The NOAA IOOS Data Integration Framework: Initial Implementation Report

Jeff de La Beaujardière
National Oceanic and Atmospheric Administration
Integrated Ocean Observing System (IOOS)
1100 Wayne Ave #1225
Silver Spring MD 20910 USA

Abstract- The US National Oceanic and Atmospheric Administration (NOAA) Integrated Ocean Observing System (IOOS) program office has begun the implementation of a Data Integration Framework (DIF) to improve management and delivery of an initial subset of ocean observations. The DIF establishes a web service layer atop key NOAA data providers, including the National Data Buoy Center (NDBC), the Center for Operational Oceanographic Products and Services (CO-OPS), and CoastWatch. Standardized service types and encodings for in situ and gridded data have been selected; this standardization allows a single client to access data from multiple sources with the same code. The DIF services will provide integrated access to data from both NOAA and regional partners. The standards and protocols used are broadly applicable, though specific decision-support tools and models relevant to harmful algal blooms, integrated ecosystem assessments, hurricane intensity, and coastal inundation have been targeted as initial customer focus areas for the DIF. The data access services are expected to be active shortly before the Oceans 2008 conference. This paper discusses the service layer, the data encoding specifications used, and the status of the implementation effort.

I. INTRODUCTION

The Integrated Ocean Observing System (IOOS) will enhance our ability to collect, deliver, and use oceanographic information. The goal is to provide sustained data on our open oceans, coastal waters, and Great Lakes in the formats, rates, and scales required by scientists, managers, businesses, governments, and the public to support research and to inform decision-making. IOOS is the oceans-and-coasts component of the US Integrated Earth Observation System (IEOS), the US contribution to the Global Ocean Observing System (GOOS), and the US contribution to the oceans-and-coasts component of the Global Earth Observation System of Systems (GEOSS). In 2007, the US National Oceanic and Atmospheric Administration (NOAA) established an office (http://ioos.noaa.gov/) to manage its contributions to IOOS. An interagency IOOS Data Management and Communications (DMAC) standards process had previously been established (http://ioosdmac.fedworx.org/).

The NOAA IOOS office has begun the implementation of a Data Integration Framework (DIF) to improve management and delivery of a subset of ocean observations (the first seven "core variables" being ocean currents, temperature, salinity, water level, waves, chlorophyll and surface winds). The DIF is intended to provide the initial operating capability for a nationwide IOOS DMAC capability, to enable the evaluation of interoperability specifications, and to demonstrate the feasibility and value of providing integrated ocean observations. In 2007, preparatory system engineering work resulted in the elaboration of DIF Functional Requirements [1] and Concept of Operations [2] documents. In 2008, establishment of this Data Integration Framework began in earnest with the implementation of a standardized, interoperable web service layer atop key NOAA data providers in order to provide integrated access to data from both NOAA and regional partners. We have used existing consensus or international standards where possible, and the standards and protocols used are meant to be broadly applicable. A working group on Web Services and Data Encoding (WSDE) has been established to guide these efforts. The WSDE working group comprises representatives from several NOAA offices and from the NOAA IOOS-funded projects that support regional observing capacity and national cross-cutting development. NOAA representation in the WSDE group includes National Ocean Service (NOS) Center for Operational Oceanographic Products and Services (CO-OPS), NOS Coastal Services Center (CSC), Office of Coast Survey (OCS) Coast Survey Development Laboratory (CSDL), National Marine Fisheries Service (NMFS) Southwest Fisheries Service Center (SWFSC) Environmental Research Division (ERD), NOS Integrated Ocean Observing System (IOOS) program office, National Environmental Satellite Data and Information Service (NESDIS) National Coastal Data Development Center (NCDDC), National Weather Service (NWS) National Data Buoy Center (NDBC), NESDIS National Geophysical Data Center (NGDC), NESDIS National Oceanographic Data Center (NODC), and Office of Oceanic and Atmospheric Research (OAR) Pacific Marine Environment Laboratory (PMEL). Non-NOAA representation includes Alaska Ocean Observing System (AOOS), Southern California Coastal Ocean Observing System (SCOOS), Gulf of Mexico Coastal Ocean Observing System (GCOOS), Northeastern Regional Association of Coastal Ocean Observing Systems (NERACOOS), Mid-Atlantic Regional Coastal Ocean Observation System (MARCOOS), Northwest Association of Networked Ocean Observing Systems (NANOOS), Southeast Coastal Ocean Observing Regional Association (SECOORA), and Southeastern Universities Research Association (SURA) Coastal Ocean Observing and Prediction (SCOOP) program.
In the following sections, we discuss the web services and encoding conventions used by the DIF, the specific implementations now underway at NOAA data providers, our work with customer applications to prepare for these data, and next steps we hope to undertake in 2009.

