NOAA

Integrated Ocean Observing System (IOOS) Program

Data Integration Framework (DIF)

Design Document

(Version 1.0)

December 4, 2008

Revisions

Version	Description of Version	Date Completed
Draft 0.4	Initial draft released to key DIF design stakeholders for comment	June 17, 2008
Draft 0.5	Added significant material including customer project summaries, key risks, requirements traceability and registry.	August 29, 2008
1.0	Document revised significantly in accordance with the DIF stakeholders comments, errors fixed.	November 2, 2008

Review & Approval

Review History

Reviewer	Version Reviewed	Signature	Date
Charles Alexander/IOOS			

Table of Contents

LIST OF FIGURES

1. DOCUMENT OVERVIEW, SCOPE AND REFERENCES

1.1. Document Organization

The Design Document is organized into the following chapters and appendices:

- **Chapter 1:** Document Overview, Scope, and References – provides an outline for the document.
- **Chapter 2:** Introduction to the DIF — briefly describes the IOOS DIF initiative and scope, and provides a background for the design. This section provides an orientation to the DIF for those not familiar with it.
- **Chapter 3:** DIF Design Concept – introduces the concept and principles that lay the foundation of the design.

Chapters 4-6 describe the DIF design in terms of the reference models identified in the Federal Enterprise Architecture Geospatial Profile [RD1], where applicable:

- **Chapter 4:** Performance Reference Model **–** describes the anticipated benefits of the DIF and provides a structure for analyzing inputs and outcomes.
- **Chapter 5:** Business Reference Model **–** describes the approaches for incorporating DIF data, services and technology into DIF customers' existing business processes.
- **Chapter 6:** Data Reference Model **–** describes the conceptual design, documentation and data sharing standards and protocols to be implemented by DIF data providers to facilitate seamless, interoperable access to data by DIF customers.

1.2. Document Purpose, Scope, and Intended Audience

This Design Document is not intended to be a detailed implementation specification. The goal of the DIF Design Document is to outline a *high level* design scope that will guide subsequent detailed design and implementation efforts. Specifically, the objectives of this document are to:

a. Identify a high level architecture for the DIF; this architecture will be used to guide detailed design efforts on specific implementation projects.
b. Leverage and validate DIF functional architecture contained in the CONOPS (Concept of Operations).
c. Describe the candidate DIF standards and conventions for data content, encoding, metadata, QC (Quality Control), and transport.
d. Define the services to be provided at each data provider location or proxy.
e. Describe how DIF components should be implemented to be consistent with applicable Department of Commerce and NOAA IT security policies and procedures.
f. Identify an approach to testing/verifying services established by data providers.
g. Summarize key risks and mitigation strategies.
h. Provide requirements traceability to the DIF Functional Requirements document.

1.3. Reference Documents

RD1. ***Federal Enterprise Architecture Geospatial Profile v 1.1*** by Architecture and Infrastructure Committee, Federal CIO Council and Federal Geographic Data Committee, January 27, 2006

RD2. ***DIF As-Is Baseline Systems Document (DRAFT Version 0.6)***, September 19, 2007

RD3. ***DIF Functional Requirements Document, Version 1.2***, October 21, 2008

RD4. ***DIF Concept of Operations Document, Version 1.0***, April 25, 2008

RD5. ***The Data Reference Model v 2.0*** by Federal Enterprise Architecture Program, November 17, 2005

RD6. ***NOAA Enterprise Architecture. Volume I: Core Architecture Descriptions v 3.0***, February 2008

RD7. ***DIF Customer Implementation Project Summary and Performance Assessment Plan, Version 1.0,*** October 16, 2008

RD8. ***CSC DTL DIF Service Testing Regimen***, Version 20080620

2. INTRODUCTION TO THE DIF

2.1. DIF Background and Context

The NOAA Integrated Ocean Observing System (IOOS) Program Office has been charged with developing a Data Integration Framework (DIF) for NOAA's Integrated Ocean Observing System.

The DIF was proposed because there are no widely accepted approaches to facilitating integration and interoperation of data from diverse, distributed sources, and meeting the geographic coverage, vertical and horizontal resolution, measurement accuracy, and timeliness requirements of the many NOAA decision-support systems. The DIF will identify useful standards and provide additional specificity as needed. It will also address gaps in data management services that serve selected NOAA ocean models, assessments, and decision-support tools such that the anticipated operational improvements and/or increases in efficiencies with respect to time and costs can be achieved and measured. The DIF will leverage appropriate data management capabilities across NOAA, other federal government agencies and regional partners.

2.2. DIF Objectives and Scope

The DIF premise is that data integration and improved access to ocean-related data will increase the data's value and effectiveness in supporting decisions related to making tools/models, and will provide better practices and other protocols of use to other IOOS partners. The DIF will not define how data providers' data holdings should be managed, but rather will define interfaces and specifications for how data should be shared.

Specifically, the objectives of the DIF are to:

1. Validate the premise that integrated data and improved access to the data has value that can be measured.
2. Develop a methodology to improve upon existing ocean data integration efforts by utilizing the principles of the IOOS DMAC. This will facilitate flexibility and extensibility to other variables, systems and decision support tools.
3. Achieve improved integration of and access to selected data sets by identifying, adopting, and adapting corresponding standards for data content, metadata, quality control, and transport and deploying these standards at selected data sources serving the selected decision support tools.
4. Maintain the DIF for a minimum period of three years, from project inception, to allow for adequate performance monitoring and assessment.
5. Provide a set of lessons learned, draft standards, implementation guidance, and other outputs that will allow the longer-term strategic ocean data integration efforts to leverage the DIF experience to the benefit of NOAA and the Nation.

The Data Integration Framework described in this document provides a scalable, flexible and extensible design to address the core DIF objectives in an initial implementation; this design can be further expanded beyond the initial set of data variables, providers, and customers as needed.

The initial four decision-support areas were selected because they address critical environmental issues aligned with NOAA mission goals: **Coastal Inundation**, **Hurricane Intensity**, **Integrated Ecosystem Assessments**, and **Harmful Algal Blooms**. The requirements of these communities, operating within NOAA, were used to guide the design and development of the DIF. The value and success of the resulting integration will be measured and evaluated by its ability to enhance the efficiency and/ or effectiveness of decision support tools for these communities. The variables were selected based on their anticipated relevance to the chosen decision support tools as well as the number of readily available data sources: these core IOOS variables are **sea water temperature**, **salinity**, **currents**, **ocean color**, **sea/water level, wind and waves**.

Integration within the DIF means improving the way the selected sources of the seven variables are made available to the four decision-support tools through the consistent application of community-based standards and protocols, such as for data content, metadata, encoding and transport. By adopting, adapting, or expanding existing standards and capabilities for data management services, the DIF will formalize a standards-based common data sharing infrastructure that is expected to facilitate and improve data integration of ocean variables across NOAA.

Initial NOAA sources of the seven core variables have been selected based on the requirements of the decision-support tools. These sources are NWS' National Data Buoy Center (NDBC), NOS' Center for Operational Oceanographic Products and Services (CO-OPS) and NESDIS' CoastWatch. Additional variables and sources may be included, if feasible, given timelines and budget constraints.

Again, while the initial DIF implementation supports seven variables, three data providers, and four decision support areas, the design is scalable, flexible and extensible and can be further expanded beyond the initial implementation, as needed.

3. DIF DESIGN CONCEPTS

This section provides a high level overview of the core DIF architecture principals and describes intersections with the NOAA Enterprise Architecture (EA).

3.1. Architectural Principles

The DIF design is based on several core architectural principles:

1. **Service-Oriented Architecture (SOA) –** in regard to the DIF as a whole, SOA is used in a broad sense as an approach to systems development and integration where functionality is grouped around business processes and packaged as interoperable services. Following that approach will ensure DIF consistency with the NOAA EA during design and implementation time.

   The SOA approach is also applicable in a more restricted sense as the means of data sharing; the DIF Data Model includes a limited number of interoperable services which will allow DIF Data Customers to retrieve and aggregate data from multiple DIF data providers without knowledge of the data providers underlying technology infrastructure. Detailed information on the services to be employed is contained in Section 6.3.2.

2. **Distributed Architecture –** the DIF is conceived as a framework with components implemented at data providers' facilities, software gateways, customer locations, or all of these. Core functionality of the DIF related to data transformation, assembly, encoding and transport will be distributed among several data providers rather than at a central data management facility (reference Section 5.3).

3. **Management Functions –** although the DIF is conceived as a distributed system, there is also a need for an overarching set of functions, such as a central data registry, central metadata management and system monitoring functions. These functions will likely be co-located with one or more data provider facilities (reference Section **Error! Reference source not found.**).

4. **Standards-based components –** the DIF achieves improved integration of and access to selected data sets by identifying, adopting, and adapting standards for data content, encoding, metadata, quality control, and transport. Detailed information on standards to be employed by the DIF is contained in Section 6.2.

5. **Data Pull Service –** the DIF is currently designed to support delivery of data through "pull" services whereby data from one or more data providers is transported to the customers as the result of a request for the data [RD4].

6. **Data Push Service –** the type of data delivery service where transfer of data is controlled by the Data Provider rather than Data Customer; is used in some real-time data distribution systems like WMO Global Telecommunications System (GTS). Although this type of service has been included into *DIF Functional Requirements Document* [RD3], there is no plan to develop the service under the initial DIF implementation.

Figure 3.1 provides a representation of the DIF architecture.

