NOAA

Integrated Ocean Observing System (IOOS) Program

Data Integration Framework (DIF)

Concept of Operations (CONOPS)

(Version 1.0)

April 25, 2008

DIF-0408-IOOS-CONOPS

Revisions

Version	Description of Version	Date Completed
Draft 0.1	Initial draft	11/12/2007
Draft 0.2	Second draft based on CONOPS Working Group review of initial draft.	12/03/2007
Draft 0.3	Draft based on IPT review.	01/17/2008
Draft 0.4	Final draft with all IPT review comments incorporated.	3/4/2008
1.0	Incorporation of editorial comments from IOOS program staff.	4/25/2008
1.1		
1.2		
1.3		
1.4		
1.5		

Review & Approval

Review History

Reviewer	Version Reviewed	Signature	Date
Christopher Moore/IOOS			

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background and Objectives

NOAA's Integrated Ocean Observing System (IOOS) Program has been charged with developing a Data Integration Framework (DIF) to meet the data collection needs of multiple NOAA ocean models, assessments, and decision-support tools.

The DIF project was proposed because there are no widely accepted and applied data format and transport standards to facilitate integration of data from diverse information and meet the geographic coverage, vertical and horizontal resolution, measurement accuracy, and timeliness requirements of the many NOAA decision-support systems. While there are numerous applicable standards, the challenge is identifying a limited set of standards with appropriate specificity to make data truly interoperable across multiple data sources. The DIF will identify a limited set of 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.

The project premise is that data integration[1], interoperability and improved access to and management of mission-critical ocean-related data, will increase the value and effectiveness of these data in supporting decision-making tools. The DIF will build a common data sharing infrastructure by adopting standards, best practices and other specifications of use for integration and transport of data to these decision-support tools and to other IOOS partners. The input parameters and customer models that have been selected are anticipated to provide a robust test of the potential of integrated data to improve models, assessments, and decision-support tools within NOAA.

Specifically, the objectives of the DIF are to:

1. Validate the premise that improved access to integrated data has value that can be measured. This premise will be tested using a minimum of five IOOS core ocean variables, from NOAA and non-NOAA sources, and four specific NOAA decision-support tools/models.

2. Develop a methodology based on IOOS Data Management and Communications (DMAC) (reference RD2) direction to improve upon existing ocean data integration efforts that 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 standards for data content, encoding, metadata, quality control, and transport.

4. Deploy these standards at selected data sources serving the four decision-support tools.

5. Maintain the DIF for a minimum period of three years, from project inception, to allow for adequate performance monitoring and assessment.

[1] Data integration refers to the process of combining data residing at different sources and providing the user with a unified access to these data. It involves the extraction, consolidation, and management of data from disparate systems to achieve broader capability by (functionally or technically) relating two or more data streams for the purposes of manipulation, analysis, and distribution. (http://en.wikipedia.org/wiki/Data_integration)

6. Provide a set of lessons learned, recommended standards and additional specificity or conventions for those standards to allow the longer-term strategic ocean data integration efforts to leverage the DIF experience for the benefit of NOAA and the Nation.

1.2 Purpose

The purpose of this document is to present the Concept of Operations (CONOPS) for the proposed DIF system. The CONOPS document is the translation of the functional requirements into operational scenarios and will provide important input into the high-level system design. The CONOPS will be used along with the Functional Requirements Document (RD5) to support the high-level design of the DIF and specific integration projects. The specific operational scenarios (referred to in this document as “Use Cases") and subcomponent functionality presented herein will be mapped to system-level requirements and high-level design.

This document presents a brief overview of the current state of operations regarding the proposed data providers and customers, a general description of the overall concept of operations, a description of the proposed system architecture from a functional standpoint, and several Use Cases describing how the DIF components will interact. This document also includes a requirements traceability matrix in Section 7 that links the requirements of the DIF Functional Requirements Document to the CONOPS Use Cases, as well as a summary of related operational, organizational, and transitional impacts of DIF implementation, and a synopsis of the proposed system.

The Use Cases are fundamental to the concept of operations and identify the operational scenarios involving DIF components, functional subcomponents, and system users. For each use case, there are step-by-step descriptions of how the proposed system will operate and interact with its users and external interfaces to perform required functions. The Use Cases are intended to describe how the basic functionality of the DIF will operate to support the identified functional requirements. They are presented in Section 5 as graphical depictions along with written descriptions using the DIF functional components and sub-components.

A single use case was developed for each major operational scenario; when the same set of high level functions are performed for several seemingly disparate activities, there is only one resultant use case. For example, if the functions required to support a request for a real-time data set are identical to the functions required to support a request for a delayed-mode dataset, only one use case is warranted. The requested data is different, but the functions required to request and receive the data are identical. Also, Use Cases are intended to be indifferent regarding data provider, user, data type, etc. The Use Cases are intended to capture the overarching, high-level operation of the DIF without presenting a separate use case for every possible permutation of actions available. However, some Use Cases are presented with real-world examples using representative customers, data providers, and data sets for demonstration purposes.

An important part of the DIF is the set of Use Cases associated with the “business processes” required to implement, operate, and update the framework. These processes, which consider the roles of offices, committees, and organizations on evaluation of systems, standards evolution, identification of new requirements, and operational deployment of enhancements, will help to maintain and evolve the DIF. Use cases related to these business processes will be addressed in subsequent versions of this document.

The Use Cases are presented using the following terminology:

1. **Goal:** The desired result of interaction with the DIF system.
2. **Use Case Diagram:** A graphical representation of the relationship between the system users and the various functional processes and/or services.
3. **Actors:** In the context of the Use Cases, actors are end-users and/or machine processes external to the DIF that interact with the DIF to initiate the events described in the Use Cases. This will mainly consist of end-users that request and receive integrated data sets and the initial data providers.

4. **Pre-existing Conditions:** Previous actions or conditions that must exist prior to executing the trigger event and intermediate steps.
5. **Trigger event:** The initiating action on the part of an actor that starts a given process.
6. **Intermediate steps:** A listing of actions performed by the DIF components to accomplish the goal.

2 Reference Documents

RD1 NOAA Integrated Ocean Observing System (IOOS) Program Office *Data Integration Framework (DIF) Master Project Plan*, Version 1.0, November 8, 2007

RD2 Data Management and Communications Plan for Research and Operational Integrated Ocean Observing Systems, The National Office for Integrated and Sustained Ocean Observations Ocean.US Publication No. 6, March 2005

RD3 IOOS Functional Requirements Baseline Document (produced by LMI), Draft, September 2007

RD4 DIF As-Is Baseline Systems Document (DRAFT Version 0.6) September 19, 2007

RD5 DIF Functional Requirements Document, Version 1.0, November 19, 2007

3 Current State of Operations

The section below provides a high-level description of how oceanographic data providers currently provide data to users, and how those users access the data.

3.1 Data Sources

Oceanographic data providers collect observational data from various platforms and sensors, assemble and process these data, and provide various output data products. Examples of data providers and some of their products include the National Data Buoy Center (NDBC) wind and current data, the Center for Operational Oceanographic Products and Services (CO-OPS) National Water Level Observation Network (NWLON) and the Physical Oceanographic Real-Time System (PORTS®), the National Environmental Satellite, Data, and Information Service (NESDIS) CoastWatch MODAS Aqua products for ocean color, and the Global Temperature and Salinity Profile Program (GTSPP) data products. Data sources may provide data in real-time, near real-time, delayed, and/or historical modes, as well as provide forecast data. Potential sources for historical data are NOAA's National Oceanographic Data Center (NODC) archive and the World Ocean Database. A likely source for real-time and forecast data for the DIF is the World Meteorological Organization's (WMO) Global Telecommunications System (GTS); the U.S. National Weather Service (NWS) serves as the GTS Regional Telecommunications Hub covering North America, Central America, and the Caribbean.

3.2 Data Access Methods

Data providers serve their data to users via numerous access methods. These methods include, but are not limited to:

- Web services, which allow user-specific customization of data requests. An example is the Open Geospatial Consortium (OGC) Web Coverage Service (WCS), Web Feature Service (WFS) and Sensor Observation Service (SOS).
- Web sites: using the Hypertext Transfer Protocol (HTTP), data providers often make their data available so that human users can browse available datasets and product services and download what is needed.
- FTP: A large amount of oceanographic data is available for download from servers via the File Transfer Protocol (FTP). FTP servers are broadly available and have been in use since well before HTTP and other Web Services.
- Automated email messages that alert users of new data.
- Really Simple Syndication (RSS): a family of Web feed formats used to publish frequently updated content. An RSS document (which is called a "feed" or "web feed"or "channel") contains either a summary of content from an associated web site or the full text.

3.3 Data Formats

Data formats, vocabularies, units, and other attributes vary widely across data providers. Interoperability tests conducted recently by the NOAA IOOS Program highlighted the disparity of formats and the difficulty in integrating data from multiple providers. Some common formats currently in use by the oceanographic community include netCDF-3, HDF4-SDS, and HDF5, comma delimited files, etc.

3.4 Customers

The DIF will initially focus on four models or and assessment/decision-support tools that ingest or otherwise use data from various data providers. Customers are defined as the NOAA offices and personnel operating and using the results of the models and assessment/decision-support tools. Each customer has a number of users, which refers to the individual persons accessing and retrieving data, performing any necessary formatting and/or data combination, and forwarding the data into the models and/or tools. The first four identified customers are the following:

- Harmful Algal Bloom Forecasting System (HAB-FS)
- Integrated Ecosystem Assessments (IEA)
- Coastal Inundation
- Hurricane Intensity

These individual customers currently each access their data in different ways, ranging from somewhat automated access to highly manual processes that require an operator to perform multiple functions.

Once all data is collected from the diverse sources, the data often require post-processing into a uniform format that can be used by the decision tool or model. Customers develop their own “normalization” routines for users to aggregate data into a usable format.

Once the initial DIF capability is achieved, and the performance impact on decision-support tools has been assessed, it is anticipated that the DIF will be expanded to include additional customers.

3.5 Current State of Interoperability

Tests conducted recently by the NOAA IOOS Program highlighted the disparity of formats and the difficulty users experience when using data from multiple providers. The tests documented inconsistencies in the application of standards and protocols across sources for selected elements of data management and communication (DMAC) and identified gaps in usage. While there is likely a significant portion of NOAA data that is already interoperable, these interoperability tests reinforced the need for more widespread integration, Some general findings common to the tests included:

- Some data are sufficiently interoperable within a given data provider/ source. However, compatibility is not extensible between sources, precluding direct integration (e.g. different data vocabularies and structures in use).
- There is a general lack of documentation provided by data providers on the standards and protocols being used (e.g. transport, vocabularies/ taxonomic conventions, data dictionaries)
- Integration of data is further hampered by an absence of metadata provided with the data. If metadata are available, they are often not available with the data themselves and are difficult to find.
- The degree of data compatibility can be seen in something as simple as the expression of time or place. For example, there is not a common standard vocabulary across providers to express timestamps or latitude/ longitude. While a single universal convention for time or latitude/longitude may not achieve interoperability, a small set of well documented conventions will allow users to more easily convert from one convention to another.