II. DATA ACCESS SERVICES AND ENCODING CONVENTIONS

No single web service type or data format will satisfy all users. The Data Integration Framework project has broadly identified three general classes of scientific information -- in situ data, gridded data, and images of data -- and has recommended a web service and encoding convention to be used in each case. These recommendations are intended to standardize a small number of data access methods and thereby to enable a single client application to obtain data from multiple providers, and to harmonize the representation of data from different providers. These services can be established either instead of or in addition to prior arrangements between individual providers and customers. The initial DIF services and encodings are summarized in Figure 1 and described in more detail below.

A. In situ data

For in situ observations such as those from buoys, piers, bottom-mounted sensors and volunteer observing ships, the DIF uses the Open Geospatial Consortium (OGC)1 Sensor Observation Service (SOS) [3] serving data encoded in Extensible Markup Language (XML) [4]. SOS defines a set of operations for software to request data or service metadata using Hypertext Transfer Protocol (HTTP) [5]. DIF data providers are implementing the SOS "core operations profile," which comprises three basic functions:

1. GetCapabilities allows users to obtain service metadata including general information about the data holdings available from a particular server. (GetCapabilities is an operation defined for all OGC web services.)
2. GetObservation allows users to retrieve data from the desired sensor(s) and time period.
3. DescribeSensor provides detailed metadata about a sensor, typically encoded in Sensor Model Language (SensorML) [6].

OGC also defines a Web Feature Service (WFS) [7] that could have been chosen instead of SOS. The two services are qualitatively similar, but WFS is general-purpose whereas SOS is explicitly specialized for use with sensor observations. SOS can be used for both in situ and remote sensors, but the remote-sensing application of SOS seems less well-developed in practice and such use has not yet been attempted as part of the DIF. The SOS specification has been submitted by the NOAA IOOS Program for consideration by the DMAC standards process.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Web Service</th>
<th>Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-situ data (buoys, piers, towed sensors)</td>
<td>OGC Sensor Observation Service (SOS)</td>
<td>XML based on OGC Observations and Measurements (O&amp;M)</td>
</tr>
<tr>
<td>Gridded data (model outputs, satellite)</td>
<td>OpenDAP and/or OGC Web Coverage Service (WCS)</td>
<td>NetCDF using Climate and Forecast (CF) conventions</td>
</tr>
<tr>
<td>Images of data</td>
<td>OGC Web Map Service (WMS)</td>
<td>GeoTIFF, PNG etc. -possibly with standardized styles</td>
</tr>
</tbody>
</table>

Figure 1 - NOAA IOOS Data Integration Framework (DIF) service types and encodings.

---

1 OGC is a non-profit, international, voluntary consensus organization that develops standards for geospatial and location-based services (http://www.opengeospatial.org); NOAA is a Principal Member.
To standardize data provided by the DIF’s Sensor Observation Services, the WSDE working group developed a draft specification for encoding *in situ* ocean observations using XML based on OGC Geography Markup Language (GML) [8] and Observations and Measurements (O&M) [9] standards. GML is an XML grammar for the transport and storage of geographic information. GML is an OGC specification and an international standard (ISO 19136)[10]. GML is general-purpose, and can be used to express features such as roads or parcel boundaries as well as observations. Consequently, GML is often specialized for a particular information community using an "Application Schema." O&M defines an application schema for expressing an observation (the act of observing a phenomenon) and measurements (numeric values that result from an observation). SOS and O&M are part of the OGC Sensor Web Enablement (SWE) suite of specifications. Figure 2 illustrates the specialization of XML for *in situ* measurements.