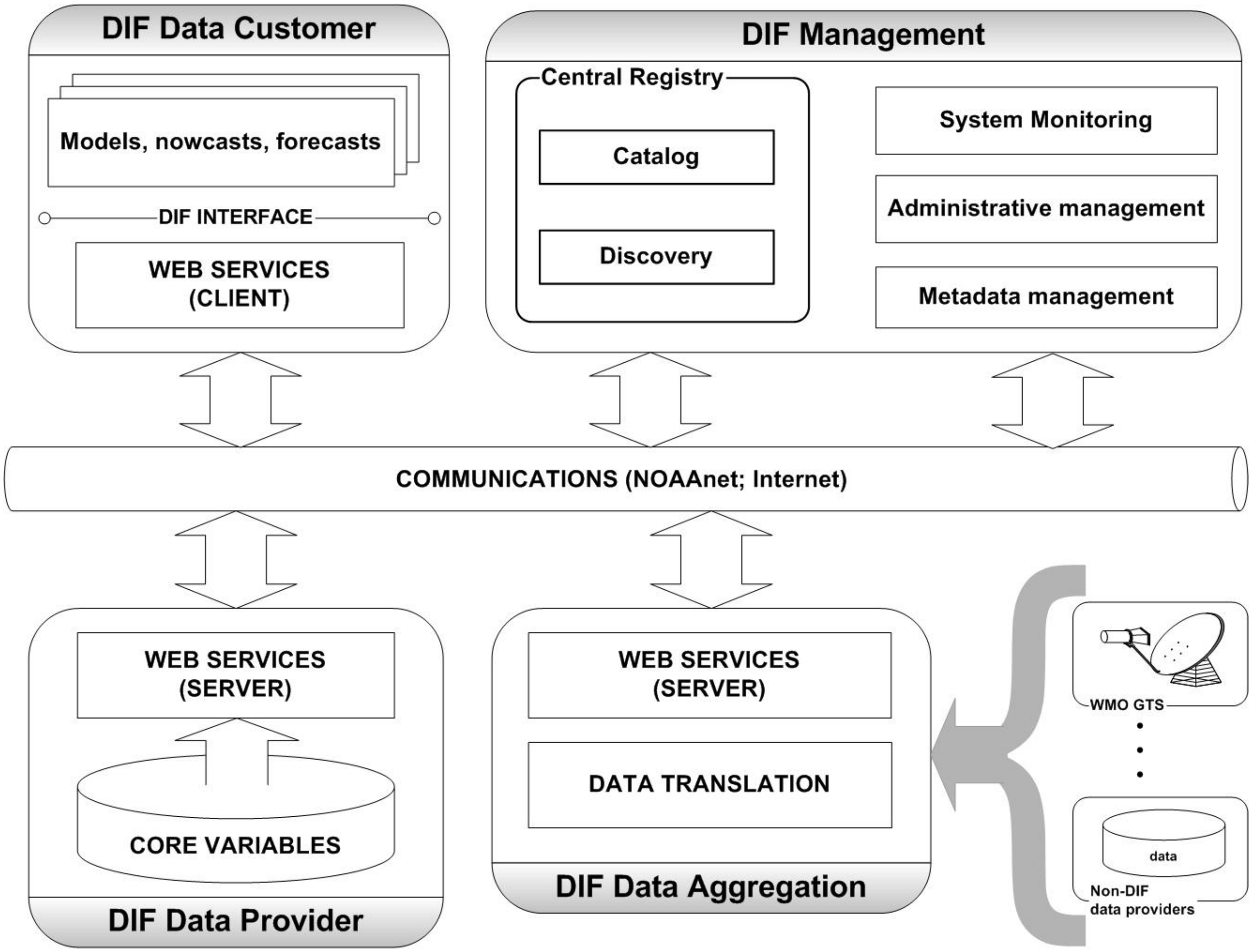

Figure 3.1: DIF Architecture

3.2. Major Components

Figure 3.1 highlights the major components of the DIF: Data Provider, Aggregation, Management, and Data Consumer, as originally described in the DIF Concept of Operations document (RD4).

A key DIF objective is to overlay services on the existing Data Providers' and Data Customers' infrastructure rather than rebuilding it. The distributed Service-Oriented Architecture model satisfies the objective, and provides a balanced and practical method for managing deployment of services in multiple Data Providers' and Data Customers' facilities.

In that architecture model each Data Customer has a DIF client component that initiates requests for data, and each Data Provider has a server component that accepts the request, processes it, and provides the data. One Data Customer may simultaneously address several Data Providers. Due to the nature of the Web Services, a Provider may also play a Customer role for another Provider if needed (see [RD4] for relevant use cases).

The DIF project intends to improve access to high-quality, integrated data by adding standardized access services, and standardized data encoding to the existing infrastructure of the data providers. As a result of that DIF customers will be able to access data from multiple sources, and seamlessly ingest it into their models and decision-support tools.

The general DIF design approach is to adopt community-recognized standards and specifications, and specialize, extend or revise them if needed for IOOS data. The standards and specifications approved by recognized standards bodies such as International Standard Organization (ISO), Open Geospatial Consortium (OGC), Unidata, etc. should be used whenever possible; modifications to these standards and specifications for adaptation to IOOS should be made only when necessary.

A Registry services will be implemented as an essential component of the distributed architecture since multiple Data Providers may carry portions of requested data. The Registry collects and manages metadata needed to discover and track the data available from various providers. Although DIF will deploy a Central Registry services serving all customers and providers, certain providers may run their own Local Registry services.

Data Aggregation services are currently anticipated in the overall design and would be deployed along with the Registry services to minimize the Data Customer’s redundant requests for data. The Data Aggregation services not only collect and convert data from various Data Providers in response to the Data Customer’s specific request, and combine it into one package for delivery. The Data Aggregation services also make provision for data spatial and temporal sub-setting, and data format conversion when needed. Certain Data Provider can also routinely use these services to aggregate data from other sub-providers along with its own data regardless of Data Customers’ requests.

An open architecture approach employed by the DIF allows free expansion in any direction, which will be considered valuable upon completion of the initial DIF implementation: more Data Customers can be served by adding more Data Providers and/or more Variables.

3.3. NOAA Enterprise Architecture Conformity

DIF design has advisedly been kept in line with the NOAA Enterprise Architecture (EA) approach; it is expected to work toward the EA goal to bridge the divide between business requirements and technical solutions, which is illustrated by Figure 3.2.

As a result, DIF’s distinctive features clearly demonstrate its conformity with the NOAA EA’s principles and requirements:

1. DIF Architecture is service-based and focused on meeting the needs of specific customers’ business priorities as well as on interoperability, scalability, and security of both NOAA line offices, and non-NOAA interested parties.
2. DIF is being developing in close collaboration with NOAA business units and the CIO office, which helps to align the DIF technical solutions with NOAA business objectives.
3. DIF establishes a reference architecture and implementation cases that can be used by project implementation teams to jump start their efforts.
4. DIF thoroughly assesses all solutions in terms of potential leverage toward DIF Target Architecture objectives, and factors this analysis into decisions for how to provision each service.

5. DIF develops and funds specific projects that are in accord with the target SOA and focused on fulfilling key business objectives.

Figure 3.2 illustrates the intersects between the NOAA Enterprise Architecture and the DIF.

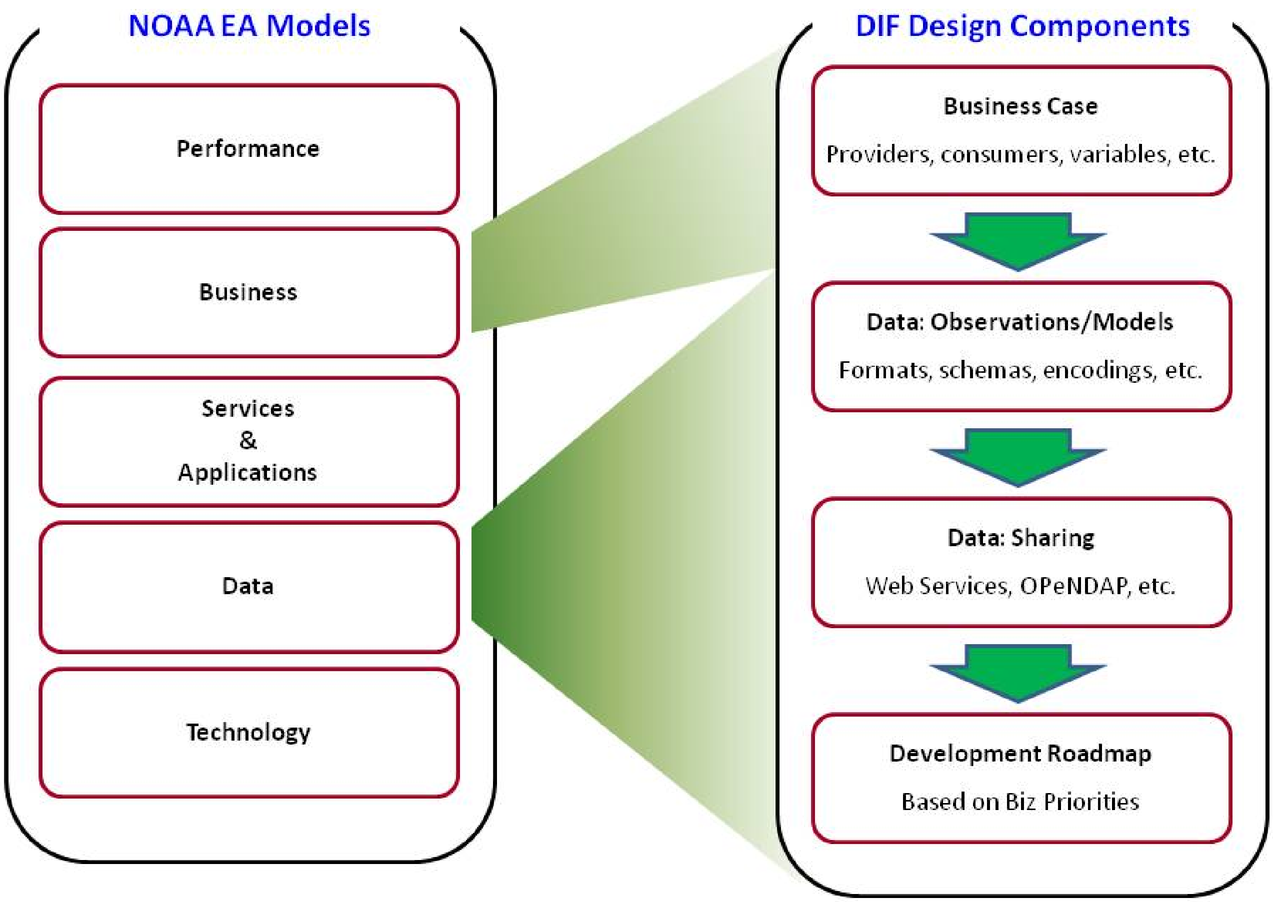

Figure 3.2: DIF conformity with NOAA EA

4. PERFORMANCE REFERENCE MODEL

This section describes the anticipated benefits of the DIF and provides a structure for analyzing inputs and outcomes.

4.1. Performance Reference Model Overview[1]

Assessment of DIF results will be performed in accord with the Performance Reference Model (PRM) approach suggested by the Federal Enterprise Architecture (FEA) recommendations. The PRM is a normative model; it sets targets for action and measures the value of improvement.

The PRM provides a framework for performance benchmarking and metrics ensuring that a common approach is applied to all DIF projects. It will allow NOAA IOOS management to better manage the DIF at a strategic level by providing a means to measure the success of investments and their impact on strategic outcomes. The PRM accomplishes these goals by establishing a common language for DIF projects to describe the outputs and measures used to achieve DIF objectives. The model articulates the link between internal business components and the achievement of business and customer-centric outputs. Most importantly, it facilitates resource-allocation decisions based on comparative determinations of relative efficiency and effectiveness of projects.

The PRM measures the performance of an activity, service, or investment by applying specific sets of indicators to relevant areas. There are four general areas for performance measurement as indicated in Figure 4.1: technology, process/activity, customer service, and business results. For each area the PRM provides a full set of measurement categories; in practice just relevant for a specific project categories are used.