In general, the lack of tiered services, where implementation details at each data provider are hidden from the data user, hampers interoperability. There is a clear need for a more service-oriented architecture where services and data formats are standardized at the point of presentation to the data user.

4 General Concept of Operations

4.1 Overview

The DIF is envisioned to be a set of services and standards that will facilitate broader and more efficient transfer of data products from a varied set of data providers to a number of customers' models and analysis tools.

The DIF will allow the identified NOAA customers to more easily and efficiently access oceanographic and other meteorological data from the data providers for input into their modeling and analysis systems. The goal of the DIF is to provide each of the customers with a uniform set of methods to access the multiple data sets that they ingest, and to make these data sets available in standardized formats. Currently, these customers tend to access these data sets using a combination of specialized automation processes and through manual retrieval from various web sites, FTP servers, and real-time delivery mechanism. The DIF will allow customers to implement more homogenous, automated processes for retrieving and using the data. This will be achieved through the selection, specialization, application, and testing of data content, encoding, transport, and metadata (including quality) standards. These standards are intended to be implemented by DIF data providers to formalize a common data sharing infrastructure, and by customers to access the data from that infrastructure. It is anticipated that the DIF will be a distributed system with components implemented at data providers, software gateways, customer locations, or all of these. Although the core functionality of the DIF related to transformation, assembly, transport, and receipt of data will be distributed primarily among data providers and customers, there is also a need for an overarching set of functions, such as a central registry, to support the overall goals and operations of the DIF. Business processes will be put in place to monitor, evaluate, and update the DIF to ensure that it continues to meet its goals and serve the interests of the data providers and customers.

4.2 Major Components

The major DIF components include the following:

1. Data provider infrastructure – implementation projects at data providers to facilitate integration of existing data sets into the DIF common data sharing infrastructure.
2. Client infrastructure– integration of software applications and services, as needed, to allow the customers to access the DIF data in an efficient way.
3. Data Aggregation Services – implementation of systems/services that aggregate data, including metadata, from the common data sharing infrastructure to provide users a single point of access to multiple providers, or thematic aggregations of data. This may be done using data from DIF participating data providers as well as data from providers that do not offer data using DIF conventions. It is anticipated that data aggregation services will aggregate data either by variable, geographic region, event or other thematic classification. Data aggregation can be performed at data provider locations, at intermediate software gateways, as part of the DIF central management function, and/or within client side applications. This concept is discussed further in Section 5.1.
4. DIF Management Functions – Data and/or system management functions cutting across all data providers and customers. This could include a central data registry, central metadata management, system monitoring, and general system administration functions.
5. Communications – various public and private communications networks used to facilitate communication between the DIF components.
6. DIF Business Processes – organizational processes, roles, and responsibilities required to implement, operate, manage and update the DIF. These processes, which consider the roles and

responsibilities of offices, committees, and organizations on evaluation of systems, standards evolution, identification of new requirements, and operational deployment of DIF enhancements, will help to maintain and evolve the framework.

The DIF component architecture and underlying functional components are presented and described in Section 5 DIF Functional Architecture.

Initial DIF operations will consist of integration projects at selected data providers and customers. These initial projects will be customized to leverage existing infrastructure and capability. Based on initial success of a few integration projects, it is anticipated that generic reference implementations for both the customer and data provider sides will be developed to facilitate integration at additional sites. The intention of the reference implementations is to provide data providers and customers the tools and mechanisms to participate in the DIF common data sharing infrastructure with minimal technical investment. These tools and mechanisms may include, but not be limited to, data vocabularies, libraries, content standards, and application programming interface specifications and software tools.

Implementation of the DIF functionality at the data provider is not intended to change or interfere with how data providers currently fulfill their missions, but to augment their service offerings by transforming their data into DIF services, standards and conventions. It is anticipated that data providers will have varying levels of available resources to implement and operate the respective DIF functionality. Therefore, the reference implementations provided will be as complete as possible, including setup and configuration tools to assist with site-specific integration. Also, it is anticipated that a number of data providers will require additional resources and that the NOAA IOOS Program will work with individual data providers to provide assistance for implementation and operation of the DIF infrastructure. Similarly, the reference implementations for the client infrastructure will be as complete as possible, and the NOAA IOOS Program will work with customers to achieve integration and operations.

4.3 Categories of Data

The DIF functionality will handle three main categories of data: observational data, metadata, and registry information. In the DIF context, these components are defined as follows:

1. **Observational Data**: geophysical measurements that are to be integrated at data providers and received by the customers' decision-support tools. Grids, times series, and profiles are three common data structures (or feature types) of observations.
2. **Metadata**: information records, including discovery, quality, collection service and other information about the data, frequently separate from the data themselves, that convey the essential characteristics of a data source. As described by the Federal Geographic Data Committee (FGDC)[2], "A geospatial metadata record includes core library catalog elements such as Title, Abstract, and Publication Data; geographic elements such as Geographic Extent and Projection Information; and database elements such as Attribute Label Definitions and Attribute Domain Values."
3. **Registry Information**: the subset of metadata pertaining to discovery of, access to, and use of data. For example registries often contain routing information such as URLs, available data formats and transport conventions and information about the data provider. Registry information may also include other metadata elements such as schemas, dictionaries, references, points of contact, etc.

[2] The Federal Geographic Data Committee (FGDC) is an interagency committee that promotes the coordinated development, use, sharing, and dissemination of geospatial data on a national basis. The FGDC is tasked by Executive Order 12906 to develop procedures and assist in the implementation of a distributed discovery mechanism for national digital geospatial data. (www.fgdc.gov)

The DIF will convey quality control (QC) information both as part of the metadata and within the observational data itself. QC information may be generated at different points in the chain such as after the observation point but before any processing, pre- or post-assembly, or further downstream as is often the case with delayed mode data. Therefore, the DIF will need to implement a flexible, distributed metadata management function that will collect all the relevant QC information for a given data set, format it into recommended standards and conventions, and pass it through to the end-user. The DIF will not perform any quality control on observational data, rather it will implement appropriate metadata management functions based on existing community-approved approaches.

The DIF will also incorporate delivery validation functions for error detection and correction, such as transport layer checksum[3], cyclical redundancy check (CRC)[4], or other delivery assurance method, to ensure that data delivered to the end-user is the same as the data coming in to the DIF and is not corrupted by the delivery process.

The DIF will handle data regardless of the time the data were collected relative to the time a customer requires it. This means that real-time, near real-time, and historical data will all be available through the DIF given that the associated data provider, or a data aggregation or brokering service, makes the subject data available in accordance with DIF conventions.

Finally, the DIF will not dictate how a data provider manages their data holdings. Rather, the DIF will define a common data sharing infrastructure that defines how the data provider will serve the data to DIF customers.

---

3 - Checksum: A value used to ensure data are stored or transmitted without error. It is created by calculating the binary values in a block of data using some algorithm and storing the results with the data. When the data are retrieved from memory or received at the other end of a network, a new checksum is computed and matched against the existing checksum. A non-match indicates an error. (CMP's TechWeb Network)

4 - Cyclical Redundancy Checking: An error checking technique used to ensure the accuracy of transmitting digital data. The transmitted messages are divided into predetermined lengths which, used as dividends, are divided by a fixed divisor. The remainder of the calculation is appended onto and sent with the message. At the receiving end, the computer recalculates the remainder. If it does not match the transmitted remainder, an error is detected. (CMP's TechWeb Network)

5 DIF Functional Architecture

This section provides an overview of the proposed DIF functional architecture and supporting logical subsystems, descriptions of the required services and functions, and a listing of the internal and external interfaces required to support DIF operations. This section is intended to present the architecture of the DIF from a functional standpoint, without referring to specific software subsystems or applications. The functional components depicted and described below are organized to support the Use Cases and are not intended to map directly to a software and/or hardware system design. The intention is for the functional architecture description, the Use Cases, and the original functional requirements to be used as a foundation for execution of specific integration projects at data provider and customer locations.

The integration of the components will be achieved wherever possible by leveraging existing community standards and protocols for data content, encoding, metadata, transport, and registries. In each of these areas, existing standards will be evaluated for use in the DIF. In cases where an existing standard and/or convention is adopted, but does not provide the level of specificity required to achieve full integration, the NOAA IOOS Program will work with key stakeholders to define the additional conventions or specificity required.

Further, wherever possible the DIF will leverage existing services or implementations to avoid development of components that already exist in whole or in part. Prior to implementation of each DIF component, an evaluation will be performed to determine if existing services, systems, or facilities are available that can be leveraged to reduce implementation cost and schedule. In this way, the DIF can be viewed as a distributed systems integration project consisting of several integration projects at data providers and customers that, as a whole, comprise the DIF.

The overall functional architecture of the DIF common data sharing infrastructure is anticipated to be a distributed set of services, with components implemented at data providers, software gateways, customer locations, and possibly a central site to house overarching DIF management functions. Figure 1 on the next page shows the overall DIF functional architecture components and the conceptual boundary of DIF functionality in relationship to the DIF data providers, customers, and potential non-DIF data providers. Following the figure is a brief description of each of the major DIF components and the anticipated responsibility for each. Subsequent sections describe these functional components in further detail including their subcomponents.