The DIF XML version 0.6 specification includes schema and data record definitions for six IOOS core variables (currents, temperature, salinity, water level, winds and waves) and a variety of sampling feature types (points, profiles, and trajectories) and collections or time series thereof. It comprises a GML application schema that extends and specializes GML and SWE schema definitions, a profile of the O&M schema, a collection of O&M observation XML documents, and an associated set of SWE XML record definitions. The resulting encoding conventions are rich enough to capture the breadth of observational data and sensor metadata that is available from NOAA DIF data providers. The XML is structured enough to be transformed by Extensible Stylesheet Language Transformations (XSLT) [11] into other useful representations including Google Earth's Keyhole Markup Language (KML) [12] (recently approved as an OGC standard), Hypertext Markup Language (HTML) [13] for web pages, and comma-separated value (CSV) text for spreadsheets. The preliminary DIF XML schema is available from the NOAA Coastal Services Center (CSC) schema repository at http://www.csc.noaa.gov/ioos/schema/IOOS-DIF/. This XML encoding specification is now being implemented and tested as described in Section III.

### B. Gridded data

For serving gridded observations (including ocean color from satellites, surface currents from high-frequency radar, and model outputs), the DIF recommends either the OGC Web Coverage Service (WCS) [14] or the Open Project for a Networked Data Access Protocol (OPeNDAP) [15]. Both protocols are suitable for accessing regular grids; OPeNDAP also supports irregular grids. WCS is explicitly called out in the GEOSS architecture and is supported by some commercial off-the-shelf (COTS) Geographic Information System (GIS) tools. OPeNDAP Data Access Protocol is under review as a recommended IOOS DMAC data transport mechanism and is well used in the scientific community. WCS has been submitted for consideration by the DMAC standards process.

WCS defines three operations for requesting data or metadata using HTTP:

1. **GetCapabilities** allows users to get service metadata including general information about the data holdings available from a particular server.
2. GetCoverage allows retrieval of coverages or subsets of coverages in the spatial or temporal domain (a "coverage" being defined as "digital geospatial information representing space-varying phenomena")[14].

3. DescribeCoverage allows a client to request full descriptions of one or more coverages served by a particular WCS server. The server responds with an XML document that fully describes the identified coverages, including the domain and range of the coverage function, supported coordinate reference systems and encoding formats, and additional metadata about the coverage.

The OPeNDAP protocol includes an intermediate data representation used to transport data from the remote source to the client, a procedure for retrieving data from remote servers, and an API consisting of OPeNDAP classes and data access calls designed to implement the protocol [15].

The DIF recommends that gridded data be encoded in Network Common Data Form (NetCDF) [16] with Climate and Forecast (CF) conventions [17]. We will document any conventions beyond CF that may be desirable for data served by the DIF.

C. Images of Data

For images of data, the DIF recommends the OGC Web Map Service (WMS) [18], which can serve maps in graphic formats such as Georeferenced Tagged Image File Format (GeoTIFF) [19]. WMS is an OGC specification and an international standard (ISO 19128) [20]. WMS has been submitted for consideration by the DMAC standards process. WMS is intended to generate visualizations upon request to the user's specifications, but can also serve static pre-generated images. WMS defines two mandatory operations:

1. GetCapabilities allows users to get service metadata including general information about the data holdings available from a particular server.
2. GetMap allows users to request an image of data of the desired size and format for a specific georeferenced bounding box and time period. By issuing GetMap requests of commensurate size and bounding box, users can overlay data from different servers and produce a composite, visually-integrated view of data.

III. DATA PROVIDER IMPLEMENTATIONS

In mid-2008, implementations of the DIF web service layer were initiated with support from IOOS at three NOAA data providers: The National Weather Service (NWS) National Data Buoy Center (NDBC), the National Ocean Service (NOS) Center for Operational Oceanographic Products and Services (CO-OPS), and the National Environmental Satellite Data and Information Service (NESDIS) CoastWatch program. These centers provide in situ or remotely-sensed data including ocean currents, temperature, salinity, water level, waves, winds and ocean color-derived chlorophyll. Specifically, as part of the DIF, NDBC will be establishing an SOS for in situ data, a WCS for gridded surface current observations from high-frequency radar (HFR) installations, and a WMS to provide images of these data. CO-OPS will be establishing an SOS for in situ data. CoastWatch will be establishing WCS and OPeNDAP services providing gridded chlorophyll concentration derived from satellite ocean color observations. Table 1 shows the breadth of feature types and variables offered by these three providers; time series of all variables will also be available.