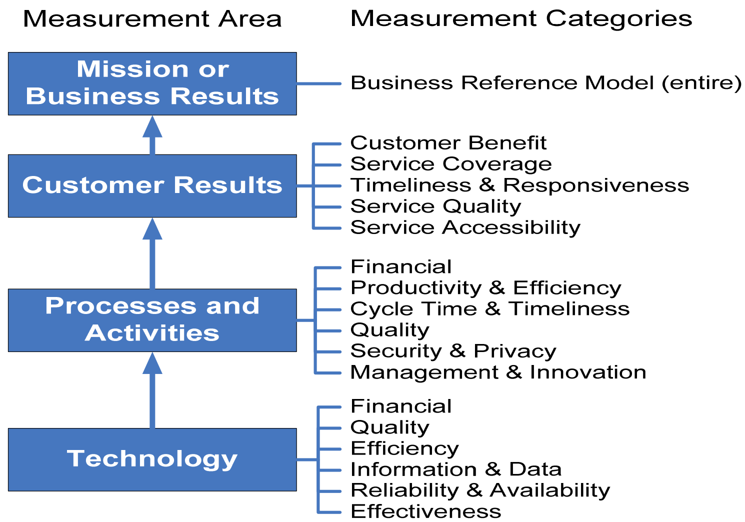

Figure 4.1: PRM Measurement Categories and Measurement Areas
(from *Federal Enterprise Architecture Geospatial Profile v 1.1*)

[1] Based on the PRM overview in the ***Federal Enterprise Architecture Geospatial Profile v 1.1*** by Architecture and Infrastructure Committee, Federal CIO Council and Federal Geographic Data Committee, January 27, 2006

4.2. Performance Assessment

The purpose of DIF performance assessment is to measure the impact of integrated data and standardized access methods on the decision-support tools. The metrics and methodology to be used for performance assessment will be specific to the customer project. While objective, quantifiable assessments are the goal, limitations in available assessment tools or resources needed to perform assessments may limit precise quantification of benefits. Detailed benchmarking and performance assessment plans are detailed in a separate document (RD7).

Based on preliminary definition of projects for the four customer areas, expected benefits of DIF integrated data includes:

- Reliability – integrated data available in standard formats and access methods provides an increased degree of data availability for decision-support tools/models
- Skill – an increase in the quality or accuracy of the output of the decision-support tools/models
- Efficiency – a decrease in the amount of time required to perform an analysis or to collect and prepare data for input into the decision-support tools/models

The FEA PRM suggests distinguishing mission results from business results, since it is assumed that mission results always benefit the general public, whereas business results reflect the result of the agency internal activity. However, in some cases the DIF mission and business results can be considered the same because improvement in business results can lead to benefits for the general public. Therefore the performance assessment will be based either on a mission results or business results, or both.

4.3. Testing and Validation

The DIF will employ testing and validation to verify that data provider and customer implementations function as expected. This testing and validation does not attempt to assess the benefit or impact of integrated data (see Section 4.2 - Performance Assessment), but rather to verify the content, format, and protocols as described below.

4.3.1. Data Provider Compliance Testing

The purpose of compliance testing is to ensure that the services, encodings, other tools, and content standards implemented at each DIF data provider are in accordance with all standards and schemas and that all agreed upon functional requirements are met. Compliance testing will make use of off-the-shelf verification and validation tools, where available, and custom test plans and procedures.

Compliance testing of the initial in-situ data providers will be performed by the NOAA CSC Data Transport Laboratory (DTL). CSC will use desktop tools such as XML Spy and Apache JMeter to make appropriate HTTP requests of the relevant services. The service requests and responses will be documented and critiqued. The DTL will consult with the DIF project team and the data providers to insure the designed data requests are reasonable and appropriate within the context of the DIF project. The DTL will implement a simple data repository to into which the data responses will be inserted in order to demonstrate the consistency, or lack thereof, of the responses from the data services. A more detailed description of the testing regime is contained in RD8.

Responsibility and methodology for compliance testing of remotely sensed data is under development.

4.3.2. DIF Client Testing

The purpose of client testing is to ensure that software clients on the decision-support tool side are able to request and receive data from one or more data providers using the standardized formats and access methods. Client testing will validate that the services implemented at the data providers are compatible with any customer side application software modifications required to request and receive the data. Custom test plans and procedures will be developed for each customer decision-support tool that will ingest DIF data. End-to-end testing will be performed by customer representatives with participation from IOOS program representatives.

5. BUSINESS REFERENCE MODEL

This section describes the business reference model for the DIF, including the approaches for incorporating DIF data, services and technology into DIF customers' existing business processes.

The DIF consists of a series of distributed implementation projects, undertaken by data providers and customers (model/analysis tools) within their own IT infrastructure. Each of these projects is being executed according to specific project plans, deliverables, and schedules.

The implementation projects described here are the initial set required to conduct an assessment of the value of the integrated DIF data on specific design support tools/models. It is anticipated that the success of the DIF will lead to expansion of the number of data providers and customers participating.

To design, and implement the DIF projects, the NOAA IOOS Program utilizes a combination of capacity and expertise resident in NOAA and contract resources. Project teams and affiliated working groups composed of cross Line Office and Goal Team representatives design, carry out, or direct the technical work and building of DIF components, and will be involved later in the testing and evaluation of the DIF.

Specific DIF components expected to be built include pilot implementations at specific data provider and customer locations and reference implementations that can be distributed as software "toolkits" to facilitate expansion of participation in the DIF beyond the initial data provider and customer groups.

The DIF design is built upon the Functional Requirements from identified customers (RD3), and approved CONOPS (RD4).

5.1. Data Customers

Four initial Data Customers have been selected for the DIF: Harmful Algal Bloom Forecast System (HAB-FS), Integrated Ecosystem Assessments (IEA), Coastal Inundation (CI) and Hurricane Intensity (HI). The selection was based on the following major criteria:

- high ROI level;
- readiness for cooperation;
- majority of data used is received from NOAA data providers.

The sections below provide a brief description of each Data Customer project and how DIF data, services and technology will be integrated into DIF customers' existing business processes. A complete description of the initial Data Customer projects is contained in RD7.

5.1.1. Harmful Algal Bloom Forecast System (HAB-FS)

This project is based on the existing operational model for the Eastern Gulf of Mexico – Harmful Algal Bloom Forecast System (HAB-FS), and the experimental model for the Western Gulf of Mexico. The existing HAB-FS output is in the form of a HAB Forecast Bulletin which contains an operational forecast of bloom extent, transport, intensification, and impact over 0-3 days. This bulletin is used by local decision makers and the general public to inform beach closure and other regional public health and safety decisions.

The project is focused on the enhancement of the transport model of HAB-FS by providing additional observations and modeled data using DIF formats and services; the following results are anticipated:

- develop an implementation strategy for serving oceanographic model data into core IOOS decision support tools;
- increase HAB forecast objectivity;
- measurably increase the spatial and temporal accuracy (skill) and precision (reproducibility) of HAB forecasts;
- develop a probability based, spatially-articulated bloom intensification estimate;
- extend forecast spatial range to the areas where the forecast has been unavailable, e.g., Tampa Bay;
- extend forecast temporal range from present 3 days to 5 days and more;
- increase probability of providing an accurate nowcast for the time periods when satellite imagery is not available due to clouds.

The project is divided into two parts (phases). Phase 1 concentrates on the capability of the existing HAB-FS bulletin software to ingest surface currents data provided by CO-OPS and NDBC using data standards and protocols identified by the DIF. Phase 1 is developed by collaborative efforts of the CSC and CO-OPS teams. Phase 1 will result in reliable access to DIF-formatted surface currents data served by NDBC and CO-OPS, and automatic ingestion of that data into the operational HAB-Forecast System bulletin generation application for the western Florida region of the Gulf of Mexico.

Phase 2 expands the available DIF data to include forecasted surface currents from CSDL. This data, along with other DIF and non-DIF data in use by the HAB-FS, will be ingested in a spatially-articulated transport model (not currently in use by HAB-FS). The partners will work together to identify historical bloom events along the West Florida Shelf and CSDL will provide modeled 2-dimensional surface current forecasts during those events.

In hindcasting mode, the enhanced transport model will be used to determine whether historical HAB nowcast and forecast quality benefits from the new transport model. IOOS/DIF, NCCOS/CCMA, and OCS/CSDL specialists will collaborate to develop Phase 2 of the project, and assess the result.

Phase 2 will make use of the General NOAA Operational Modeling Environment (GNOME) transport model (GNOME was originally developed by the Office of Response and Restoration (OR&R) to predict oil spill trajectories in the marine environment, and HAB and oil spill behavior have demonstrated an apparent similarity). GNOME's "Diagnostic Mode" will be invoked to forecast bloom position along the West Florida shelf, predicting how currents move and spread blooms.

5.1.2. Coastal Inundation (CI)

This project is intended to showcase the benefits of the IOOS DIF by providing enhancements to how National Weather Service's Storm Surge Team at Tropical Prediction Center (TPC) along with local Weather Forecast Offices create operational Sea Lake and Overland Surges from Hurricanes (SLOSH) forecasts, and by allowing users of the SLOSH Display program to utilize IOOS DIF real time water level observations, tidal predictions and some statistics in operational forecasts and other decision making.

In addition this project will provide an enhance briefing capability for TPC to support live local to national level media interviews, and assist TPC and WFOs in their briefings to emergency managers and other decision makers at all levels of state and federal government during a hurricane landfall threat.

The project will enhance the existing SLOSH Display application to allow integration of the IOOS DIF wind and water level observations, tidal predictions, and other water level products with the surge forecast. This will allow the forecaster to compare the observations to the forecasts and to have additional information to determine total water level, which in turn will assist them with their local statements. In addition, the SLOSH Display presentation capability will be enhanced in order to use it for media and other user briefings: the application was not initially designed to be used in these briefings, but proposed visualization improvement will assure that valuable information is graphically depicted in a simple and clear manner to better inform decision makers.

TPC and regional WFOs are both expected to benefit from the project, however, due to different goals and objectives, the anticipated benefits are also different.

Anticipated benefits to TPC come to the following ones:

- direct automatic integration of the real-time data into SLOSH program instead of "web scraping" in order to facilitate model initialization and potentially yield efficiency in the pre- and post-storm analysis.
- ability to communicate the actual situation before landfall to emergency managers and general public as the storm progresses (currently, TPC does not communicate real-time water levels to external audiences).

- dramatic improvement of visual presentation of the information provided, to better inform emergency managers and external users.

Anticipated benefits to WFOs boil down to the ability to increase the accuracy and reliability of local water surge guidance, and hence increase informational awareness of local emergency managers.

5.1.3. Hurricane Intensity (HI)

The goal of this project is to evaluate the benefits of integration of ocean data in DIF standards into an air-sea numerical model, to aid the scientific and operational community to improve Atlantic hurricane intensity forecasts. If successful the same tools and methodologies applied to this effort could be applied to improve the forecast of tropical cyclone intensity in all basins where tropical cyclones occur.

NOAA/NWS's National Centers for Environmental Prediction (NCEP) will collaborate with NOAA/OAR's Atlantic Oceanographic Marine Laboratory (AOML) to conduct hindcast evaluations for three intense Category 4-5 hurricanes from 2005 – Rita and Wilma in the Gulf of Mexico and Emily in the Caribbean Sea. Research indicates that tropical cyclones intensify over warm mesoscale features located in the open ocean. The two-fold purpose of this project is to evaluate this effect by incorporating new ocean temperature data into the forecast models and to assess the operational requirements of DIF-formatted data for these purposes.

Currently, hurricane predictions are developed at NCEP using an atmospheric component (HWRF) coupled to an ocean model (HYCOM). For this project, the ocean component – via HYCOM – will be initialized and "forced" at the boundaries with data from the operational RTOFS (Real time Ocean Forecast System)-Atlantic. New temperature data – in DIF formats – will be developed by AOML from various sources and provided to NCEP to be assimilated in RTOFS-Atlantic. Before and after scenarios will be conducted for Rita, Wilma, and Emily to evaluate forecast accuracies with and without the new temperature data.