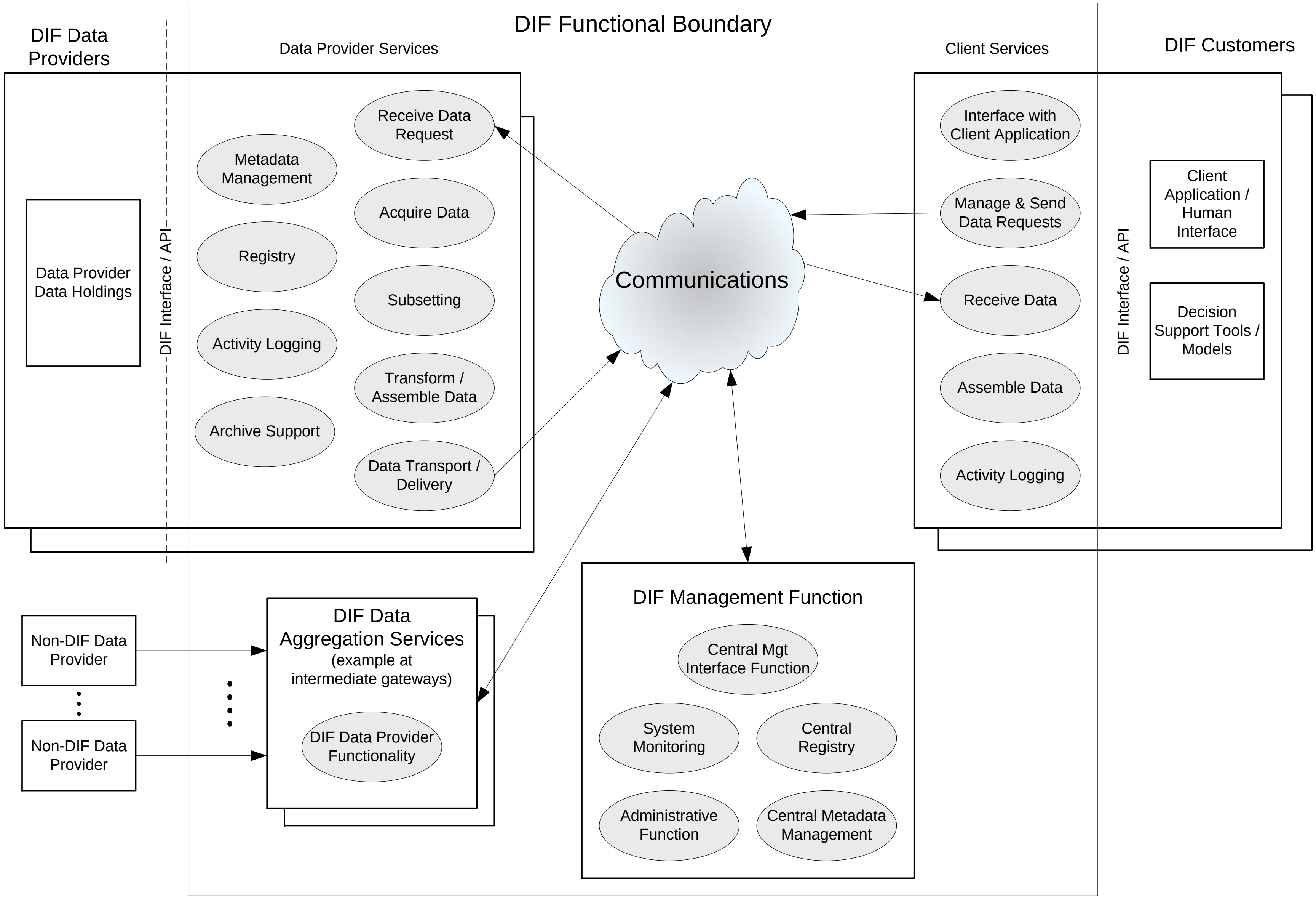

Figure 1 – Overall DIF Functional Architecture

The primary DIF components include the following:

1. **Data provider services** – Implemented at data providers and comprised of a number of functional components to support data providers' efforts to integrate their source data sets into standard DIF conventions and to offer the data for transport to end-users. Implementation and operation of any required services at the data provider locations will be the responsibility of the data provider themselves, with service requirements, standards, and other non-systems framework support provided by the NOAA IOOS Program.

2. **Client services** – Software applications and/or services and functional components that allow the customers to access data efficiently through the DIF, including the capability to search available data sets, view metadata, and retrieve data both manually and automatically. Implementation and operation of any required services at the client locations will be the responsibility of the client, with service requirements, standards, and other non-systems framework support provided by the NOAA IOOS Program.

3. **Data Aggregation Services** – Systems/services and functional components that aggregate data, including metadata, from the common data sharing infrastructure to provide users a single point of access. It is anticipated that data aggregation services will aggregate data by data type, variable, or community of interest. Data aggregation services may be performed at data provider locations, at intermediate software gateways, as part of the DIF central management function, and/or within client side applications. Aggregation services will typically retrieve and aggregate data from multiple DIF data providers, however, they may also be capable of retrieving data from providers that do not offer data in accordance with DIF conventions, performing DIF integration on these data, and including these data in the aggregation function. Responsibility for implementation and operation of data aggregation will vary based on which model is employed. In general, the responsibility for implementation and operation will lie with the organization providing the aggregation service. In some cases, the NOAA IOOS Program itself may implement and operate an aggregation service if there is a particular unmet need where the NOAA IOOS Program can add value.

4. **DIF Management Function** – Data and/or system management functions cutting across all data providers and customers. This will include, but not necessarily be limited to, functionality supporting a central data registry, central metadata management, system monitoring, and general system administration. The most likely scenario is that the NOAA IOOS Program would sponsor the implementation and operation of these over-arching management functions, although the systems could be co-located with one or more data provider facilities.

5. **Communications** – Various public and private communications networks currently used to exchange oceanographic data. In general, the Internet will be used as the primary mechanism for communications for delayed-mode data, although other direct or private network links can be employed as required by individual data providers and customers. Real-time data will more likely make use of dedicated private networks with service level guarantees. As NOAANet[5] matures it may become a valuable resource for these communications. Responsibility for the communications infrastructure will rest with the individual data providers and clients that require the connectivity.

---

[5] NOAAnet is an enterprise-wide wide area network based on Multi Protocol Label Switch System (MPLS) technology. NOAANet. NOAANet will initially be deployed to Weather Service locations, with subsequent expansion to other NOAA locations.

5.1 Component Architecture

This section describes the functional architecture of the four major components of the overall DIF architecture and lists the functional subcomponents required to support each. The functionality of each subcomponent is described in a corresponding subsection in Section 5.2.

5.1.1 Data Provider Services

The data provider services consist of a set of functional subcomponents that will link the data providers' existing systems/services with the DIF. It will provide the services and functionality required for end-users to access data according to DIF-adopted conventions.

DIF functionality at a given data provider is not intended to disrupt or change how the provider operates and offers data currently, it is intended to be implemented as an additional service layer.

The functional subcomponents and precise implementation approach will be determined by the data provider. It is envisioned that common toolkits, or reference implementations, will be developed and made available to new DIF data providers to ease implementation of services. Standards, application-specific profiles, and service definitions will define methods by which the data provider will serve data in a manner consistent with both the individual data provider’s IT security policies and the IT security requirements of NOAA.

The functional subcomponents of the data provider infrastructure comprise the following functions:

1. Receive Data Requests
2. Acquire Data
3. Subsetting
4. Transform/Assemble Data
5. Data Transport Delivery
6. Metadata
7. Registry
8. Activity Logging
9. Archive Support

These functional subcomponents will be implemented using software applications and services that will be defined in the subsequent system design phase.

5.1.2 Client Services

The client services consist of a set of functional subcomponents that would be integrated with the client’s decision-support tool or model. These services will interface with the clients' systems to facilitate browsing available data sets, viewing metadata, manually requesting data sets and metadata, and setting up automated delivery of DIF data into client tools.

It is envisioned that common toolkits, or reference implementations, will be developed and made available to DIF customers to ease implementation of services. Standards and service definitions will define methods by which the DIF infrastructure will provide data to the clients’ systems in a manner consistent with both the individual client’s IT security policies and the IT security requirements of NOAA.

The client service will be able to support end-user interfaces, as well as machine-machine interfaces which input data directly into clients' decision-support tools/models.

The functional subcomponents of the client infrastructure comprise the following functions:

1. Interface with Client Application
2. Manage and Send Data Requests
3. Receive Data
4. Assemble Data
5. Activity Logging

These functional subcomponents will be implemented using software applications and services that will be defined in the subsequent system design phase.

5.1.3 Data Aggregation Services

Data Aggregation Services are anticipated to consist of the basic DIF Data Provider Functionality in order to aggregate data from multiple data providers for delivery to DIF customers. There will likely be no physical restrictions on aggregation services as they may be performed at data provider locations, at intermediate software gateways, within client-side applications, or even within the DIF Management Function. Figure 1 represents the situation where the aggregation services reside at intermediate software gateways that may interface with DIF data providers (through the communications function) and non-DIF data providers to retrieve and integrate data.

Aggregation services will include two functions based on whether the data are retrieved from participating DIF data providers or from data providers who are unwilling or unable to provide data in accordance with DIF conventions. In the first case, a user has requested data that originates from more than one DIF data provider and requires a single collection point for the data. This requires an aggregation service to retrieve and assemble the data from different providers into a single product for delivery to the end-user. In the second case, the user has requested data from one or more sources that do not serve data according to DIF conventions. This requires an aggregation service to retrieve the requested data from the distributed sources, transform them and their metadata into DIF conventions, and perform any necessary assembly of multiple data sets for delivery to the end-user. Either approach may incorporate filtering or subsetting with respect to data type, geographical range, time period, etc.

With aggregation services distributed among the various components of the DIF, certain requests for data from a user may end up being fulfilled through more than one stage of aggregation. An example is when a user concurrently requests data from data sources that are both DIF compliant and non-DIF compliant. In such a case, the data from the DIF data providers may be aggregated at an intermediate gateway service or forwarded directly to the client-side application, and data from non-DIF providers may be aggregated at an intermediate gateway or the DIF Central Management function, and then forwarded to the client-side application, which would then perform aggregation upon all the data received.

Although the aggregation services were not intended to be data management centers with high-volume storage capabilities, the technical design of the aggregation services may include the capability to support various levels of storage of aggregated data sets.

5.1.4 DIF Management Function

The DIF Management Function consists of a set of functional subcomponents to perform a central system management role for the various distributed DIF components. The actual implementation may be a distributed set of services, or it may be consolidated in a central location, however for the purposes of the functional architecture description it will be treated as though it is centralized. Even if it is distributed, its functionality and user interaction with it will likely appear centralized. Portions of the DIF Management Function may be hosted on stand-alone systems at one or more primary data provider centers (ie, a particular data provider could host the central registry portion), and/or a separate central location where all management functions are consolidated.

Existing services and standards will be used wherever possible in the DIF Management Function. For example, the Central Registry will be based on an existing registry service definition (such as OGC's Catalogue Services for the web specification, or CS-W, or the FGDC clearinghouse)

The functional subcomponents of the DIF Management Function comprise the following functions:

1. Central Management Interface
2. System Monitoring
3. Central Registry
4. Central Metadata Management
5. Administrative Functions

These functional subcomponents will be implemented using software applications and services that will be defined in the subsequent system design phase.

5.2 Subcomponent Functionality

This section describes the individual subcomponent functionality for each entity within the four major components of the DIF. This subcomponent functionality is designed to directly support the functional requirements as implemented by the Use Cases presented in Section 6. Again, these are not intended to map directly to application or services modules since a single functional subcomponent may be implemented using more than one application or service module and conversely, a single application or service module may implement more than one functional subcomponent.