<table>
<thead>
<tr>
<th>Table 1: Variables and feature types to be offered by DIF data providers.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currents</td>
</tr>
<tr>
<td>Point (single station)</td>
</tr>
<tr>
<td>Profile (vertical or horizontal)</td>
</tr>
<tr>
<td>Collection (group of stations)</td>
</tr>
<tr>
<td>2D grid</td>
</tr>
</tbody>
</table>

The SOS implementation at NDBC is of particular interest. For the first time, a single service layer will provide access to national and regional data from the four Data Assembly Centers (DACs) at NDBC: the NDBC DAC (data from NDBC-operated stations), the IOOS DAC (data from stations operated by regional coastal ocean observing systems (RCOOS) and transmitted to NDBC), the Deep-ocean Assessment and Reporting of Tsunamis (DART) DAC, and the Tropical Atmosphere Ocean (TAO) DAC. CO-OPS, meanwhile, will provide integrated access to data from its NWLON and PORTS stations. Figure 3 presents a conceptual illustration of this initial implementation. The use of a standardized service and output format allows for a single client application to access data from multiple IOOS DIF SOS sources with the same code.
The IOOS DAC at NDBC includes a subset of the observations gathered by the RCOOS. In order to make all those observations available and interoperable, the NOAA IOOS office is encouraging its regional partners to implement SOS and to offer in situ data encoded according to the DIF XML conventions.

As of this writing, the NDBC and CO-OPS SOS implementations are not yet complete. The SOS GetObservation operation is being established for each variable. GetCapabilities is partially implemented. DescribeSensor has not yet been implemented, pending elaboration of SensorML descriptions for the various sensors. If completion occurs on schedule, these services will be ready by the time of the Oceans 2008 conference.

In addition to the SOS at NDBC and CO-OPS, another SOS implementation is pending. The Observing System Monitoring Center (OSMC) [21] software developed at NOAA Pacific Marine Environment Laboratory (PMEL) in support of NDBC and the NOAA Office of Climate Observations (OCO) will be enhanced to provide an SOS interface to data that OSMC caches from the World Meteorological Organization (WMO) Global Telecommunications System (GTS). OSMC will offer in situ data encoded according to the DIF XML schema, and will also experiment with encodings based on Climate Science Modelling Language (CSML) [22].

IV. CUSTOMERS

The NOAA IOOS Program selected several application areas of particular interest in order to serve as initial customers for interoperable data available through the DIF:

1. Harmful Algal Blooms
2. Coastal Inundation
3. Hurricane Intensification
4. Integrated Ecosystem Assessments

Preparatory work is ongoing with these customers' Decision Support Tools (DSTs). However, at the time of this writing the data access services are still being established, so none of these customers are yet ingesting data via the DIF. The IOOS Program and WSDE working group are also working closely with CSC to develop client modules to access and use of data from DIF web services (e.g., ingest by COTS GIS applications). Once completed, the web services being established by the NOAA IOOS Program Office’s DIF Project will serve not only the four customer applications above but the broad IOOS user community by facilitating access to and use of these important observations.

V. NEXT STEPS

By September 2008, the initial SOS implementations at NDBC and CO-OPS should be complete. The CoastWatch chlorophyll WCS and the NDBC WMS and WCS should be complete by October 2008. However, these are only the first steps, and additional work is expected in FY 2009 and beyond. The following is a sampling of possible next steps; the actual work performed will, of course, depend on project requirements and available resources.
**Testing, evaluation and refinement of encoding specifications:** The NOAA IOOS Program Office and CSC will perform interoperability testing to confirm that the NDBC and CO-OPS SOS implementations are compatible. The goal is that client software applications should be able to request a particular data type from both servers and ingest the data with no server-specific coding. Beyond this testing, performance and usability will be assessed to evaluate these encoding specifications. Also, we would like to harmonize with ongoing OGC work to evolve the SWE schema and with the evolving Climate Science Modelling Language (CSML) [22], which is derived in part from OGC O&M. The DIF XML encoding specification will be submitted for consideration by the DMAC standards process.

**Registry implementation:** A Registry should be established to provide a catalog of the available services and data holdings. The OGC Catalog Services for Web (CS/W) [23] profile seems promising. We hope to be able to make use of the GEOSS registry rather than building our own.

**Metadata:** Good data documentation is critical to finding data, and especially to being able to use data once found. We have already begun an assessment of metadata implementations at the NOAA data providers, and will assess how to map available metadata to existing metadata standards and assist providers in improving their metadata.

**Quality assessment:** The data served by NDBC, CO-OPS and CoastWatch undergo a variety of quality-control measures. We will continue our efforts with the Quality Assurance of Real-Time Ocean Data (QARTOD) to OGC (Q20) project to determine how best to represent QA/QC information in SensorML descriptions.

