardizing that code base has benefits but is a secondary goal relative to migrating to the new standards implementation.

Why is this important?

- A nationally uniform set of data access services form the foundation of an integrated system of Regional and National Data Providers;
- This foundation enables easier development of additional value added DMAC services including a richly-populated Catalog, and data discovery services and easier integration with other national programs;
- Providing SOS reference implementation software and an IOOS encoding profile, and maintaining this software at the national level, allows for rapid expansion of IOOS data providers;
- By maintaining version control within the US IOOS Program Office we will maintain integration;
- Assisting federal and regional partners with deployment of the software reference implementation ensures there will be a robust network of interoperable data providers during the FY12-FY13 timeframe;
- This integration allows anyone to access US IOOS data, across US coastal waters and the Great Lakes, for all IOOS variables; this data can then be immediately used within a model, or decision support product.

How will we execute?

- A working group comprised of developers from NERACOOS, MARACOOS, SECOORA, NANOOS, AOOS, CO-OPS, NDBC and USACE have been collaborating via phone, Google docs and in person to define the new IOOS encoding profile for *in situ* mode data. This collaboration will culminate in early February 2012 with a workshop to finalize all the open technical issues and ultimately results in publication of the encoding profile.
- Axiom, the data management lead for AOOS and under contract through the AOOS cooperative agreement, will adapt and expand existing open source software components to be compatible with IOOS-adopted standards for serving *in situ* data.
- IOOS will publish the new encoding profile and will require its use with partner data providers; partners will implement the profile, either on their own or by using IOOS' software reference implementation.
- IOOS will develop a process to assist interested Regional Associations with installation of the reference implementation.

What will be the results in the next 24 months?

- Publication of the IOOS encoding profile for in situ data [FY12]
- Implementation of the IOOS encoding profile by NDBC, USACE (Mobile District), CO-OPS and the five regions listed above [FY12]
- Define timeline for adoption of IOOS encoding profile by the remaining Regional Associations [FY12]
- Release of the SOS reference implementation, fully documented such that it can be implemented by any data provider [FY12]

- Implementation of the SOS reference implementation at several data providers (GLOS, GCOOS and the NOAA Chesapeake Bay Office have volunteered to install the reference implementation software); [FY12]
- Develop at least one data product, such as a dash board or mashup, that shows the utility of this interoperability – e.g. daily SST from all coastal and Great Lakes assets [FY13]
- Complete implementation of SOS IOOS encoding profile by all 11 Regions [FY13]

2. Building an enhanced U.S. IOOS Data Portal

What are we doing?

Work with the Gulf of Maine Research Institute (GMRI) and the National Geophysical Data Center (NGDC) to enhance the IOOS Data Portal (aka the Catalog), which functions as an Internet portal for users to discover data and metadata for IOOS assets. The IOOS Data Portal is actually a system of DMAC components comprised of the Service Registry, Data Catalog, System Viewer, Data Integration Service, and System Monitor. Enhancements include:

- Online, form-based method for data providers to register their services and metadata in the Portal
- Automation of standards compliance testing to provide data providers with feedback regarding their level of compliance
- Improved metadata evaluation tools to help data providers improve metadata
- Expanded metadata management and data discovery via integration with ESRI's GeoPortal server
- Basic aggregation of *in situ* data across multiple data providers
- More reliable data provider service monitoring and reporting
- IT infrastructure performance improvements and support for the new IOOS encoding profile
- Identification of additional Portal requirements to be implemented during FY13

Why is this important?

- The Portal is the public face of the national U.S. IOOS enterprise. It is the primary method for users to visualize IOOS as an integrated national "system" and because the Portal rests on top of the web services and standards IOOS has adopted, it is an example of the interoperability a functional DMAC subsystem can provide.
- The Portal also has an important role in enabling active management of the system. For example, among the more critical goals for the Portal in the near term, is to enable the automation of the annual observations asset inventory.

How will we execute?

- IOOS has funded GMRI and NGDC to make the necessary enhancements. Their work is guided by detailed Statements of Work (SOW) to complete the tasks listed above.
- Open source software, including ESRI's Geoportal server, is leveraged wherever possible to maximize cost effectiveness and interoperability with other NOAA and IOOS community initiatives

What will be the results in the next 24 months?

- Specific guidance for authoring the metadata that drives the Catalog functions and registers sensors and data providers [FY12]
- Self-registration function/page that enables system growth and simplifies overall operation and maintenance [FY12]
- Simplified map-based interface to illustrate the breadth of IOOS observing assets including direct links to data sources [FY12]
- More robust and functional IOOS Portal with an expanded set of data services and assets [FY12]
- Enhancements for FY13 work will be identified by April 2012 [FY12]
- Automated update of the asset inventory [FY13]

3. Sustaining our technical collaboration with NDBC and CO-OPS

What are we doing?