5.2.1 Data Provider Infrastructure Subcomponent Functionality

5.2.1.1 Receive Data Requests

The data provider's Receive Data Request function accepts incoming requests for data including requests for observational data, metadata, and registry information. Requests handled can be from manual processes initiated through client services and can be automated from either clients or the DIF management function. Requests for observational data from DIF customers constitute data pull actions and resulting data provider actions are considered the response to pull requests. The Receive Data Request function validates the data request with respect to format and content, and if the request is for restricted data, it verifies authentication information within the request. Once validated, the function passes the request to the appropriate functional entity for fulfillment, such as to the Acquire Data, Subsetting, Transform/Assemble Data, Data Transport/Delivery, Metadata, and Registry functions. The Receive Data Request may also log the request with the Activity Logging function.

5.2.1.2 Acquire Data

The Acquire Data function receives a request from the Receive Data Request function for observational data and analyzes the contents of the request for the identification and location of the requested data and for subsetting information, such as temporal and geographical limits. The function then uses this information to call up the requested data sets from either data provider storage, archives, etc. If subsetting information is included, the function passes this information to the Subsetting function. The Acquire Data function then forwards relevant request parameters regarding formatting and assembly requirements to the Transform/Assemble Data function and forward routing and transport requirements to the Data Transport/Delivery function. The Acquire Data function then manages the transfer of the requested data set(s) from the appropriate data provider storage location(s) to the Subsetting Function, if required, or to the Data Transport/Delivery function.

5.2.1.3 Subsetting

The Subsetting function receives the subsetting information from the Acquire Data function when subsetting is required and buffers observational data transferred from data provider storage by the Acquire Data function. The Subsetting function then applies the subsetting criteria from the user request to the observational data creating a new data set that it then forwards to the Transform/Assemble Data function.

5.2.1.4 Transform/Assemble Data

The Transform/Assemble Data function receives information from the Acquire Data function regarding data format and assembly requirements of the end-user. The function also receives observational data transferred from data provider storage by the Acquire Data function and/or from the Subsetting function. The function then acts on the observational data in accordance with the request requirements, such as formatting changes and assembling data from multiple internal sources into integrated sets. The function also retrieves required metadata for the observational data and includes this in the integrated sets in accordance with DIF conventions. The integrated data product(s) is then forwarded to the Data Transport and Delivery function.

5.2.1.5 Data Transport and Delivery

The Data Transport and Delivery receives formatted and assembled data and metadata, including required QC data, from the Transform/Assemble Data function along with transport protocol and routing information from the Receive Data Request function. The function then applies a checksum, CRC, or other delivery assurance method, formats the data into the required transport protocol, opens a session with the customer Receive Data Function, delivers data through the Communications Layer, and forwards transaction information to the data provider’s Activity Logging function.

5.2.1.6 Metadata

The Metadata function receives metadata retrieval requests from the Transform/Assembly Data function in conjunction with a request from an end-user for either metadata by itself or to accompany a request for observational data, then locates required metadata, and returns this to the assembly function. The Metadata function also responds to requests from the DIF Central Metadata Management function and returns metadata definitions for repository within the DIF Management Function. The Metadata function may also initiate a push of metadata definitions to the Central Metadata Management function when updated metadata information becomes available at the data provider.

5.2.1.7 Registry

The Registry function collects and assembles discovery metadata into a single registry directory. This information may include, but not be limited to, basic descriptive information about available data sets, data set names, data format, transport and access methods, and URLs and/or IP routing as well as possibly geographic coverage information to assist users in determining the applicability of a particular data set. The Registry function also responds to requests from the DIF Central Registry function and returns discovery metadata for repository within the DIF Central Registry. The local Registry function may also initiate a push of discovery metadata to the Central Registry function when updated metadata becomes available. The DIF registry function will be based on an existing registry standard or specification.

5.2.1.8 Activity Logging

The Activity Logging function will interface with the other functional subcomponents to collect and store records of selected DIF transactions. Which transactions are stored in the log will be determined as part of specific integration projects. The function may be implemented to push log information to the DIF Management Function and/or to respond to requests from the DIF Management Function for download of log information.

5.2.1.9 Archive Support

The Archive Support function will provide the capability for the transmission of DIF data sets to archive centers. This may occur automatically as a pushed service or in response to pull requests from the archive centers.

5.2.2 Client Subcomponent Functionality

5.2.2.1 Interface with Client Application

The Interface with Client Application function will provide the basic interface functionality required between the DIF common data sharing infrastructure and the client application. The client application may be one of a number of applications such as browsers, GIS applications, or custom applications. The interface function will manage the communications between the DIF and the client, to support end-user interaction and delivery of observational data and metadata. The interface function will support administrative functions such as user authentication, serving of interaction functions to support such activities as manual browsing and request of data, and will manage the delivery of returned observational data and metadata into the client application for display, storage, and/or delivery into the front end of the customers’ decision-support tools / models. The interface function will interact with the other functional subcomponents of the client infrastructure.

5.2.2.2 Manage and Send Data Requests

The Manage and Send Data Requests function will interact with the Client Application Interface to manage and process both manual and automated requests for observational data and metadata from the DIF. The function will validate the format and content of requests based on request type, and forward these requests through the communications layer to the appropriate destination within the DIF functional infrastructure. The function will also manage automated data requests by storing and periodically processing requests set up by end-users, effectively brokering the automated pull process for retrieval of observational data. The function will also be able to accept and deliver automated data requests originating from the external client application representing pull requests from outside the DIF interface boundary.

5.2.2.3 Receive Data

The Receive Data function will respond to requests from data providers' data delivery function to open communications sessions to receive requested observational data and metadata. The function may include a buffering capability, and will serve to forward requested data to the Interface with Client Application function for delivery to external end-users. If compatible data sets requiring assembly are received from multiple DIF data providers, the Receive Data function will buffer these as required and forward them to the Assemble Data function.

5.2.2.4 Assemble Data

The Assemble Data function may receive instruction from the Interface with Client Application function that returned data sets collected from multiple data providers need to be assembled, or integrated, prior to delivery to client application. In this case, the function will receive Data forwarded from the Receive Data function and will assemble the data sets into the desired end product. To the extent feasible during automated assembly, the function will analyze data sets for duplicate records and will discard any detected duplicate records in the final data set.

5.2.2.5 Activity Logging

The Activity Logging function will interface with the other functional subcomponents to collect and store records of selected DIF transactions. Which transactions are stored in the log will be determined during detailed system design. The function may be implemented to push log information to the DIF Management Function and/or to respond to requests from the DIF Management Function for download of log information.

5.2.3 Data Aggregation Services Functionality

DIF Data Aggregation Services will respond to requests to integrate data from multiple data sources when the capability to do so for the specific data sets requested does not reside elsewhere within the DIF functionality. Aggregation services will aggregate data on demand in response to specific requests from end-users and perform retrieval, transformation, assembly, and transport of observational data and/or metadata between multiple data providers and end-users. Data may be from one or more DIF data providers and/or non-DIF data providers.

5.2.4 The DIF Management Functionality

5.2.4.1 Central Management Interface

The Central Management Interface function will be a communications service that manages communications between the functional subcomponents of the DIF Management Function and the rest of the distributed DIF infrastructure located at data providers, customer sites, and aggregation services gateways. The function will perform routing and IT security functions as required to manage these communications services and will provide an end-user interface for administrative activities. The function will also respond to download requests for files supporting the data provider and client reference implementations, including, but not limited to, software applications and toolkits, data libraries and vocabularies, specifications, and associated documentation.

5.2.4.2 System Monitoring

The System Monitoring function will provide basic operational monitoring of all DIF components, such as data provider services. The monitoring function will periodically poll communications services and DIF application components to determine if services are currently available or not, and will maintain up-to-date status on each component. Status information can be broadcast to parties that have registered interest, or authorized users can access the DIF management interface to view the status of systems and services throughout the DIF infrastructure. At least for the initial DIF implementation, system monitoring is intended to simply monitor and report the availability of DIF services and is not to be a comprehensive performance or traffic monitoring system.

5.2.4.3 Central Registry

The Central Registry function serves as a comprehensive directory of registry information about all data sets available through the DIF common data sharing infrastructure. This information may include discovery metadata such as data set names, data format, transport and access methods, and URLs and/or IP routing. The function may periodically poll DIF data providers' local Registry functions for metadata updates (pull) and will receive unsolicited registry information updates pushed from the data providers.

The function will respond to client-initiated requests for registry data and return this information to the clients.

5.2.4.4 Central Metadata Management

The Central Metadata Management function serves as a comprehensive directory of metadata definitions for all data sets available through the DIF infrastructure. The function may periodically poll DIF data providers' local Metadata Management functions for updates (pull) and will receive unsolicited metadata updates pushed from the data providers. The function will respond to client-initiated requests for metadata and return this information to the clients.

5.2.4.5 Administrative Function

The DIF central Administrative Function will implement typical overarching administrative functions that may include, but not be limited to, management of user accounts, IT security functions, activity logging, system reporting, and system maintenance. The function will also manage the DIF infrastructure configuration by keeping a full record of system-wide infrastructure components and by tracking and recording instances of infrastructure components and services at newly established data providers, clients, and aggregation services. The function will send notification to system users of changes within the DIF infrastructure.

5.3 Internal and External Interfaces

Internal DIF interfaces will likely be internal application interfaces and local network interfaces within a particular data provider infrastructure, client infrastructure, DIF management, or aggregation services installation. These distributed components of the DIF will likely interface through wide area networks (WANs) implemented using public and/or private networks, such as the Internet or NOAANet, and may be implemented using secure communications channels such as those provided by virtual private network (VPN) services.

External DIF interfaces are anticipated to consist of the application programming interfaces (API) implemented between the reference implementations, and the respective DIF data providers and DIF customers. Additional interfaces may exist between non-DIF data providers and DIF Aggregation Services and would be addressed on a case-by-case basis.

6 Operational Scenarios (Use Cases)

Execution of the Use Cases presented in this section is supported by the DIF functional components presented in Section 5 DIF Functional Architecture. The functional components are depicted in the use case diagrams by ovals. When reading the Use Cases, Section 5 should be referenced when further explanation of a functional component is required.

Overall, the Use Cases are intended to be independent of data type, the temporal nature of the data, the data provider, and the end-user. This is due to the fundamental nature of the DIF concept, which is a standards-based framework that should support any data sets as long as the data provider and data customer are following agreed-upon DIF conventions. Further, each use case is intended to be generic in nature and can be applied to specific cases of data set, data provider, and end-user. The DIF conventions effectively harmonize the inputs and outputs of the system and allow a single set of processes and/or services to fulfill the needs of multiple end-users and data providers. Although the Use Cases are presented in this somewhat generic manner, a few examples of how a use case will apply to specific users are included for demonstration.

Additionally, the Use Cases incorporate the functional requirements in a form that supports an arbitrary number of data providers and data consumers. They are not intended to capture broader “business process” scenarios of the DIF such as incorporation of new standards or data providers into the framework. While it is important to model these business processes, it is not addressed in this version of the document.