**Data integration service:** The DIF SOS implementations at NDBC and CO-OPS will provide integrated access to their respective holdings—e.g., data from four NDBC DACs can be obtained with a single SOS GetObservation request. Integration of data records from multiple independent SOS can occur at the visual level simply by showing the data records. As yet, however, no service exists that can aggregate data records from multiple independent SOS into a single record while removing duplicates and ensuring consistency. The standardized access services and encodings used by the DIF will enable that level of aggregation, and we would like to establish such an integration service.

**Data translation service:** The NOAA National Marine Fisheries Service (NMFS) Southwest Fisheries Service Center (SWFSC) Environmental Research Division (ERD) has developed a data translation and visualization service known as ERDDAP. ERDDAP is able to access data in a variety of formats and protocols, and to transform those data on-the-fly to other formats or representations desired by the user. We wish to have ERDDAP able to read and translate from the SOS implementations that use the DIF XML encoding specification. We also hope to work with one of the projects funded by the National Science Foundation (NSF) Ocean Observatories Initiative (OOI) Cyberinfrastructure effort to establish a distributed, reliable, fault-tolerant translation service to test this approach to providing seamless access to data.

**Customer development:** We will work with the already-identified customers to assist them in accessing data through the servers established by the DIF. We would also like to identify other possible customer areas such as climate change and marine transportation.

**Additional variables:** We plan to add additional physical variables and perhaps biological variables to broaden the applicability of the DIF. Also, we would like to add other sources of similar variables (e.g., satellite sea surface temperature and sea surface height to complement in situ temperature and water level).

**Additional providers:** Within NOAA, we hope to work with other data providers and archive data centers to provide access to more data. As stated above, we will continue to work with regional partners to promote service and data interoperability. Finally, we hope to collaborate with other US federal agencies that offer marine or estuarine data to promote interoperability with those sources.
The NOAA IOOS program has initiated the development of a Data Integration Framework to provide interoperable access to several sources of oceanographic data. The present scope of the effort is modest, focusing on a small number of variables and providers, but the methodology is generally applicable to a wide variety of observations, providers and customers. Figure 4 indicates the scope of the DIF services we hope to establish as part of this project. Our intent is to evolve these approaches to maximize interoperability among all ocean observing groups and data providers, and we welcome participation in this effort by other data providers.

ACKNOWLEDGMENTS

A project of this magnitude cannot be accomplished without many participants. The in situ XML schema was developed by Paul Daisey, one of the authors of the GML specification. Implementation has been managed and performed at NDBC by (in alphabetical order) Bill Burnett, Darrell Duncan, Mike Garcia, Bill Smith and Chris Taylor, at CO-OPS by Andrea Hardy, Mohamed Chaouchi, and Armin Pruessner, and at CoastWatch by Paul DiGiacomo, Dave Foley, Phil Keegstra, and Avichal Mehra. Interoperability testing will be performed by Daniel Martin and John Ulmer. The OSMC and ERDDAP work will be led by Steve Hankin and Roy Mendelssohn, respectively. Metadata work has been performed by Julie Bosch and Ted Habermann. Jack Harlan leads the HFR surface currents work. WSDE working group members not already mentioned include Luis Bermudez, Rick Blair, Julie Bosch, Jen Bowen, Tess Brandon, Eric Bridger, Grant Cameron, John Cartwright, Rob Cermak, Jesse Cleary, Jeremy Cothran, Janet Fredericks, Charlton Galvarino, Felimon Gayanilo, Matt Howard, Bill Howe, Eoin Howlett, Stuart McClean, Rich Patchen, Paul Reuter, Craig Risien, Stephen Uczekaj, and Darren Wright. On the customer application side, I appreciate the current and especially future efforts by Stephen Baig, Rob Bassett, Charles Carleton, Lynn deWitt, Lori Fernstemacher, Kathleen Fisher, Gustavo Goni, Nicole Kurkoski, Rebecca Love, Sharon Mesick, Eric Roby, Frank Schwing, Bob Simons, Peter Stone, Rick Stumpf, Arthur Taylor, Michelle Tomlinson, Mark Vincent, and Yanlin Ye. Finally, in the NOAA IOOS program office, I thank Charles Alexander, CAPT Christopher Moore and Zdenka Willis for their leadership, and Alexander Birger, Jack Oliva, Carmel Ortiz, Rob Ragsdale, Rebecca Shuford and Marcia Weak for all their work to establish the Data Integration Framework.
REFERENCES