New temperature data provided by AOML will consist of temperature profiles obtained from regional XBTs, AXBTs, profiling floats, thermistor chains, and moorings. Some of these data are archived at NODC, while other data are available from AOML. In addition, AOML will develop synthetic water temperature profiles using historical statistical relationships between altimetry-derived sea surface height and hydrographically-derived depths of selected isotherms, and forcing the temperature of the mixed later to the satellite-derived sea surface temperature. The altimeter data will be provided by AVISO and will include the constellation of all available altimeters, including NASA's Jason-1, ESA's Envisat, and US Navy's GFO. Hydrographic and synthetic data will be converted to DIF standards.

This project will assess the value of ocean data in tropical cyclone intensity prediction. A positive outcome may result in transitioning new data to operational forecast models per reducing error in hurricane intensity forecasts. Synthetic temperature profiles may also provide qualitative information on how to improve tropical cyclone prediction studies via ocean observing capabilities.

5.1.4. Integrated Ecosystem Assessments (IEA)

This project involves development of tools to transform various data to and from standard IOOS-DIF standards and to serve the IEA decision support model by achieving improved integration of selected data sets. An IEA is a formal synthesis and quantitative analysis of existing information on relevant natural and socio-economic factors in relation to specified ecosystem management objectives.

A useful IEA must provide an efficient, transparent means of summarizing the status of ecosystem components, screening and prioritizing potential risks, and evaluating alternative management strategies against a backdrop of environmental (e.g., climatic, oceanographic, seasonal) variability. To accomplish these goals, it must be possible to access a wide-spectrum of ocean observing data, and to transform these data into formats of use to the systems involved in the IEA analyses.

The project should result in modification of ERDDAP software to provide enhanced integration with selected IOOS DIF data services and, in collaboration with the Ecosystem Goal Team, to prototype the implementation of these services into the IEA model for the Gulf of Mexico and California Current Regions.

It is expected that the development of ERDDAP will allow a seamless access to the data being served by NOAA providers as well as any provider that serves data via any Web SOS service. That will benefit IEAs, as more data will be available to use in models, calculations and assessments. A definitive benefit of ERDDAP usage should include a resulting social and/or economic gain of IEA, and is anticipated to be significant. However that kind of result is completely IEA type dependant and IEA-to-IEA variation are substantial.

5.2. Core Variables

Seven initial Core Variables have been selected for the project implementation: Sea Temperature, Water Level, Ocean Color, Currents, Water Salinity, Winds, and Waves. It is anticipated that these initial variables are of high value for the selected Data Customers and can be made readily available from a limited number of the selected Data Providers.

5.3. Data Providers

Three initial major NOAA Data Providers have been selected for the initial implementation of the DIF: NDBC, CO-OPS, and CoastWatch. The selection is based on the fact that these providers together represent a significant amount of data for the selected Core Variables, and therefore they are the most reasonable candidate for the DIF projects. It is anticipated that the experiences and lessons learned from the initial Data Provider implementations will lead to expansion in the number of data providers, include NOAA and non-NOAA sources of data.

The Figure 5.1 below indicates the Core Variables that will be served by DIF Data Providers, and presents their feature types:

	Currents	Water Level	Sea Temp	Salinity	Surface Winds	Waves	Ocean color (chlorophyll)
Single station	NDBC, CO-OPS	NDBC, CO-OPS	NDBC, CO-OPS	NDBC, CO-OPS	NDBC, CO-OPS	NDBC	n/a
Group of stations	NDBC	NDBC	NDBC	NDBC	NDBC	NDBC	n/a
Profile	NDBC, CO-OPS	n/a	NDBC, CO-OPS	NDBC, CO-OPS	n/a	n/a	n/a
2D grid	NDBC						CoastWatch

Figure 5.1: Variables and feature types to be offered by DIF data providers.

5.3.1. NDBC

NDBC collects, processes, and formats meteorological and oceanographic data from the NWS-Buoy and the Coastal-Marine Automated Network (C-MAN) systems. It also serves as the IOOS Data Assembly Center (DAC). As the IOOS DAC, NDBC collects data from regional ocean observing systems, quality controls the data, and distributes it via the WMO GTS in real-time. NDBC also makes the quality controlled data available on their web site and via netCDF files on its OPeNDAP server.

Three out of four DIF data customers are interested in data provided by NDBC:

- HI modeling;
- IEA modeling; and
- HAB-FS modeling.

5.3.2. CO-OPS

CO-OPS collects, analyzes, and distributes historical and real-time observations and predictions of water levels, coastal currents, and other meteorological and oceanographic data. CO-OPS provides data in use by the customer models via two systems: the National Water Level Observation Network (NWLON) and the Physical Oceanographic Real-Time System (PORTS).

Two DIF data customers are interested in data provided by CO-OPS:

- Coastal Inundation modeling; and
- Hurricane Intensity modeling.

5.3.3. CoastWatch

The CoastWatch program at NESDIS provides timely access to the data collected by NASA's Aqua satellite. The DIF core variable of ocean color is the result of collecting water-leaving radiance data via the Moderate Resolution Imaging Spectroradiometer (MODIS) system, and post-processing it by the CoastWatch's OKEANOS computer system.

CoastWatch provides MODIS Aqua ocean color data through the CoastWatch website via HTTP. Color data and images can also be accessed via FTP from OKEANOS servers.

Two DIF data customers are interested in the ocean color data provided by CoastWatch:

- Hurricane Intensity modeling; and
- Harmful Algal Bloom Forecast System (currently, the HABs FA customer uses SeaWiFS data as a primary source and access MODIS data as a backup. However, with uncertain access in the future to near real-time SeaWiFS data, MODIS data has become more important to the HAB bulletin process and the analysis tool has been updated to accommodate MODIS data.)

5.4. DIF Projects Oversight and Management

Each customer and data provider implementation project will be managed through their individual Statements of Work (SOWs) which define project scope, tasking, schedule, milestones and deliverables. The NOAA IOOS program representatives will provide technical oversight of individual projects through periodic technical reviews, management meetings, and progress reports. The NOAA IOOS program will maintain a master schedule of all projects and dependencies and NOAA IOOS program representatives will work together to ensure that budget, schedule and technical expectations are met for all projects. Any discrepancies from the SOWs will be managed and escalated as needed to minimize impacts on the overall program.

5.5. Key Risks and Mitigation Strategies

Figure 5.2 below identifies the key technical risks for the DIF project, and describes mitigation strategies for each.

Area	Risk	Mitigation Strategy
Standards	Migrating into a standards-based framework may be difficult because the NOAA data community has a limited history of effectively using standards, and selected standards may not be palatable to all members of the community.	The Working Group on Standards and Data Encodings (WSDE) has broad representation to ensure all viewpoints are heard. Collaborations with NOAA Enterprise Architects, GEO-IDE and GEOSS used as inputs to the selection process.
Standards - SOS	SOS is a fairly new service which is not yet widely adopted and somewhat unproven.	a. Work closely with OGC to ensure timely addressing and correction of issues. b. Constantly analyzing alternative technologies for different approaches with similar outcome.
Services – Performance	Services implemented by DIF data providers may not provide the performance (speed of data retrieval) expected by customers.	c. Work with data providers to optimize performance. d. Look for alternative technologies to provide the necessary performance level
Standards – Encodings	XML encodings and conventions selected for in-situ data are fairly rich and verbose which may make it difficult to transport the data and transform them into other formats.	Work out more limited sets of initial encodings without detriment to functionality.
IT Security	IT security requirements may limit DIF service availability.	Work closely with NOAA IT security stakeholders to define a core set of requirements which meet IT security and service availability and performance needs.
Performance Assessment	Data Customer projects do not yield quantifiable performance metrics.	Identify performance metrics or other benchmarks that can describe the benefits subjectively or anecdotally.
Performance Assessment	Results of performance assessments do not show significant value of integrated data.	Work closely with Data Customers to identify projects with high likelihood of success; perform requirements analysis to identify key needs.
Data providers	Selected data providers do not provide data of importance to customers.	Identify customer projects that can make use of data from selected data providers.

Figure 5.2: Key Risks and Mitigation Strategies

5.6. Requirements Traceability

Requirements traceability was conducted to compare the requirements defined in [RD3] with the design concepts documented herein. A detailed requirements traceability matrix is contained in Section7; a summary of some of the major requirements gaps is presented below:

1. **Required data sets –** the requirements for data sets needed to support the four Data Customers were developed through several interviews and meetings with potential customers in the target model areas (HABs, Coastal Inundation, etc), before specific customer projects were identified.

   The DIF design and implementation approach described herein does not provide all of the required data sets identified in [RD3]. Resource constraints required selection of only a few DIF data providers, and all required data sets are not available through those providers. Customer projects have been identified to maximize the use of data sets that are available through the planned DIF implementation. Further, some of the data sets initially identified as requirements to support target model areas are not required for the specific customer projects selected.

2. **Metadata –** the requirements state that each DIF logical dataset or data service shall have at least one FGDC or ISO metadata record. The current metadata design is not definitive in this regard.

3. **Push services –** the DIF requirements call for both pull and push services from data providers. The current design is focused on pull services; a clear approach for push services has not yet been defined.

6. DATA REFERENCE MODEL

This section provides an overview of the initial DIF Data Model, including discussion on how it relates to the Federal Enterprise Architecture Data Reference Model (FEA DRM), and describes the data encoding and transport specifications to be employed.

6.1. DIF Data Model Overview[2]

The DIF Data Model is built in accordance with the FEA DRM. The DRM was used as a template for the DIF Data Model first and foremost because their goals and objectives demonstrate a high level of similarity:

- FEA DRM is a framework with a primary goal to enable information sharing and reuse across the federal government via the standard description and discovery of common data and the promotion of uniform data management practices. DIF is a framework with the same goal within IOOS.
- The DRM provides a flexible and standards-based approach to accomplish its goal. DIF pursues the very same objective.
- The scope of the DRM is broad, as it may be applied within a single agency, within a Community of Interest (COI), or cross-COI. While the initial implementation scope of DIF is somewhat limited, the framework concept is broad and it is anticipated that the initial DIF implementation will be leveraged for the broader IOOS.

The DRM provides a standard means by which data may be described, categorized, and shared. These are reflected within three areas, as illustrated in Figure 6.1:

- **Data Description**, which provides means for a COI to unify the structure (syntax) and meaning (semantics) of the data that it uses.
- **Data Context**, which provides additional meaning to data in order to relate it to the purposes for which it was created and used; it facilitates discovery of data through an approach to the categorization of data according to taxonomies.
- **Data Sharing**, which facilitates the access and exchange of data.

A projection of the DIF Data Model core functional components onto DRM standardization areas is presented in Figure 6.2, and described in more details in the subsequent sections.