- Integrating DMAC developments from federal partners with those from the Regional Associations by including all parties in development efforts like the IOOS SOS reference implementation development team;
- Both will also work on additional DMAC technical developments including developing improved metadata describing QA/QC procedures and data processing history;
- Continuing to expand the data sets and volumes served through their data access services;
- Increasing the breadth and clarity of their service documentation on their web sites to make it easier to use their services;
- Continued Operations and Maintenance of their IOOS services.

Why is this important?

- NDBC and CO-OPS are vital operational data providers of ocean observations and early adopters of US IOOS DMAC protocols and services.

How will we execute?

- IOOS has funded NDBC and CO-OPS to make enhancements to their Sensor Observation Services consistent with the reference implementation work described previously; detailed Statements of Work (SOWs), requirements and frequent meetings are used to manage the progress of development;
- Representatives from NDBC and CO-OPS are participating in the reference implementation working group along with regional developers
- NDBC and CO-OPS routinely participate in the annual RA DMAC workshops

What will be the results in the next 24 months?

- NDBC and CO-OPS will implement IOOS profile of the SWE data encodings in their Sensor Observation Services [FY12]
- Data delivered through their SOS services will be accompanied by quality control flags [FY12]
- CO-OPS will be adding High and Low Tide predictions, Relative Humidity and Rainfall data to their SOS and NDBC will be adding Volunteer Observing Ship data to theirs. [FY12]
- NDBC will develop a utility web service that translates between WMO identifiers and regionally specific platform identifiers so that users can identify duplicate data.

- NDBC will develop a performance testing function that will provide compliance and monitoring information for all of the SOS services in the IOOS enterprise. [FY12]
- Both will be developing and/or documenting client access tools to demonstrate how to use their data access services, particularly their SOS services. [FY12]

4. Improving use of controlled vocabularies

What are we doing?

- Writing a policy for the present use, future extension and maintenance of controlled vocabularies and the process for implementation.

Why is this important?

- Controlled vocabularies are a managed set of terms and definitions that communicate concepts germane to a community of interest.
- Controlled vocabularies provide organizations and structure that make data easier to find, use and relate between different disciplines, organization, or communities (e.g. atmospheric, oceanic, terrestrial), which facilitates interoperability.
- Controlled vocabularies form the basis for advanced discovery capabilities that require translations between communities (for example, one user may search for *wtemp* and another for *sea_water_temperature* but both are interested in data describing the temperature of water)

How will we execute?

- A working group of vocabulary experts will collaborate on the draft policy document.
- A finalized process will be published on the U.S. IOOS Program Website for use by our IOOS Partners.
- The IOOS Program will engage the IOOS RA DMAC Coordinators to implement this policy.

What will be the results in the next 24 months?

- A community developed policy that governs the present and future use of controlled vocabularies [FY12]
- A document with clear guidance for DMAC developers [FY12]
- Implementation of the process in the Regional Associations [FY13]

5. Establishing QARTOD for U.S. IOOS – a sustainable quality assurance/quality control process

What are we doing?

- Starting in February 2012, US IOOS will initiate a sustainable, community-based process to establish authoritative procedures for quality assurance and quality control (QA/QC) for the measurement of ocean conditions.
- This process replaces, and expands and improves, an existing community-based QA/QC process known as QARTOD and adopts the same name.
- Activities will be based on a detailed work plan developed in close collaboration with the US IOOS community and previous QARTOD leadership.
- The initial focus will be real-time observations from instruments.

- Establish QA/QC procedures for each of the 26 IOOS core variables, when necessary, including detailed information about the sensors and procedures used to measure the variables;
- From the list of individual QA/QC procedures developed, define a minimum set of QA/QC procedures that will be used for certification of RA data providers. These procedures will eventually be incorporated into data provider certification guidelines in an appropriate manner as they are developed.

Why is this important?

- All of the known QA/QC programs in existence today provide parts to the solution, but none accomplishes the task of putting it all together.
- Reliable and practical community-based QA/QC procedures are an essential component of a successful observation program.
- These procedures are also anticipated as components to the data-provider certification process required by the ICOOS Act.

How will we execute?