Each use case includes a listing of pre-existing conditions that must be in place to successfully complete the scenario. Even though not expressly included as a pre-existing condition, it is assumed throughout the Use Cases that the required DIF software and services have been correctly installed and configured for each participating customer, data provider, aggregation service, and central management function.

6.1 Use Cases for End-users

These Use Cases describe scenarios in which the end-user is a human data consumer of a particular customer's model, assessment tool, and/or decision-support tool. Throughout these Use Cases, the end-user is assumed to be accessing DIF functionality through their own client application that resides outside the DIF functional boundary. This application interacts with the DIF through standardized services. Therefore, since the client application is the entity that directly interfaces with the DIF, the client application is considered to be the actor in each of these Use Cases while supporting the respective end-user.

Example: In the case of the HAB-FS decision-support tool, the client application is the HABs Bulletin generation software, which will ultimately ingest the integrated DIF data.

6.1.1 View Available Data (Discovery Metadata)

1. **Goal:** To interactively view the inventory of data available through the Central Registry resident in the DIF Central Management Function. This scenario is not intended to view the data themselves, but to view discovery metadata for available DIF data sets.
2. **Use Case Diagram:**

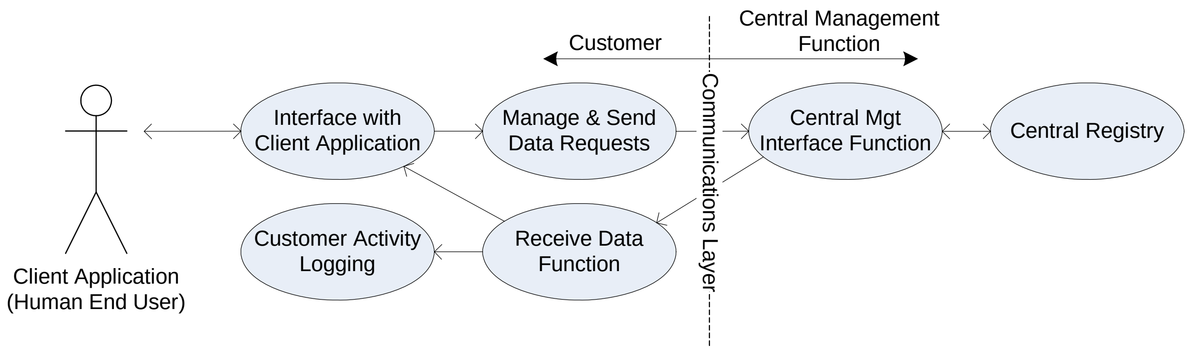

Figure 2 – Use Case: Search Available Data through a Central Registry

3. **Actors:** Client application supporting end-user
4. **Pre-existing Conditions:**
   a. End-user has accessed DIF functionality through local Client Application and requested to view registry information
   b. DIF system has assembled a registry of information (discovery metadata) regarding the latest available data listings into the Central Registry.
5. **Trigger event(s):** Client Application sends request for listing of available data
6. **Intermediate steps:**
   a. DIF Interface with Client Application accepts request and passes it to the Manage and Send Data Requests function.
   b. Manage and Send Data Requests function determines that request will be fulfilled by accessing the DIF Central Registry
   c. Manage and Send Data Requests function opens a session with the DIF Central Management Interface Function and relays request
   d. DIF Central Management Interface Function passes the request to the Central Registry, which returns results based on request parameters
   e. DIF Central Management Interface Function passes the registry information back to the customer-side Receive Data Function
   f. Receive Data Function receives registry information or data set information, passes data to Client Application Interface, and logs activity. NOTE: The Receive Data Function handles request for all types of DIF data, i.e., Observational Data, Metadata and Registry Information.
   g. Client Application Interface passes registry information to client application for display

6.1.2 Manually Request Observational Data (Pull)

1. **Goal:** Receive selected observational data from one or more DIF data providers into the client application and/or decision-support tools as a result of a manual request, or pull, by an end-user.

   Example: For the HAB-FS decision-support tool, this use case describes the request for surface currents observational data, with temporal and/or geographic subsetting, from NDBC and CO-OPS.

2. **Use Case Diagram:**

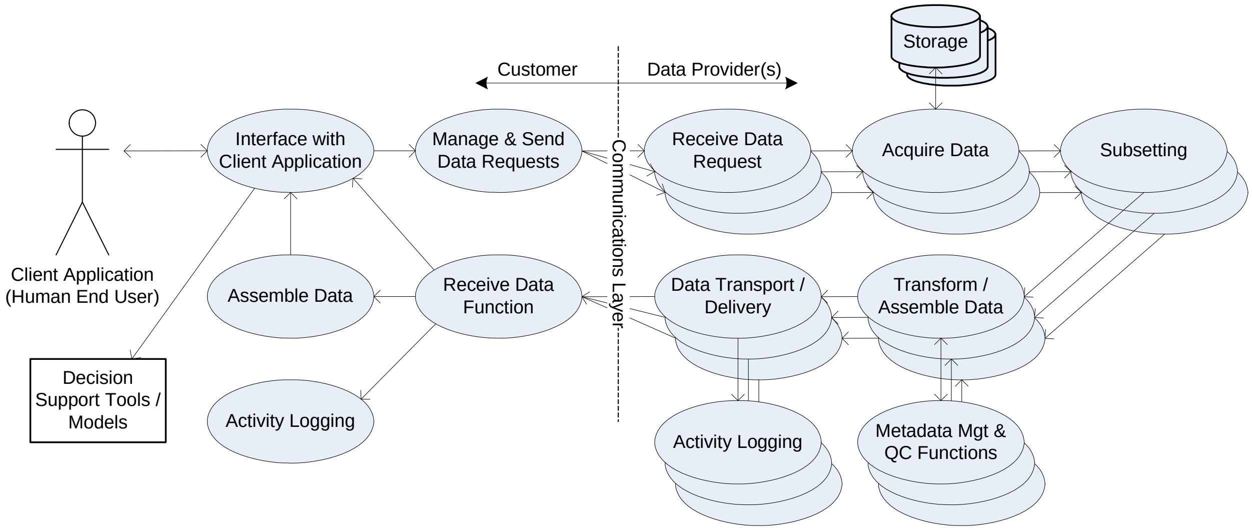

Figure 3 – Use Case: Manually Request Data

3. **Actors:** Client Application, such as a browser or GIS application, supporting end-users

4. **Pre-existing Conditions:**

   a. End-user has accessed DIF functionality through local Client Application and has selected one or more data sets for retrieval.

   b. DIF system has assembled information regarding the latest available data into the Central Registry.

   c. Data provider(s) or DIF functionality at data provider(s) is capable of serving data in accordance with DIF common data sharing infrastructure standards.

5. **Trigger event:** Client Application sends a manual request for data to the DIF Client Application Interface

   Example: For the HAB-FS decision-support tool, the trigger event is the end-user executing a script designed to pull the currents data.

6. **Intermediate steps:**

   a. DIF Client Application Interface displays the observational data that was the subject of the manual request.

   b. User selects data set(s) for download.

   c. DIF Client Application Interface retrieves registry information about the selected data set(s) from Central Registry using mechanisms described in 6.1.1

d. DIF Client Application Interface displays options available such as geographical coverage area, data source provider(s), data temporal frequency, available spatial resolution, time period of available data, etc.

e. User selects desired bounding conditions for data, i.e., subsetting options.

f. User selects whether data should be returned to the client application for review and manual routing, sent to the input function of the customer's decision-support tools/models, or to both.

g. DIF Client Application interface forwards the request parameters to the Manage and Send Data Requests function, which formats data request with bounding conditions and transmits request to the appropriate data provider(s) in accordance with addressing/routing information received as part of the registry information.

h. DIF data provider Receive Data Request function accepts and validates the data request and forwards the request to the Acquire Data function and sends relevant request information regarding processing and routing to the Subsetting, Transform/Assemble Data, and Data Transport/Delivery functions.

i. The Acquire Data function accesses each data provider's data holdings to retrieve the requested data.

j. Requested data are retrieved from data provider(s) storage and is subsetted by the data provider(s), if required, according to the user's criteria.

k. The data provider(s) then performs any necessary transformation and assembly for delivery, including retrieval and appending of any associated metadata to be included with the data product delivery, and removal of duplicate records to the extent feasible by the automated DIF services.

l. The data package is forwarded to the data transport and delivery function, which calculates a checksum, cyclical redundancy check (CRC), or other delivery assurance method, formats the data into the required transport protocol, opens a session with the customer Receive Data Function, delivers data through the Communications Layer, and forwards transaction information to the data provider's Activity Logging function.

m. DIF Receive Data Function validates delivery of data using chosen delivery assurance method, performs any necessary transport-level extraction for delivery to customer, logs activity, and forwards the data to the Client Application Interface if data is already compiled in a single data set, or forwards multiple data sets to the customer-side Assemble Data function in the case in which data is received from more than one data provider.

n. If required by having received like data sets from more than one source, the Assemble Data function buffers the multiple data sets and combines the data into a single product.

   Example: For the HAB-FS decision-support tool, this is the point where the ocean currents data from NDBC would be combined with the ocean currents data from CO-OPS.

o. To the extent feasible by the automated DIF services, the Assemble Data function then removes any detected duplicate records and then forwards the data for delivery through the Client Application Interface

p. DIF Client Application Interface forwards requested data into Client Application and/or to the front end of the customer's Decision-support Tools / Models in accordance with user selection of data destination(s)

6.1.3 Request Complete Metadata

1. **Goal:** Request and view complete metadata in Client Application for selected data sets
2. **Use Case Diagram:**