[2] Based on the DRM overview in ***The Data Reference Model v 2.0*** by Federal Enterprise Architecture Program, November 17, 2005

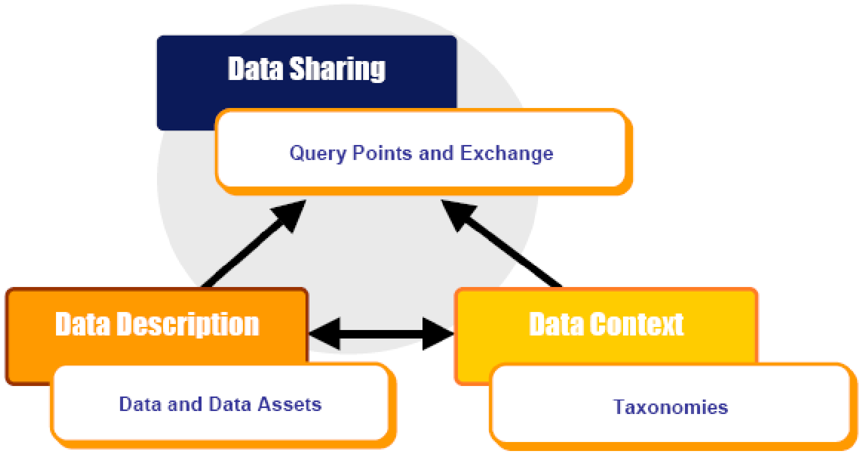

Figure 6.1: DRM Standardization Areas
(from *The Data Reference Model v 2.0*)

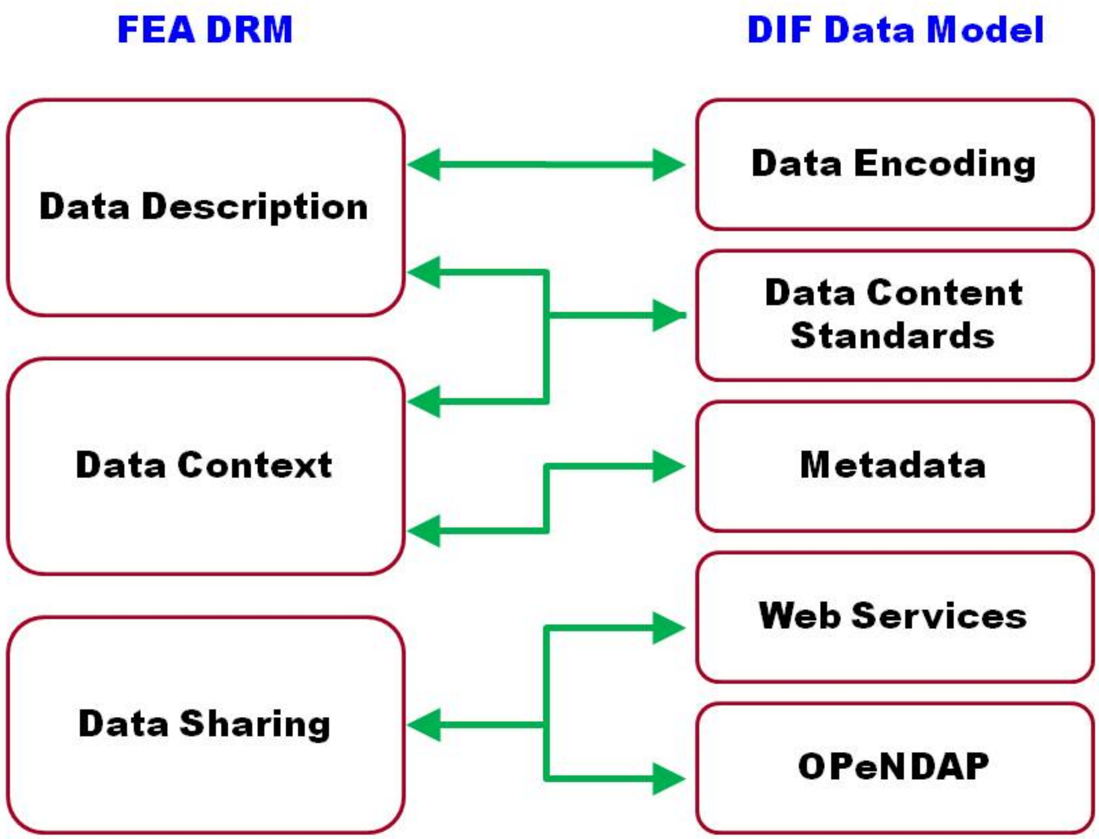

Figure 6.2: DIF Data Model projection on the FEA DRM

6.2. Data Description and Context

DIF data description and context is achieved through adoption and extension of a series of data content standards and encodings. The content standards and encodings must support observational (*in situ* and remotely sensed), modeled data, and metadata.

Both observational and modeled data may be point/site specific or spatial (gridded, contour, pictorial).

Data types, services and encodings for the initial DIF Core Variables are described in more details in subsequent sections and separate documents. Figure 3.1 provides an overview of the selected services and encodings by data type.

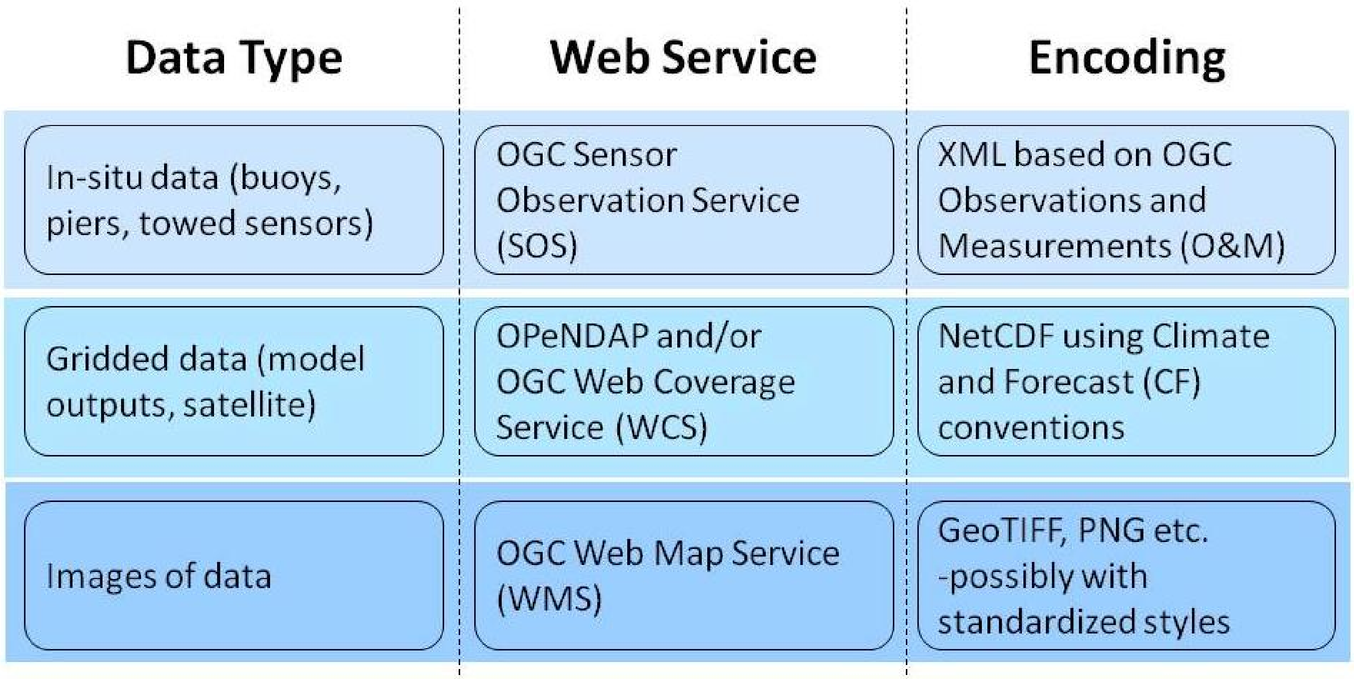

Figure 6.3: DIF Data types, Web Services and Data Encodings

6.2.1. In Situ Data Content Standard & Encoding

For encoding in-situ featured data, the DIF will initially use Geography Markup Language (GML). Certain Extensible Markup Language (XML) Schema Definitions (XSD), and corresponding sample XML instances, based on the GML application profile of the OGC Observations and Measurements (O&M) specification version 1.0.0, will be adopted, refined or developed.

The corresponding XML Schemas and encodings will be published in DIF data content standard and encoding documents.

6.2.2. Gridded Data Content Standard & Encoding

DIF gridded data will be encoded in Network Common Data Form version 4 (NetCDF-4) with Climate and Forecast (CF) conventions. The conventions may be constrained even further in the process of design in order to better define and represent data sets, and mitigate the risks.

6.2.3. Pictorial Data Content Standard & Encoding

For encoding spatial geographical data as a digital image file suitable for direct display on a computer screen, general pictorial formats such as PNG, GIF, JPEG or TIFF will be used.

6.2.4. Metadata[3]

For Metadata encoding and processing, the objective is implementation of a single standard (or family of standards); either FGDC or ISO 19115 and ISO 19115-2. These standards establish a common framework for communicating information about geospatial data sets. They include mandatory elements as well as recommended or optional elements. This information includes:

- identification of sources and stewards;
- details about the data's organization including number and type of features, spatial reference, and attributes (with a description of each) and definitions of acceptable ranges of values;
- descriptions of data quality;
- use constraints;
- information needed to successfully access, transfer and process the data.

The existing Metadata components of the NOAA Data Providers will be investigated and recommendations will be provided, to the extent practicable, for conversion to the accepted standards. Metadata activities will be applied to all DIF datasets and Data Providers.

With regard to metadata the DIF project will, at a minimum:

- Evaluate existing documentation and define actions needed for adoption and adaptation of international standards for metadata;
- Assist data providers in identifying and/or creating appropriate metadata;
- Provide information such as lessons learned for other decision support tools within NOAA.

Ultimately, a metadata standard for the DIF will be established, which the DIF data providers will be encouraged to implement.

In an effort to identify high-quality, standards compliant metadata, the NOAA IOOS Program has compared existing metadata from NWS/ NDBC, NOS/CO-OPS and NESDIS/CoastWatch against the Federal Geographic Data Committee (FGDC) and ISO 19115/19115-2 standards. At present, the majority of metadata available from these data providers is not compliant with either of these standards. The IOOS Program recognizes that the ISO standards provide for more comprehensive metadata, allow for greater international use and exchange and may offer greater machine-to-machine interoperability. For those reasons, we expect that the DIF partners will agree that the ISO Standards should be the target for DIF. In addition, the DIF efforts will need to adhere to NOAA/NOS metadata policy and the use of FGDC metadata standards, the future of which includes the North American Profile of ISO 19115. The investigation efforts described above will lead to a decision on a standard the DIF can formally adopt.

[3] Excerpted from Ted Haberman's & Rob Ragsdale's Metadata Guidance for NOAA IOOS Program's Data Integration Framework.

The Marine Community Metadata Profile of ISO 19115 was reviewed to determine if this profile should be adopted by the DIF. Through the review, it was found that the profile would probably not be useful for DIF data providers or users. It excludes elements of 19115 that add value and includes others that would limit the profile's usefulness. Thus, this profile of ISO 19115 is not presently being considered for use by the DIF.

6.3. Data Sharing

6.3.1. Service-Oriented Architecture (SOA)

A service-oriented architecture has been selected as the means for implementing data sharing across the DIF. Web services are self-contained, self-describing, modular applications that can be published, located, and invoked across the Web. Once a Web Service is deployed, other applications (as well as other Web services) can discover and invoke the deployed service as well as data. The Web Services provide a uniform and widely accessible interface to traditional middleware platforms.