- Engage the Federal Agencies and IOOS Regions that are part of or contribute to US IOOS that will use the established QA/QC procedures;
- Develop a matrix of the full spectrum of procedures needed (IOOS core variables and sensors);
- Determine procedures that can be immediately adopted that are already being used by operational entities e.g. CO-OPS, NDBC, USACE (CDIP), US Environmental Protection Agency;
- Assess previous QARTOD work and move them to completed procedures as appropriate;
- Prioritize new QA/QC procedures needed based on the matrix developed above;
- Embark on developing new QA/QC procedures:
    - Review of existing practices across U.S. IOOS community partners and international efforts;
    - Convene a set of subject matter experts to develop the QA/QC procedures;
    - Write the QA/QC procedures manual; and,
    - Publish the QA/QC procedures, via a technical memorandum signed by the US IOOS Program Director.

What will be the results in the next 24 months?

- Initial assessment of existing documentation of QA/QC procedures relevant to the IOOS core variables. [FY12]
- Prioritization of the needs based on the assessment which will result in a work plan for QARTOD into future years. [FY12]
- Delivery of at least one manual providing explicit instruction for QC tests that are ready for data providers to implement. [FY13]

6. Implementing DMAC for biological observations in US IOOS regions

What are we doing?

- Expanding on the FY10-11 work that established common data content standards for IOOS biological core variables (i.e. fish species and fish abundance) and common access standards using ERDDAP (the Environmental Research Division's Data Access Program) service as the access and delivery services and Environmental Data Connector service (EDC) as the client to user application service.
- These biological data services have been implemented in PacIOOS and are easing access to and integration of biological data by consolidating data formats, and providing timely access to integrated data. The data providers involved are NOAA Fisheries/Pacific Islands Fisheries Science Center, National Park Services, Pacific Island Network and NOAA Papah*ā*naumoku*ā*kea Marine National Monument). The end users of this integrated data are population assessment scientists from NOAA National Marine Fisheries Service and Fisheries Regional Management Council.
- Implementing these biological data services in US IOOS regions including active regional engagement to access new sources of biological observations.

Why is this important?

- Integrating biological variables within US IOOS is a priority.
- Establishing broadly agreed upon standards for biological variables and integration with physical observing variables will greatly facilitate a much broader use of these data, for example:
    - to better understand species distribution and to learn how they might be shifting in response to climate change; and
    - to develop ecosystem indicators and targets critical for advancing integrated ecosystem assessment and ecosystem based management a priority for the National Ocean Policy.

How will we execute?

- Provide training via webinars to all US IOOS regions on the results and potential use and application of DMAC services for biological observations completed in FY11 in collaboration with NOAA Fisheries, OBIS USA, PacIOOS and other regional partners.
- Convene a series of meetings with partners in the GCOOS and SECOORA Regions to identify and develop requirements to implement biological data services that facilitate access/use of NOAA Fisheries' Southeast Area Monitoring and Assessment Program (SEAMAP) database. SEAMAP is a state/federal program designed to collect, manage and disseminate fishery-independent data in the SE Atlantic, Gulf of Mexico and Caribbean and has been in operation since 1983.
- These efforts will include continuing technical collaborations with USGS/OBIS (US Geological survey/Ocean Biogeography Information System) and NOAA Fisheries/SWFSC/ERD (South West Fisheries Science Center, Environmental Research Division) and possible new

partnerships through GCOOS and SECOORA with NOAA/NODC's National Coastal Data Development Center, NOAA. Fisheries /SEFSC (South East Fisheries Science Center and the Gulf States Marine Fisheries Commission and Atlantic States Marine Fisheries Commission which manages SEAMAP surveys data.

What will be the results in the next 24 months?

- Implementation of the DMAC biological data services in GCOOS and SECOORA to integrate biological observations [FY12]
- Implementation of the biological data services in remaining RAs [FY13]

7. Improving access to animal telemetry observations by establishing data standards

What are we doing?

- Conducting two projects to demonstrate how the application of IOOS-based data interoperability tools will lead to improved access to animal telemetry observations and reconcile animal telemetry data standards by:
    - Project 1: enabling ocean modelers access to historic physical oceanographic observations collected from sensors on marine animal (e.g. Seals and Sharks) placed by Stanford University's Tagging of Pacific Predators (TOPP) Program; an
    - Project 2: collaborating with the Pacific Ocean Shelf Tracking project (POST) in developing standards and access services for data from acoustic sensors placed in endangered and threatened fish species such as salmon and green sturgeon.

Project 1. TOPP data to ocean models

Why is this important?

- It will improve accuracy of ocean condition forecasts in poorly sampled ocean regions.

How will we execute?