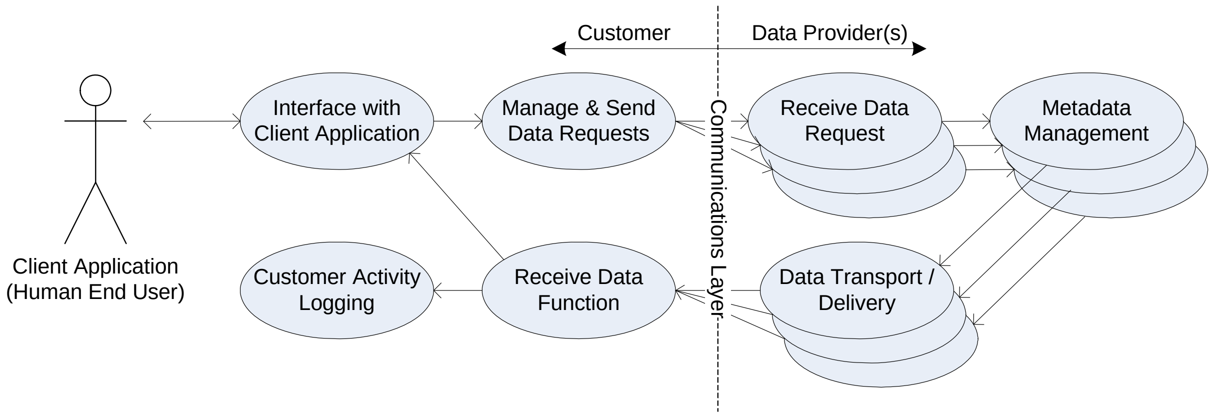

Figure 4 – Use Case: Request Metadata

3. **Actors:** Client application supporting end-user
4. **Pre-existing Conditions:**
   a. End-user has accessed DIF functionality through local Client Application and requested option to view metadata for selected data
   b. Metadata registry information has been compiled within the DIF Central Registry
5. **Trigger event:** Client Application sends request to DIF Client Application Interface to return metadata
6. **Intermediate steps:**
   a. DIF Client Application Interface passes metadata request to the Manage and Send Data Requests function
   b. Manage and Send Data Requests function opens session with DIF Central Management Function that retrieves addressing/routing information for selected data from the Central Registry
   c. Manage and Send Data Requests function opens necessary sessions with DIF data providers and relays requests using addressing/routing information from the Central Registry
   d. DIF data provider(s) Receive Data Request function accesses local Metadata function and returns results based on request parameters. NOTE: The Receive Data Request function handles request for all types of DIF data, i.e., Observational Data, Metadata and Registry Information.
   e. Data provider(s) Data Transport / Delivery function sends resulting information back to client Receive Data function
   f. Receive Data function receives metadata, passes it to the Client Application Interface, and logs activity

g. Client Application Interface passes metadata to Client Application for display

6.1.4 Configure Automated Data Request (Pull)

1. **Goal:** Create or manage custom data requests that will be fulfilled via automatic schedules, i.e., a scheduled, customer initiated, pull request

   Example: This could be used by the analysts for the HABs Bulletin to set up an automated, periodic pull of data and pre-defined times or intervals. That way, data would be available locally and there would be no need for individual manual requests prior to generation of each bulletin.

2. **Use Case Diagram:**

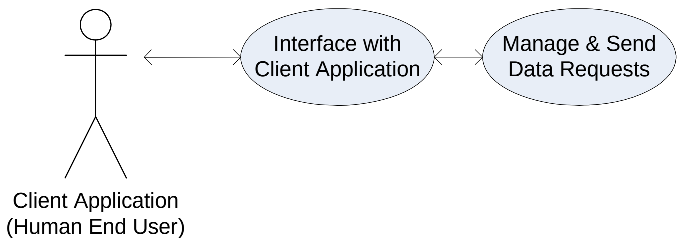

Figure 5 – Use Case: Configure Automated Data Request (Pull)

3. **Actors:** Client Application supporting end-users. For the purposes of this use case, the Client Application could be the decision-support tools or models themselves that have an interface configured with the DIF Client Application Interface.

4. **Specific Pre-existing Conditions:**

   a. End-user has accessed DIF functionality through local Client Application and has selected option to configure automated data requests.

   b. DIF Central Management Function has assembled registry information regarding the latest available data in the Central Registry.

5. **Trigger event:** Client Application sends request to configure automated data requests to the DIF Client Application Interface

6. **Intermediate steps:**

   a. DIF Client Application Interface responds to request to display the automated data delivery request function.

   b. User selects data set(s) for inclusion in new automated delivery request(s) or modifies existing automated request listings.

   c. DIF Client Application Interface retrieves information about the selected data set(s) from the Central Registry using mechanisms described in 6.1.1.

   d. DIF Client Application Interface displays options available such as geographical coverage area, data source provider, data temporal frequency, available spatial resolution, time period of available data, etc.

   e. User selects desired bounding conditions, i.e., subsetting options, for requested data sets.

   f. User selects frequency / schedule for data delivery.

   g. User selects destination for delivered data products.

h. User saves automated delivery request.

i. DIF Client Application Interface forwards parameters to the Manage and Send Data Requests function, which formats the data request per DIF conventions.

j. DIF Manage and Send Data Requests function returns the formatted request through the Client Application Interface to the customer's Client Application, which then enters the request(s) into an appropriate local queue for automated processing.

6.2 Use Cases for Machine End-user

These Use Cases describe scenarios in which the end-user is a computer/machine, operator (human) of a particular customer's model, assessment tool, and/or decision-support tool. Throughout these Use Cases, the machine end-user is assumed to be accessing DIF functionality through an automated client application that resides outside the DIF functional boundary and interacts with the DIF through standardized services. Therefore, since the automated client application is the entity that directly interfaces with the DIF, the automated client application is considered to be the actor in each of these Use Cases while supporting the respective machine end-user, and ultimately the human customer.

6.2.1 Automatically Request Data (Pull)

1. **Goal:** Automatically deliver selected data from DIF data providers into the customer decision-support tool / model using locally stored and processed data pull requests.

2. **Use Case Diagram:**

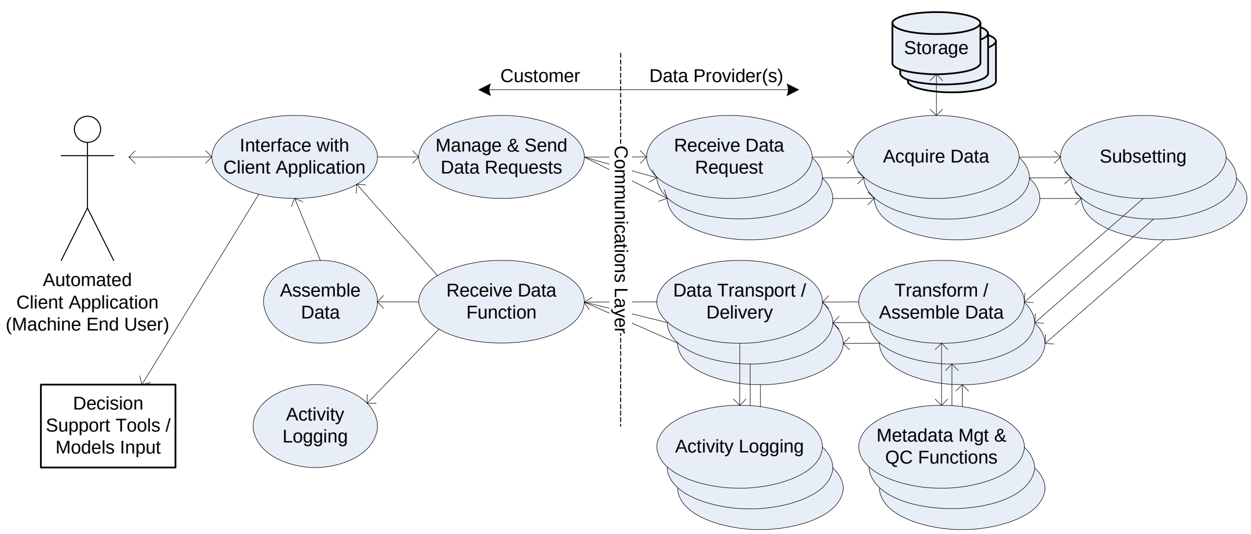

Figure 6 – Use Case: Automatically Request Data (Pull)

3. **Actors:** Automated Client Application computer/machine) that operates between the identified decision-support tools and the DIF Client Application Interface

4. **Pre-existing Conditions:**

   a. End-user has configured one or more automated data requests

   b. One or more automated data requests are “active” in the automated client application.

c. DIF system has assembled registry information regarding the latest available data into the Central Registry.

5. **Trigger event:** Automated Client Application processes automated data request active in queue

6. **Intermediate steps:**

   a. End-user machine initiates an automated data request session with the DIF Client Application Interface and forwards parameters of the automated data request.

   b. DIF Client Application Interface forwards the request parameters to the Manage and Send Data Requests function, which transmits request to the appropriate data provider(s) using addressing/routing information from the Central Registry.

   c. DIF data provider(s) accepts and validates the data request and the Acquire Data function accesses data provider(s) storage to retrieve the requested data.

   d. Requested data are retrieved from data provider(s) storage and subsetted, if required, according to the automated request criteria.

   e. Subsetted data then undergo any necessary transformation and assembly for delivery, including retrieval and appending any associated metadata to be included with the data product delivery, and removal of duplicate records to the extent feasible by the automated DIF services.

   f. The data package is forwarded to the data transport and delivery function, which applies a checksum, CRC, or other delivery assurance method, formats the data into the required transport protocol, opens a session with the customer receive data function, delivers data through the communications layer, and forwards transaction information to the data provider’s activity logging function.

   g. DIF Receive Data Function validates delivery of data using chosen delivery assurance method, performs any necessary transport-level extraction for delivery to customer, logs activity, and forwards the data to the Client Application Interface if data is already compiled in a single data set, or forwards multiple data sets to the data assembly function.

   h. If required, the data assembly function buffers multiple data sets and combines them into a single product for delivery through the Client Application Interface.

   i. DIF Client Application Interface forwards requested data into Automated Client Application and/or to the front end of the customer’s decision-support Tools / Models in accordance with the automated request’s designation of data destination(s)

6.3 Use Cases for Data Providers

These Use Cases describe scenarios in which the primary actor is a DIF data provider, typically a machine. In general, machine end-users act to support their own human end-users operating the identified data centers. For our purposes, these human end-users are outside the scope of these DIF Use Cases since the subject machine at the data provider(s) will be serving the role of the actor that initiates each process.