The standardized access could have been achieved by implementation of certain middleware platforms like RMI, Jini, CORBA, DCOM, etc. However, none of them was chosen as a single standard middleware platform for the DIF because it is very difficult, if not impossible, for large enterprise like NOAA to standardize on a single middleware platform. More than one platform may be required because different departments have different requirements, and there is also a need to interoperate with non-NOAA data providers in a heterogeneous environment.

A very special aspect of the problem is IT security. It is important to organize a smooth operation of the geographically and logically distributed system through a number of firewalls, and keep a high level of security intact.

Web Services allows interoperation in complex heterogeneous environment because they are based on open specifications and protocols like XML, HTTP, SOAP, WSDL, and UDDI. The use of XML greatly simplifies design and deployment of the services, and HTTP usage resolves the problem with firewalls as HTTP ports are usually opened. However, it is not mandatory to use HTTP transport for Web Services, FTP or SMTP can be used instead.

Figures 6.3 and 6.4 indicate the scope of the services established by the DIF. The details of the services are described in the subsequent sections.

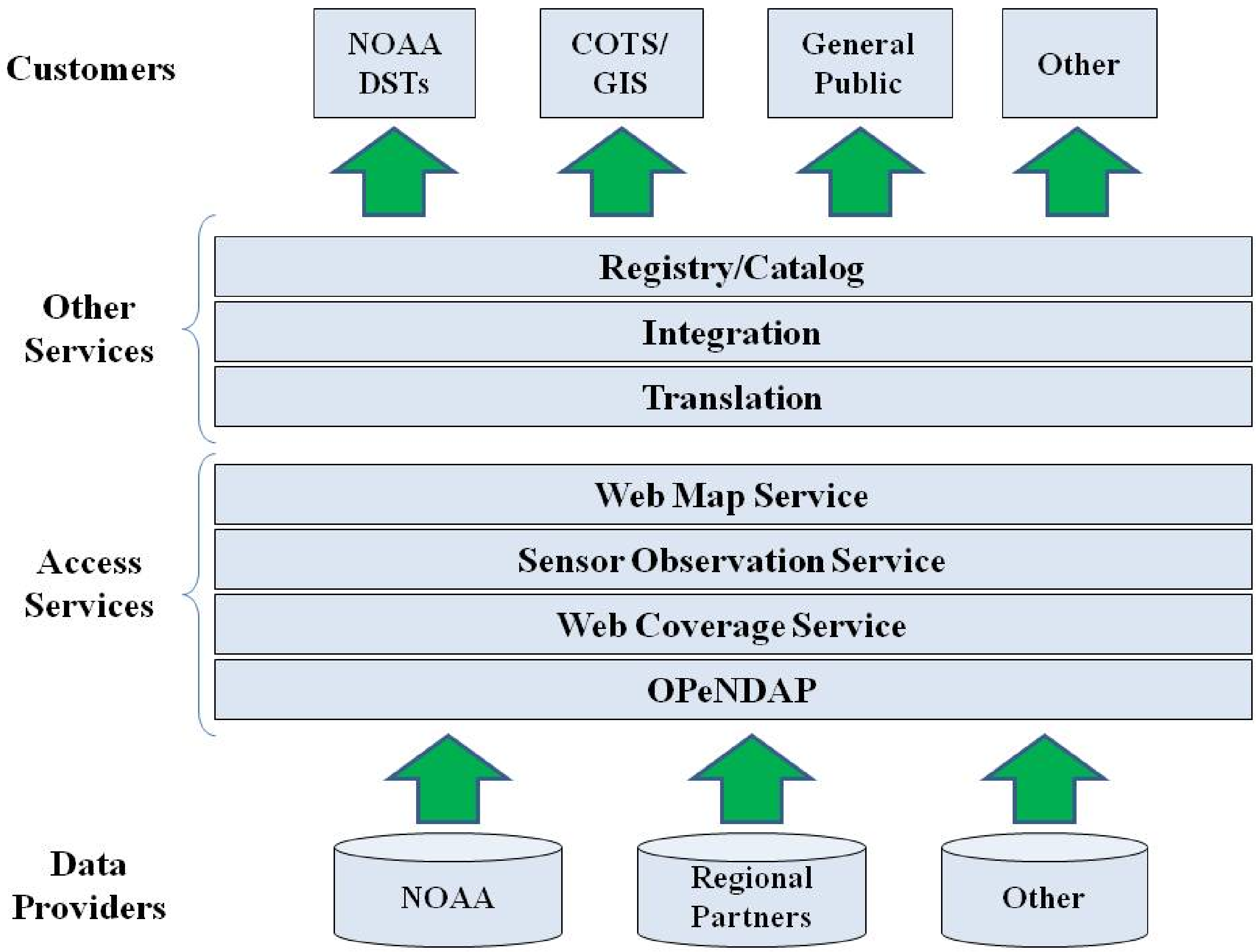

Figure 6.4: DIF Service Layers
(from Jeff de La Beaujardière "The NOAA IOOS Data Integration Framework: Initial Implementation Report", Ocean-2008, Quebec, Canada, 2008)

6.3.2. Data Access Services: OGC Web Services

Within the Web Services environment, OGC Web Services represent a mature, standards-based framework that enables seamless integration of a variety of online data. It allows different systems to communicate with each other across the Web using well established technologies such as XML and HTTP. OGC Web Services provide a vendor-neutral, interoperable framework for web-based discovery, access, integration, analysis, exploitation and visualization of multiple online geodata sources, sensor-derived information, and geo-processing capabilities.

The OGC Web Services concept is based on a model of a distributed service-oriented architecture operated by data stewards, which is in line with NOAA Enterprise Architecture principles.

The following Web Services, protocols and data encoding specifications were adopted in the initial phase of DIF for operating with different types of data:

- for “in-situ” data (buoys, piers, towed sensors) – OGC Sensor Observation Service (SOS), version 1.0.0 ("core operations profile" only), and XML encoding based on OGC Observations and Measurements (O&M) specifications.
- for gridded data (model outputs, satellite) – OGC Web Coverage Service (WCS), and NetCDF encoding with Climate and Forecast (CF) conventions.
- for images of data (dynamically serving pictorial maps) – OGC Web Map Service (WMS), and any general picture encoding format, e.g., GIF, PNG, JPEG, etc.

For gridded data, the DIF also recommends the Open-source Project for a Network Data Access Protocol (OPeNDAP). Unlike the WCS, which can work only with regular grids, OPeNDAP also supports irregular grids. OPeNDAP is under review as a recommended IOOS DMAC data transport mechanism and is well used in the scientific community.

6.3.2.1. Sensor Observation Service (SOS)

OGC’s Sensor Observation Service (SOS) provides an API for managing deployed sensors and retrieving sensor data. The deployed sensors of various types may be also grouped into several “constellations”, and then accessed through SOS.

SOS has three mandatory core operations: **GetObservation, DescribeSensor**, and **GetCapabilities**.

The **GetObservation** operation provides access to sensor observations and measurement data via a spatio-temporal query that can be filtered by phenomena. A **GetObservation** message contains one or more elements that constrain the observations to be retrieved from a Sensor Observation Service. Each **GetObservation** query element has mandatory attributes of service and version. The mandatory version element attribute must correspond to the specific service interface version negotiated between the service and client during the service binding process.

The **DescribeSensor** operation retrieves detailed information about the sensors and the processes generating those measurements.

The **GetCapabilities** operation provides the means to access SOS service metadata.

Several optional, non-mandatory operations have also been defined. There are two operations to support transactions, **RegisterSensor** and **InsertObservation**, and six enhanced operations, including **GetResult**, **GetFeatureOfInterest**, **GetFeatureOfInterestTime**, **DescribeFeatureOfInterest**, **DescribeObservationType**, and **DescribeResultModel**.

Used in conjunction with other OGC specifications, the SOS provides a broad range of interoperable capability for discovering, binding to and interrogating individual sensors, sensor platforms, or networked constellations of sensors in real-time, archived or simulated environments.

6.3.2.2. Web Coverage Service (WCS)

OGC’s Web Coverage Service (WCS) supports retrieval of “coverages”, i.e., digital geospatial information representing space-varying phenomena. WCS is explicitly called out in the Global Earth Observation System of Systems (GEOSS) architecture and is supported by some commercial off-the-shelf (COTS) Geographic Information System (GIS) tools.

A WCS provides access to detailed sets of geospatial information in forms that are useful for client-side rendering, multi-valued coverages, and input into scientific models and other clients. The WCS may be compared to the OGC Web Map Service (WMS) and the Web Feature Service (WFS); like those it allows clients to choose portions of a server's information holdings based on spatial constraints and other criteria. However, unlike the WMS and WFS, the WCS provides rich sets of data with its original semantics (instead of pictures) together with their detailed descriptions, and operates with a spatio-temporal domain and multidimensional sets of properties (instead of discrete geospatial features).

The WCS provides three operations: **GetCapabilities**, **DescribeCoverage**, and **GetCoverage**.

The **GetCapabilities** operation returns an XML document describing the service and coverages that clients may request. When the **GetCapabilities** operation does not return such descriptions, then equivalent information must be available from a separate source, such as an image catalog (registry).

The **DescribeCoverage** operation lets clients request a full description of one or more coverages served by a particular WCS server. The server responds with an XML document that fully describes the identified coverages.

The **GetCoverage** operation is normally run after **GetCapabilities** and **DescribeCoverage** operation responses have shown what requests are allowed and what data are available. The **GetCoverage** operation allows retrieval of subsets of coverages. A WCS server processes a **GetCoverage** operation request and returns a response to the client that either contains or references the requested coverage(s). The **GetCoverage** operation returns coverage data, encoded in a well-known format (e.g., netCDF, HDF-EOS, or GeoTIFF).

6.3.2.3. Web Map Service (WMS)

A Web Map Service (WMS) produces maps directly from spatial geographical data as a digital geolocated image file suitable for display on a computer screen. WMS-produced maps are rendered in pictorial formats such as PNG, GIF, or JPEG.

The three operations defined for a WMS are **GetCapabilities**, **GetMap**, and **GetFeatureInfo**. The G**etFeatureInfo** is optional.

The purpose of the mandatory **GetCapabilities** operation is to obtain service metadata, which is a machine- as well as human-readable description of the server's information content and acceptable request parameter values.

The **GetMap** operation returns a map. Upon receiving a **GetMap** request, a WMS either satisfies the request or issues a service exception.

GetFeatureInfo is an optional operation. The **GetFeatureInfo** operation is designed to provide clients of a WMS with more information about features in the pictures of maps that were returned by previous Map requests.

6.3.2.4. OPeNDAP

The OPeNDAP provides a way to retrieve data from anywhere on the network, and deliver data to a wide variety of customer applications. The OPeNDAP 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.

The OPeNDAP architecture uses a client/server model, with a client that sends requests for data out onto the network to some server, which answers with the requested data. This model is similar to the one used by the Web Services where client programs submit requests to web servers for the data that make up web pages. In fact, Web browser can serve as an OPeNDAP client with limited functionality. Usually OPeNDAP client can do much more than just browse the data; it is capable of retrieving subsets of data, and also has the ability to aggregate data from several sources in one transfer operation.

6.3.3. Other Services

6.3.3.1. Registry and Catalog Services

It is intended that the DIF project will implement a Registry service to document the full range of available services and data holdings; in addition, a Catalog service should be developed to provide a catalog of the resources to the Data Customers. A specific technical approach for the implementation has not yet been determined. It is likely that the DIF Registry Service will be developed on base of the existing service, e.g. IOOS or GEOSS registry, rather than built from scratch. The OGC Catalog Services for Web (CS/W) profile seems also promising as one of the candidates.