- Jointly funded (FY11) by grants from U.S. IOOS and the U.S. Navy Office of Naval Research Marine Mammals and Biological Oceanography Program.
- Project team includes scientists from Stanford, U.S. IOOS, and Office of Naval Research (ONR) in collaboration with data "customers" at the U.S. Naval Oceanographic Office (NAVO) and NOAA National Weather Service's National Centers for Environmental Prediction (NCEP).
- Project objectives are to:
    - document customer requirements (data formats, access, etc.).
    - apply existing data access services to ease modeler assimilation of TOPP data. This will also include evaluating similar templates and formats developed for Australia's Integrated Marine Observing System (IMOS).

What will be the results in the next 24 months?

- Ingestion of TOPP historical physical oceanographic data into NAVO and NCEP ocean models [FY12]
- In addition, if resources allow, we will explore enabling access to TOPP real-time/near-real time observations [FY13]

Project 2. POST Project

Why is this important?

- Improved accessibility to these data will improve understanding and management of endangered and threatened species such as salmon, sturgeon and other mid-sized fish species as agreed upon at the US IOOS Animal Telemetry Observations Workshop (March 2011).
- Documenting standard access requirements for animal telemetry observations will facilitate a broader discussion on enabling improved access to and use of animal telemetry data across the national and international animal telemetry community; a significant advance for stewards of ocean observations

How will we execute?

- FY12 grant from the U.S. IOOS Program Office to NANOOS collaborating with NERACOOS, several NOAA Fisheries Science Centers, Australia's IMOS and the Ocean Tracking Network (OTN).
- Project objectives are to:
    - extend, adopt and implement data and metadata standards and data services for animal acoustic telemetry observations;
    - implement a standard approach for exposing these observations and information via web services; and
    - develop regional, accessible online products for browsing and visualizing such data, based on a customer-specified set of requirements.
- To ensure relevance and future extensibility, this project team will involve potential customers drawn from POST and NANOOS stakeholders (particularly fisheries state and federal agencies).
- Anticipated regional customers for these data include: NOAA Fisheries offices in West-coast, state fisheries agencies in Washington, Oregon, California and Alaska, several Native American tribes (including the Puyallup, Nisqually, and Squaxin), the Bonneville Power Administration (operator of the Columbia River dams), US Army Corps of Engineers, Seattle City Light (Power Company)
- Project team will seek broad community participation in technical discussions and priority-setting exercises related to the execution of project objectives. IOOS and NANOOS leadership will ensure adherence to IOOS DMAC recommendations and extensibility to other groups, datasets and regions.

What will be the results in the next 24 months?

- A data schema suitable to incorporate animal acoustic telemetry data and provide interoperability between animal telemetry data and physical observing data [FY12]
- Scalable web services to facilitate the data flow to customer applications [FY12]
- Implementation of animal acoustic telemetry data services in NANOOS servers [FY13]

In addition if resources allow, regional accessible online products for browsing and visualizing of tracking animal data [FY13]

8. Establishing baseline standards for profiling glider data

What are we doing?

- Establishing data formats, metadata and other requirements to enable improved and interoperable access to glider observation data
- Identifying priority issues/challenges associated with operating glider fleets and providing a collaborative process for their resolution.

Why is this important?

- The non-Federal partners of US IOOS currently operate over 80 gliders. (US IOOS partners deployed 7 of 9 gliders during the DWH oil spill).
- Standards for glider operations, particularly protocols for data description, format and associated services for access and visualization, will significantly magnify the value of these observations to the ocean science community.
- As with HF Radar, gliders offer a niche observation opportunity and US IOOS can lead the nation in the development of a National Federal and non-Federal glider fleet to improve national ocean observing coverage.

How are we executing?

- Establishing a small working group led by the US IOOS Program Office with representatives from MARACOOS, SCCOOS, NANOOS and CeNCOOS. Other federal agencies and role players will be brought in as appropriate.
- Forming a technical steering committee and organizing a community meeting to identify issues/challenges associated with integrating glider data, and operating gliders.
- Coordinating with NSF's OOI Cyberinfrastructure team, which has mutual goals for glider data structure development in FY2012
- Providing limited start-up funding to support these efforts.

What will be the results in the next 24 months?

- A dynamic, web-based National Glider Asset map on the US IOOS web site front page to highlight the broad use of this technology [FY12]
- An outline of a National Glider Plan, including data standards and formats issues, which can be promoted at the December AGU meeting [FY12]
- Completion of a National Plan [FY13]
- Expansion of the glider program in accordance with the National Plan [FY13]