6.3.1 Automatically Deliver Data (Push)

1. **Goal:** Automatically deliver newly available observational data that is designated for push to one or more end-users.
2. **Use Case Diagram:**

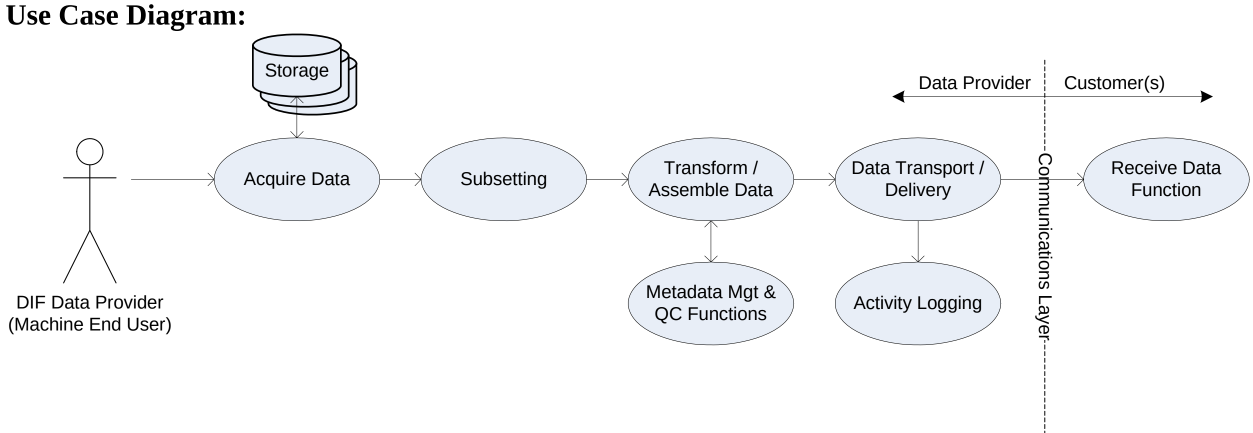

Figure 7 – Use Case: Automatically Deliver Data (Push)

3. **Actors:** Automated DIF data provider application managing pushed data services and requests.
4. **Pre-existing Conditions:**
    a. DIF data provider is capable of and configured to provide pushed services for observational data into or through the DIF Data Provider Infrastructure.
    b. One or more DIF customers have registered with the data provider for available pushed services through the DIF.
    c. This data provider has accepted and logged the customer's request.
5. **Trigger event:** Updated observational data has been made available in a data set that is included for push services through the DIF.
6. **Intermediate steps:**
    a. DIF data provider forwards request parameters for data delivery to the Acquire Data function.
    b. Acquire Data function accesses data provider storage to retrieve the required data
    c. Required data are retrieved from data provider storage and subsetted, if required, according to the push request criteria.
    d. Subsetted data then undergo any necessary transformation and assembly for delivery, including retrieval and appending of QC data and any associated metadata to be included

with the data product delivery, and removal of duplicate records to the extent feasible by the automated DIF services.

e. The data package is forwarded to the data transport and delivery function, which applies a checksum, CRC, or other delivery assurance method, formats the data into the required transport protocol, opens a session with the customer receive data function, delivers data through the communications layer, and forwards transaction information to the data provider's activity logging function.

f. DIF Receive Data Function validates delivery of data using chosen delivery assurance method, performs any necessary transport-level extraction to complete delivery

6.3.2 Push Registry Information to DIF Management Function

This represents a "push" case where the DIF data providers send updated registry information for available data products to the DIF Management Function. There is a reciprocal use case (Section 6.4.1) in which the Management Function "pulls" updated registry information from the data providers.

1. **Goal:** To send updated registry information for the data provider's available DIF data products to the DIF Management Function for inclusion in the Central Registry.
2. **Use Case Diagram:**

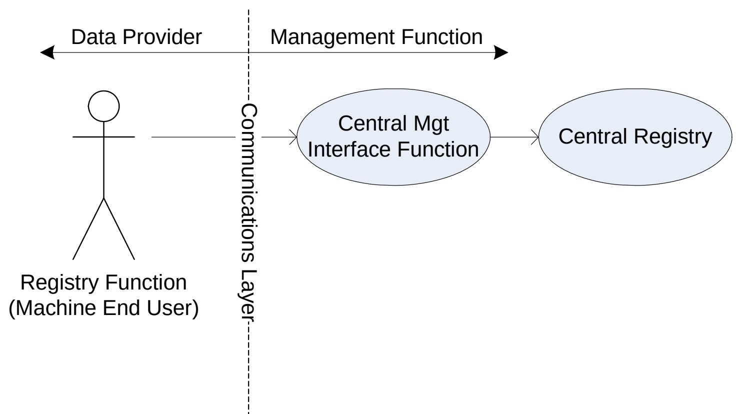

Figure 8 – Use Case: Push Registry Information to DIF Management Function

3. **Actors:** Registry function within the DIF data provider
4. **Pre-existing Conditions:**
   a. DIF data providers is registered with the Central Registry to automatically push updated registry information
5. **Trigger event(s):** DIF data provider's Registry function receives updated registry information for one or more of its DIF data sets
6. **Intermediate steps:**
   a. Data provider's Registry function assembles updated registry information and formats it according to DIF conventions.
   b. Registry opens a communications session with Central Management Interface Function and forwards updated registry information

c. Central Management Interface Function registry receives information and forwards it to the Central Registry, which updates the Central Registry directory accordingly, and notifies end-users as appropriate that updated registry information is available.

6.3.3 Push Metadata to DIF Management Function

This represents a "push" case where the DIF data providers send updated metadata for available data products to the DIF Management Function. There is a reciprocal use case (Section 6.4.2) in which the Management Function "pulls" updated metadata from the data providers.

1. **Goal:** To send updated metadata for the data provider's available DIF data products to the DIF Management Function for inclusion in the Central Metadata Management function.
2. **Use Case Diagram:**

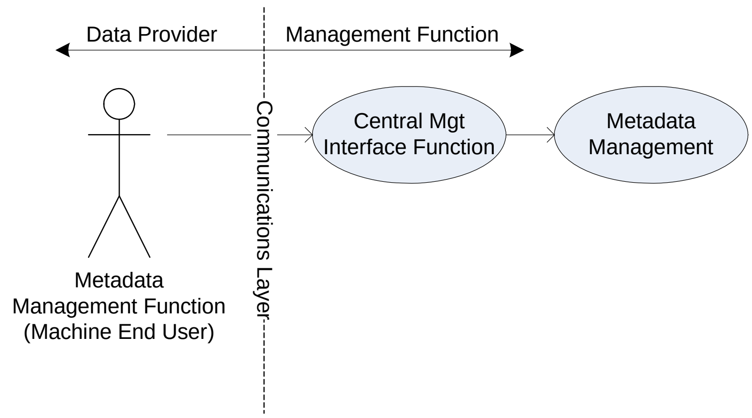

Figure 9 – Use Case: Push Metadata to DIF Management Function

3. **Actors:** Metadata Management function within the DIF data provider
4. **Pre-existing Conditions:**
   a. DIF data providers is registered with the Central Metadata Management function to automatically push updated metadata
5. **Trigger event(s):** DIF data provider's Metadata Management function receives updated metadata definitions for one or more of its DIF data sets
6. **Intermediate steps:**
   a. Data provider's Metadata Management function assembles updated metadata and formats it according to DIF conventions.
   b. Metadata Management function opens a communications session with Central Management Interface Function and forwards updated metadata
   c. Central Management Interface Function receives metadata and forwards it to the Central Metadata Management, which updates the central metadata directory accordingly, and notifies end-users as appropriate that updated metadata is available.

6.4 Use Cases for DIF Management Function

These Use Cases describe scenarios in which the primary actor is a machine end-user that is a data consumer, or customer, of the DIF. In general, machine end-users act to support their own human end-users operating the identified customer decision-support tools.

6.4.1 Retrieve Registry Information from Data Providers

This represents a "pull" case where the DIF Management Function polls the participating DIF data providers for updated registry information for available data products. There is a reciprocal use case (Section 6.3.2) in which the data provider "pushes" updated registry information to the DIF Management Function.

1. **Goal:** To collect and catalog the registry data for the inventory of data products available through the DIF from participating DIF data providers.
2. **Use Case Diagram:**

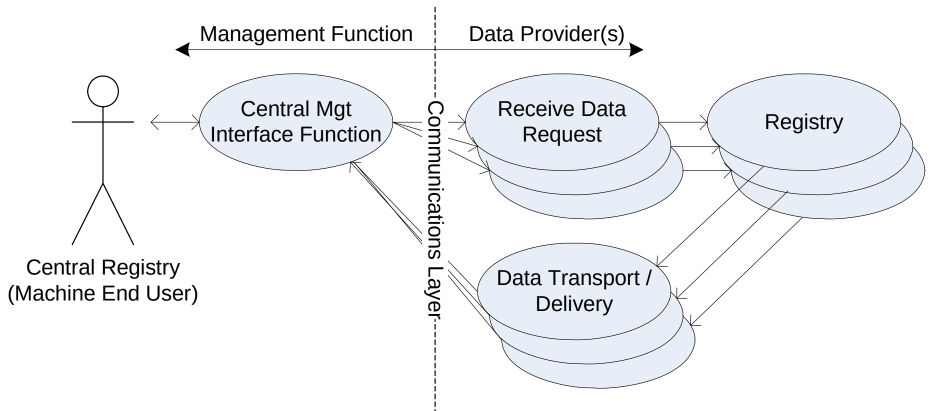

Figure 10 – Use Case: Retrieve Registry Information from Data Provider(s)

3. **Actors:** Central Registry function within the DIF Management Function
4. **Pre-existing Conditions:**
   a. One or more DIF data providers are registered with the Central Registry to be periodically polled for registry information updates
5. **Trigger event(s):** Central Registry sends request(s) to DIF Central Management Interface Function to retrieve registry information from one or more data providers
6. **Intermediate steps:**
   a. Central Management Interface Function opens a communications session with data provider(s) and forwards request to the Receive Data Request function
   b. Receive Data Request function validates the request and forwards request parameters to local Registry
   c. Registry retrieves registry information based on request parameters and forwards it to the Data Transport/Delivery function
   d. Data provider(s) Data Transport/Delivery function sends resulting registry information back to the Central Management Interface Function

e. Central Management Interface Function forwards the registry information to the Central Registry, which updates the Central Registry directory accordingly, and notifies end-users as appropriate that updated registry information is available.

6.4.2 Retrieve Metadata from Data Providers

This represents a "pull" case where the DIF Management Function polls the participating DIF data providers for updated metadata for available data products. There is a reciprocal use case (Section 6.3.3) in which the data provider "pushes" updated metadata to the DIF Management Function.

1. **Goal:** To collect and catalog metadata for the inventory of data products available through the DIF from participating DIF data providers.
2. **Use Case Diagram:**

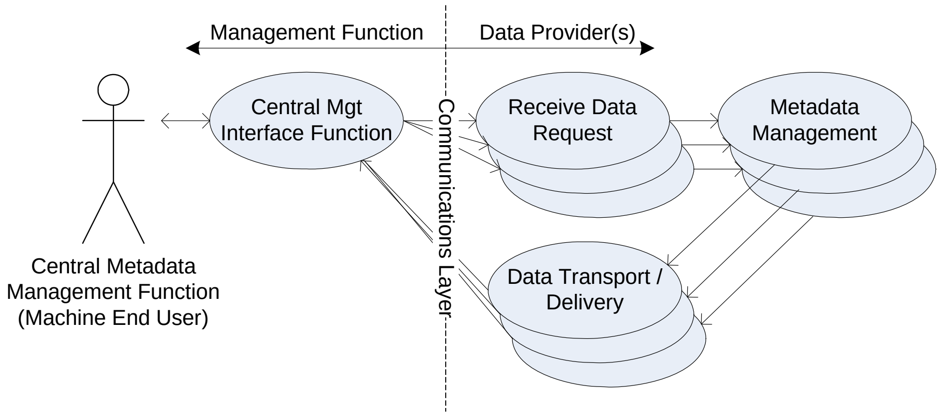

Figure 11 – Use Case: Retrieve Metadata from Data Providers

3. **Actors:** Central Metadata Management function within the DIF Management Function
4. **Pre-existing Conditions:**
   a. One or more DIF data providers are registered with the Central Metadata Management function to be periodically polled for metadata updates
5. **Trigger event(s):** Central Metadata Management function sends request(s) to DIF Central Management Interface Function to retrieve metadata from one or more data providers
6. **Intermediate steps:**
   a. Central Management Interface Function opens a communications session with data provider and forwards request to the Receive Data Request function
   b. Receive Data Request function validates the request and forwards request parameters to local Metadata Management function
   c. Metadata Management function retrieves metadata based on request parameters and forwards it to the Data Transport/Delivery function
   d. Data provider Data Transport/Delivery function sends resulting metadata back to the Central Management Interface Function
   e. Central Management Interface Function forwards the metadata to the Central Metadata Management function, which updates the central metadata directory accordingly, and notifies end-users as appropriate that updated metadata is available.