Implementation of a DIF registry is not critical to support the four Data Customer projects, but will help expand the utility of the DIF services to a wider group of Data Customers.

6.3.3.2. Data Integration and Translation Services

Although the DIF SOS Web Service implementations at the Data Providers (i.e. NDBC and CO-OPS) are expected to provide integrated access to multiple data holdings of these Providers (e.g. Data Customers should be able to retrieve data from all NDBC Data Assembly Centers with a single SOS request), no specification for such a “Data Integration Service” that should aggregate data records from multiple independent SOS into a single record, has been yet developed. However, it is intended that the DIF will enable that level of data aggregation later because the standardized access services and encodings utilized by the DIF are capable to support it.

A data translation and visualization service known as ERDDAP will be used as a foundation for the DIF Data Translation service implementation. The ERDDAP has been developed by the NOAA National Marine Fisheries Service (NMFS) Southwest Fisheries Service Center (SWFSC) Environmental Research Division (ERD). 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 requested by the Data Customer. ERDDAP will be capable of reading and translating from the SOS implementations that use the DIF XML encoding specification and will also translate non DIF formatted data into a DIF compliance SOS service. That capability along with other ERDDAP features is expected to carve the way to a full-scale distributed, reliable, fault-tolerant translation service.

6.3.4. IT Security

6.3.4.1. SOA Security Overview

The flexibility of service-oriented architecture has a cost – it brings some additional concerns over security due to the number of communities, domains, and platforms that may be crossed in executing a business process based on SOA.

The NIST Special Publication 800-95, "Guide to Secure Web Services", addresses some specific security issues:

- confidentiality and integrity of data that is transmitted via Web services protocols in service-to-service transactions, including data that traverses intermediary services;
- functional integrity of the Web services that requires the establishment of trust between services on a transaction-by-transaction basis;
- availability in the face of denial of service attacks that exploit vulnerabilities unique to Web service technologies, especially targeting core services, such as discovery service, on which other services rely;
- SOAs are dynamic and can seldom be fully constrained to the physical boundaries of a single network;
- access protocol messages as well as data are usually transmitted over HTTP, which is allowed to flow without restriction through most firewalls;
- Transport Layer Security (TLS), which is used to authenticate and encrypt Web-based messages, cannot accommodate Web services' inherent ability to forward messages to multiple other Web services simultaneously.

On the other hand, Web services rely on existing protocols and will coexist with other network applications in the same environment. Therefore, the majority of Web service security problems can be mitigated by using traditional security tools, such as firewalls, intrusion detection systems (IDS), and secured operating systems, which have been in effect before implementation of Web services applications, along with adhering to the strong IT security policy.

The NIST SP 800-95 recommends enforcing the following measures to improve the security of Web services transactions:

- authentication and identity management across domains and environments;
- authorization and confidentiality (access control);
- integrity (no inappropriate modifications are made);
- availability (reliable service, no denial of service);
- non-repudiation (positive identification and inability to deny providing or receiving services);
- auditing and monitoring;
- security administration and policy management.

6.3.4.2. DIF IT Security Implementation Approach

Because of its distributed nature, the DIF approach to IT security will also be distributed. The DIF is a framework, comprised of numerous distributed components, each with their own IT security boundary. Because of this, it is not anticipated that the DIF itself will need its own system Certification & Accreditation (C&A). Rather, C&As will be done locally at Data Provider and Data Customer locations in accordance with local IT security office guidelines.

7. REQUIREMENTS TRACEABILITY MATRIX

The matrix below provides requirements traceability between the design presented in this document and [RD3].

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
GEN045	All DIF data, including model outputs, defined in Section 5.2 shall adhere to a common data model or set of models and conventions to ensure consistency of content and format. The common data model(s) and conventions shall adhere to DIF-approved standards.	H		1. Correct in regard to CI, IEA, and HI (?) models 2. Not entirely correct in the HAB-FS case: currently some data is supplied in different formats (e.g. HDF)
RDS001	The DIF shall provide MMS ADCP Currents data from NDBC.	H	Y	
RDS035	The DIF shall provide Texas Automated Buoy Systems (TABS) Currents data from NDBC.	H	Y	Not all of them are included in the current list of stations served by NDBC
RDS040	The DIF shall provide Coastal Ocean Monitoring and Prediction System (COMPS) Buoy Currents data from University of South Florida.	H	Y	Not clear as of Jul'08 when the data will be served, who will serve the data, and if the data will be served at all within DIF timeframe
RDS100	The DIF shall provide Gulf Currents data from HF Radar (CODAR) in NetCDF format.	H	Y	
RDS115	The DIF shall provide Current data from shipboard ADCP profiles. GAP: Verify data source provider.	H	?	No plans of providing that data as of Jul'08
MTD001	Each DIF logical dataset or data service shall have at least one FGDC or ISO metadata record.	H		1. DIF Data Content Standards for Core Variables include important metadata & QC information in accord with ISO and FGDC standard requirements. 2. NGDC metadata working group analyzed existing situation with metadata in NDBC, CO-OPS and CW with the goal of accessing standards compliance and improving metadata
MTD002	DIF metadata standards shall adequately define data attribute details (e.g. unit of measure, reporting convention, precision, code definitions)	H		
MTD015	The DIF shall encode all metadata in valid XML and make it available for public access.	H		

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
MTD016	When publishing data utilizing a service interface, the DIF shall report metadata for the service (e.g. using the GetCapabilities record for an OGC service).	H		quality. The final report is due to Aug 15, 2008. 3. Metadata development requires significant funds (about $900K for FY08-09), which have not been secured as of Jul'08, i.e. DIF metadata requirements may not be met in full.
QC015	The DIF shall document quality procedures in the metadata record.	H		
XPT010	The DIF shall provide the capability for the transmission of DIF data sets to customers and archive centers in real-time using DMAC standards and protocols.	H		1. All DIF data content standards and delivery services are developing within DMAC framework. 2. The difference between delayed and real time modes should be clearly defined and documented (i.e. whether real-time refers just to "push" service or to quasi real-time "pull" service as well).
XPT015	The DIF shall provide the capability for the transmission of DIF data sets to customers and archive centers in delayed mode using DMAC standards and protocols.	H	Y	
FILT010	DIF data providers shall assemble compatible distributed data sources, along geographic and temporal boundaries, into seamless logical collections differentiated by variable, to allow users to exploit data from distributed sensor arrays.	H	Y	1. Both NDBC & CO-OPS will provide assembled data, however NDBC will implement a "box" selection method in FY08, and CO-OPS will not through FY08/09.
GEN001	The DIF shall retrieve, process, and deliver to end users data containing the core variable of Temperature.	H	Y	Web Service publicly accessible
GEN005	The DIF shall retrieve, process, and deliver to end users data containing the core variable of Salinity.	H	Y	Web Service publicly accessible
GEN010	The DIF shall retrieve, process, and deliver to end users data containing the core variable of Currents.	H	Y	Web Service publicly accessible
GEN015	The DIF shall retrieve, process, and deliver to end users data containing the core variable of Sea Level.	H	Y	Web Service publicly accessible
GEN020	The DIF shall retrieve, process, and deliver to end users data containing the core variable of Ocean Color.	H	Y	The Data Content Standard acceptance process was completed on July 7, 2008, which may cause a delay in implementation.
GEN021	The DIF shall retrieve, process, and deliver to end users data containing the core variable of Winds.	H	Y	Web Service publicly accessible

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
GEN022	The DIF shall retrieve, process, and deliver to end users data containing the core variable of Waves.	H	Y	Web Service publicly accessible
GEN025	The DIF shall support access and delivery (transport) of real time data, as applicable, for the data sets defined in Section 5.2.	H		
GEN026	The DIF shall support access and delivery (transport) of delayed mode data, as applicable, for the data sets defined in Section 5.2.	H		
GEN030	The DIF shall support data in the form of time series.	H	Y	
GEN035	The DIF shall support data in the form of profiles.	H	Y	1. CO-OPS SOW for FY09 has not been approved yet 2. OSMC SOW has not been completed as of Jul'08
GEN040	The DIF shall support data in the form of gridded data.	H	Y	Web Service publicly accessible
RDS030	The DIF shall provide TGLO 3-D circulation model Currents data from Texas A&M. NOTE: this is not deemed an operational system by NOAA.	L – Not operational	Y	Not confirmed yet
RDS040	The DIF shall provide NOS/Dynalysis NGOM 2D & 3D circulation Model Currents data. Note:	L – Not operational	Y	Not confirmed yet
RDS060	The DIF shall provide NAM model Forecasted Winds data from NCEP sites.	M		No SOW as of Jul'08
RDS070	The DIF shall provide NDFD Forecasted Winds data.	M		No SOW as of Jul'08
RDS180	The DIF shall provide Water Temperature data from CO-OPS systems.	M	Y	Web Service publicly accessible
RDS181	The DIF shall provide Water Temperature data from NERRS.	M	Y	1. Web Service publicly accessible 2. Not all obs may be available (33 stations)
RDS182	The DIF shall provide Water Temperature data from NEP systems.	M		

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
RDS185	The DIF shall provide Water Temperature data from NDBC.	M	Y	Web Service publicly accessible
RDS186	The DIF shall provide Water Level data from CO-OPS systems.	M	Y	Web Service publicly accessible
RDS187	The DIF shall provide Water Level data from NDBC systems.	M	Y	Web Service publicly accessible
RDS190	The DIF shall provide Sea Level data from USGS portable water level gauges via NHC has data.	M		
RDS200	The DIF shall provide Winds data from NDBC	M	Y	Web Service publicly accessible
RDS200	The DIF shall provide Winds data from CO-OPS	M	Y	Web Service publicly accessible
RDS210	The DIF shall provide Waves data from CO-OPS	M		
VAL005	The DIF shall filter and delete duplicate records for data sets collected from multiple sources. **GAP:** Identify precise method for determining which records are duplicates.	H		
TRAN001	The DIF shall provide the capability and tools to transform data, as needed, into the format consistent with the DIF common data model and content standards.	H	Y	SOW is planned to be developed by mid-Aug'08. 1. NDBC aggregates data from different sources and serves them in DIF format 2. Web Service publicly accessible
TRAN020	The DIF shall provide mechanism for aggregation of data of the same data type from multiple data providers. Such aggregation shall be provided over useful geographic and temporal boundaries to allow users to exploit data of the same type from multiple sources.	H	Y	SOW is planned to be developed by mid-Aug'08. 1. NDBC aggregates data from different sources and serves them in DIF format 2. Web Service publicly accessible
MTD003	DIF metadata standards shall adequately convey data file formats or structures.	H		DIF Metadata development process is at the very early stage, and no traceable deliveries have been scheduled yet.
MTD004	DIF metadata standards shall adequately capture data quality information (e.g. QC tests applied, QC flags and flag definitions)	H		

Requirement ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
MTD005	DIF metadata standards shall convey all transformations (e.g. unit conversions, format conversions, sub-setting) that have occurred to data from the entry point to the DIF to the output/delivery to the data user.	H		
MTD010	The DIF shall ensure that the linkages between data and metadata are maintained with great reliability.	H		
MTD020	The DIF shall provide tools to enable end users and data providers to increase their capability in metadata generation and management.	H		
MTD025	The DIF shall establish a standard glossary for use by customers and data providers.	H		
MTD030	The DIF shall provide a central metadata registry.	M		
QC001	The DIF shall ensure that quality control operations are a partnership among data observation/collection components, processors, analysts, other users, and the DMAC.	H	Y	1. For DIF Phase 1 DIF relies on the QC processes and measures that have been implemented and utilized by data providers. 2. DIF does not impose new requirements apart from those included in the Data Content Standards.
QC002	The DIF shall identify, adopt, and adapt existing standards for quality control.	H		
QC005	The DIF shall provide a mechanism for ensuring that data are of known and documented quality.	H		
QC010	The DIF shall ensure that all DIF data products should include readily accessible documentation of sensor characteristics, processing, calibration, provenance, quality control, accuracy, precision, and other data and information required to satisfy user needs.	H		
QC020	The DIF shall transport QC and error characteristics, flags, through from data provider to data customer. **Gap:** will need to identify all error and QC data per integrated data set.	H		
QC030	The DIF shall accurately report lat/long and time of measurement.	H		

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
QC035	The DIF shall consistently flag missing data.	H		
QC040	The DIF shall provide evaluated QC'd Sea Level products and elevation contours from data sources. **GAP:** Identify data sources.	M		
QC050	The DIF shall implement data quality flags consistently across the data community.	H		
QC055	The DIF shall provide raw data with data quality flags rather than cleaned data. **GAP:** Identify which variables and data sources.	M		
QC060	The DIF shall deliver data of known quality, possibly using NWLON QC procedure, for all Sea Level data. **GAP:** Identify which providers specifically need improved QC.	H		
QC065	The DIF shall provide QC of Current data from shipboard ADCP profiles.	H		
DSA005	The DIF shall facilitate contribution of standards-based data to NODC archive from ecosystem community.	L - Not in DIF scope		No information on this one.
XPT001	The DIF shall support transport protocol(s) to allow machine-to-machine data transport.	H	Y	Required functionality is implied by design.
XPT005	The DIF shall support transport protocol(s) to allow data retrieval manually by a user.	M	Y	
XPT020	The DIF shall provide for the transport of metadata associated with all data transported.	H	Y	
XPT025	Data transport standards applicable to the DIF shall be identified and adapted as needed to support DIF functional requirements.	H	Y	
XPT040	The DIF shall provide automated delivery, “push”, of data to a customer model data input.	M	Y	No plans to implement, should be considered for removal from the list of requirements.
XPT045	The DIF shall allow users to manually select data for delivery.	M	Y	Required functionality is implied by design.

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
XPT050	The DIF data and metadata access functions shall be seamlessly integrated.	H	Y	
FILT001	The DIF shall allow for server-side subsetting of data for all relevant data.	H	Y	
FILT002	The DIF shall allow subsetting of data based on any field in the relevant data set.	H	Y	
SEC001	The DIF shall implement applicable NOAA requirements for IT security.	H		1. The distributed SOA accepted by DIF assumes that: a. general NOAA IT security guidance is applicable as DIF implements well known access services and transport protocols; b. no access at all to classified data will be provided; c. security measures are localized and implemented by all data providers according to the NOAA line offices IT security policies and procedures;
SEC005	The DIF shall ensure that interfaces to external systems are secure from unauthorized access.	H		
SEC010	The DIF shall separate user functionality (including user interface services) from information system management functionality.	H		
SEC015	The DIF shall isolate security functions from non-security functions.	H		
SEC020	The DIF shall protect the confidentiality of security-relevant system information and integrity of all transmitted information.	H		
SEC025	The DIF shall terminate network connections at the end of each session or after a configurable period of inactivity.	H		
SEC030	The DIF shall provide mechanisms to protect the authenticity of communications sessions.	H		
SEC035	The DIF shall incorporate antivirus protections with automatic updates.	H		
SEC040	The DIF shall check information inputs for accuracy, completeness, and validity.	H		
SEC045	The DIF shall authenticate that data sets received at client locations are from a valid DIF data source	H		

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
ADM001	The DIF shall provide a mechanism for continually monitoring results and publication of those results as a method for evaluating improvements in performance as a result of data integration.	H	Y	1. Each customer SOW describes expected benefits, and benchmarks and metrics to measure them. 2. One of the lessons already learned is that in many cases it is very difficult to do a quantitative analysis, and only qualitative one is possible.
ADM005	The DIF shall provide a mechanism for detection of status of DIF data provider interfaces and report the status externally.	H	?	Not clear what exactly this means.
ADM010	The DIF shall provide a mechanism for system extensibility.	H	Y	Required functionality is implied by design.
ADM015	The DIF shall establish and publicize policies for data availability.	H	?	No information available
ADM020	The DIF shall provide a mechanism for soliciting and responding to user feedback.	H	Y	
ADM025	The DIF shall establish and maintain international linkages.	H	Y	
ADM030	The DIF shall provide a registry or other means for determining what data are available within the DIF based upon queries that may be issued by others or by other machines.	M	Y	SOW has not been completed yet
RDS005	The DIF shall provide CTD offshore Salinity data from NDBC buoys.	M		
RDS010	The DIF shall provide C-MAN real-time Wind Speed and Direction data from NDBC.	M		
RDS011	The DIF shall provide NWLON real-time Wind Speed and Direction data from NDBC.	M		
RDS012	The DIF shall provide NDBC Buoy real-time Wind Speed and Direction data from NDBC.	M		

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
RDS020	The DIF shall provide near real-time SeaWiFS Ocean Color data from CoastWatch servers.	M		
RDS025	The DIF shall provide near real-time MODIS Aqua Ocean Color data from CoastWatch servers.	M	Y	
RDS045	The DIF shall provide estuarine Salinity data from NERRS.	M	Y	1. Web Service publicly accessible 2. Not all obs may be available (33 stations)
RDS050	The DIF shall provide estuarine Salinity data from the CO-OPS PORTS® data sets.	M	Y	Web Service publicly accessible
RDS055	The DIF shall provide estuarine Salinity data from the EPA National Estuary Program (NEP)	H		
RDS065	The DIF shall provide Forecasted Winds data from NWS regional marine text forecasts. GAP: Verify source provider of forecast data.	M		
RDS070	The DIF shall provide Cell Count data from various state partners including, but not limited to, Florida Fish and Wildlife Research Institute, Mote Marine Laboratory, Collier County Pollution Control and Prevention Department, Sarasota County Health Department, Alabama Department of Public Health.	L- Not core variable		
RDS075	The DIF shall provide Bathymetry data from the NGDC digital elevation model (DEM) Coastal Relief Model.	L- Not core variable		
RDS080	The DIF shall provide Bathymetry data from NOS Hydrographic Surveys.	L- Not core variable		
RDS090	The DIF shall provide USGS data to replace current web scraping. GAP: Determine variables, platforms, sources and formats.	L		
RDS095	The DIF shall provide offshore Currents and other data from existing AUVs. GAP: Verify data source provider.	L-data may not be readily available		

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
RDS105	The DIF shall provide Topography data from the USGS National Elevation Database.	L- Not core variable		
RDS110	The DIF shall provide Topography data LiDAR. GAP: Verify data source providers, formats.	L- Not core variable		
RDS120	The DIF shall provide subsurface Pressure data from available sources. GAP: Identify platforms, data source provider(s), formats.	L - Not core variable		
RDS125	The DIF shall provide sub-surface and near-shore Salinity data from regional observing systems. GAP: Verify data source providers, access methods, formats.	L	Y	Highly dependant on funds and human resources availability in FY09.
RDS130	Increase spatial coverage of current data, or access to in situ currents. GAP: Identify specific sources.	L		Exact meaning of the requirement is not clear.
RDS140	Obtain additional surface and subsurface Temperature data at various resolutions. GAP: Determine data availability and sources.	L		No certain project
RDS145	The DIF shall provide access to Temperature interpolations for Sea Surface and Subsurface Temperatures from models. GAP: identify models and source provider.	L		No information.
RDS150	The DIF shall provide access to historical data for all time series data provided through the GTS. GAP: Identify variables and observation platforms.	L	Y	SOW has not been completed yet
RDS155	The DIF shall provide Subsurface Temperature data to address the Gulf of Mexico data gaps. GAP: Identify observation platforms and data source providers.	L		
RDS160	The DIF shall provide Salinity data to address Gulf of Mexico data gaps. GAP: Identify observation platforms and data source providers.	H		
RDS165	The DIF shall provide Forecasted Winds from regional sources. GAP: Need to identify specific sources.	M	Y	Highly dependant on funds and human resources availability in FY09.

Require-ment ID	Requirement description	DIF Priority (H)igh, (M)edium, or (L)ow	Requirement addressed in DIF design	Comment
RDS170	The DIF shall provide post-storm High Water Mark data from USGS storm gauges.	L - Not core variable	Y	The requirement may be met as a part of the SLOSH Display program enhancement project
RDS175	The DIF shall provide the Ocean Heat content product produced at Tropical Prediction Center.	L		

8. LIST OF ACRONYMS

CI	Coastal Inundation
CONOPS	Concept of Operations
CO-OPS	Center for Operational Oceanographic Products and Services
CSC	Coastal Services Center
CSDL	Coast Survey Development Lab
CTD	Conductivity, Temperature, Depth
DIF	Data Integration Framework
DMAC	Data Management and Communications
DMIT	Data Management Integration Team
EPA	Environmental Protection Agency
ETSS	Extratropical Storm Surge
FGDC	Federal Geographic Data Committee
FIPS	Federal Information Processing Standards
FTP	File Transfer Protocol
GOM	Gulf of Mexico
GTS	WMO Global Telecommunications System
GTSPP	Global Temperature and Salinity Profile Program
HAB	Harmful Algal Bloom
HF	High frequency (radar)
HI	Hurricane Intensity
HTTP	Hyper Text Transfer Protocol
IEA	Integrated Ecosystem Assessments
IGOSS	Integrated Global Ocean Services System
IMS	Information Management System
IOOS	Integrated Ocean Observing System
IP	Internet Protocol
IT	Information Technology
MMS	Minerals Management Service
MODIS	Moderate Resolution Imaging Spectroradiometer
NCCOS	National Centers for Coastal Ocean Science
NDBC	National Data Buoy Center
NEP	National Estuary Program
NERRS	National Estuarine Research Reserve System
NESDIS	National Environmental Satellite, Data, and Information Service
NetCDF	Network Common Data Form
NGOM	Northern Gulf of Mexico
NODC	US National Oceanographic Data Center
NOS	National Ocean Service
NWLON	National Water Level Observation Network
NWS	National Weather Service
NWSTG	NWS Telecommunications Gateway
OPeNDAP	Open-source Project for a Network Data Access Protocol

PORTS	Physical Oceanographic Real-Time System
QC	Quality Control
RD	Reference Document
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
SLOSH	Sea, Lake, and Overland Surges from Hurricanes
TABS	Texas Automated Buoy System
TGLO	Texas General Land Office
URL	Uniform Resource Locator
USF	University of South Florida
USGS	United States Geological Survey
VPN	Virtual Private Network
WAN	Wide Area Network
WMO	World Meteorological Organization