6.5 Use Cases for Aggregation Services

These Use Cases describe scenarios in which the primary actor can be either an end-user or a machine, representing a customer of the DIF, that has requested data from multiple sources requiring assembly and/or transformation by an intermediate aggregation service.

6.5.1 Aggregate and Deliver Data from Multiple DIF Providers

1. **Goal:** Deliver selected data from multiple DIF data providers into the customer decision-support tool / model where data is assembled / transformed by aggregation services.
2. **Use Case Diagram:**

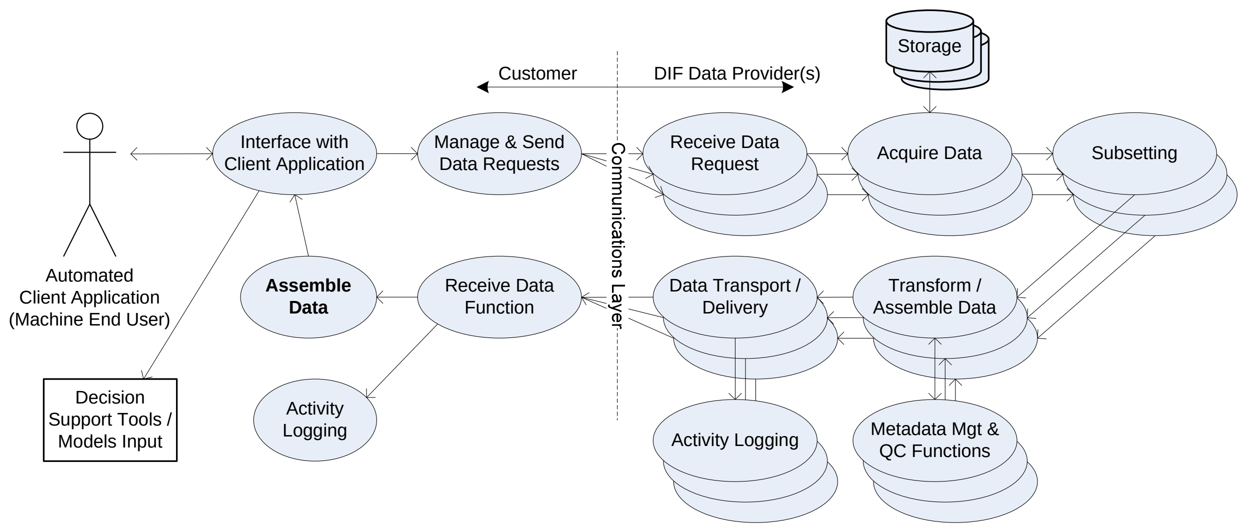

Figure 12 – Use Case: Aggregate and Deliver Data from Multiple DIF Providers

3. **Actors:** Client Application (machine) that interfaces with the identified decision-support tools and the DIF common data sharing infrastructure
4. **Pre-existing Conditions:**
    a. End-user or machine has requested data that is available from multiple data sources requiring aggregation services.
    b. DIF system has assembled registry information regarding the latest available data into the Central Registry.
5. **Trigger event:** Client Application processes automated data request active in queue
6. **Intermediate steps:**
    a. End-user machine initiates an automated data request session with the DIF Client Application Interface and forwards parameters of the automated data request.
    b. DIF Client Application Interface forwards the requested parameters to the Manage and Send Data Requests function, which transmits request to the appropriate data provider(s) using addressing/routing information from the Central Registry.
    c. DIF data provider accepts and validates the data request and the Acquire Data function accesses data provider storage to retrieve the requested data.
    d. Requested data are retrieved from data provider storage and subsetted, if required, according to the automated request criteria.

e. Subsetted data then undergo any necessary transformation and assembly for delivery, including retrieval and appending of any associated metadata to be included with the data product delivery, and removal of duplicate records to the extent feasible by the automated DIF services.

f. The data package is forwarded to the data transport and delivery function, which applies a checksum, CRC, or other delivery assurance method, formats the data into the required transport protocol, opens a session with the customer receive data function, delivers data through the communications layer, and forwards transaction information to the data provider's activity logging function.

g. DIF Receive Data Function validates delivery of data using chosen delivery assurance method, performs any necessary transport-level extraction for delivery to customer, logs activity, and forwards the data to the Client Application Interface if data is already compiled in a single data set, or forwards multiple data sets to the data assembly function.

h. If required, the data assembly function buffers multiple data sets and combines them into a single product for delivery through the Client Application Interface.

i. DIF Client Application Interface forwards requested data into Automated Client Application and/or to the front end of the customer's decision-support tools/models in accordance with the automated request's designation of data destination(s)

6.5.2 Aggregate and Deliver Data from Non-DIF Providers

1. **Goal:** Deliver selected data from one or more data providers that do not serve data in accordance with DIF conventions into the customer decision-support tool/model where data is assembled/transformed by aggregation services.

2. **Use Case Diagram:**

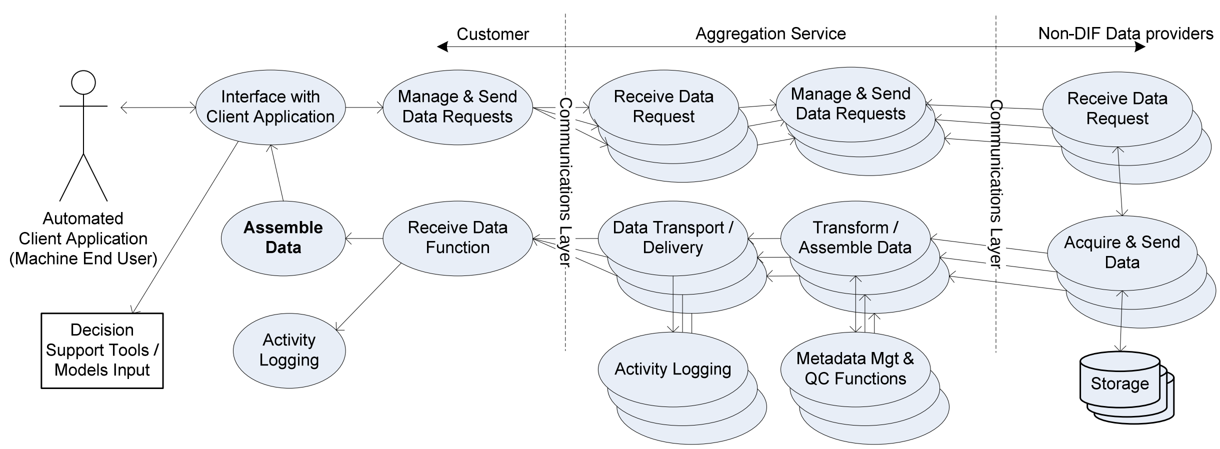

Figure 13 – Use Case: Aggregate and Deliver Data from Non-DIF Providers

3. **Actors:** Automated Client Application computer (machine) that operates between the identified decision-support tools and the DIF Client Application Interface

4. **Pre-existing Conditions:**

    a. End-user or machine has requested data that is available from multiple non-DIF data sources requiring aggregation services.

b. DIF system has assembled registry information regarding the latest available data into the Central Registry.

5. **Trigger event:** Automated Client Application processes automated data request active in queue

6. **Intermediate steps:**

a. End-user machine initiates an automated data request session with the DIF Client Application Interface and forwards parameters of the automated data request.

b. DIF Client Application Interface forwards the request parameters to the Manage and Send Data Requests function, which transmits request to the appropriate data provider(s) using addressing/routing information from the Central Registry.

c. DIF data provider(s) accepts and validates the data request and the Acquire Data function accesses data provider(s) storage to retrieve the requested data.

d. Requested data are retrieved from data provider(s) storage and subsetted, if required, according to the automated request criteria.

e. Subsetted data then undergo any necessary transformation and assembly for delivery, including retrieval and appending of QC data and any associated metadata to be included with the data product delivery, and removal of duplicate records to the extent feasible by the automated DIF services.

f. The data package is forwarded to the data transport and delivery function, which applies a checksum, CRC, or other delivery assurance method, formats the data into the required transport protocol, opens a session with the customer receive data function, delivers data through the communications layer, and forwards transaction information to the data provider's activity logging function.

g. DIF Receive Data Function validates delivery of data using chosen delivery assurance method, performs any necessary transport-level extraction for delivery to customer, logs activity, and forwards the data to the Client Application Interface if data is already compiled in a single data set, or forwards multiple data sets to the data assembly function.

h. If required, the data assembly function buffers multiple data sets and combines them into a single product for delivery through the Client Application Interface.

i. DIF Client Application Interface forwards requested data into Automated Client Application and/or to the front end of the customer's decision-support Tools / Models in accordance with the automated request's designation of data destination(s)

7 Requirements Traceability

This section contains tables comprising the Requirements Traceability Matrices that cross-reference functional requirements from Reference Document RD5 *DIF Functional Requirements Document* to Use Cases and/or Subcomponent Functionalities described above. These are organized in three tables corresponding to the three development phases referenced in the Functional Requirements Document. In general, the phases correspond to specific discrete implementation timelines, according to priorities.

7.1 Phase I Requirements

Table 1 – Phase I Requirements Traceability Matrix

ID	Subject	Requirement	Applicable Use Cases
GEN045	Common Data Model	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.	All
RDS001	Currents / ADCP / NDBC	The DIF shall provide MMS ADCP Currents data from NDBC.	6.1.2 6.2.1
RDS035	Currents / TABS / NDBC	The DIF shall provide Texas Automated Buoy Systems (TABS) Currents data from NDBC.	6.1.2 6.2.1
RDS040	Currents / COMPS / USF (NDBC?)	The DIF shall provide Coastal Ocean Monitoring and Prediction System (COMPS) Buoy Currents data from University of South Florida.	6.1.2 6.2.1
RDS100	Currents / HF Radar / NDBC?	The DIF shall provide currents data from HF Radar in NetCDF format. GAP: Verify data source provider.	6.1.2 6.2.1
RDS115	Currents / ADCP / NDBC?	The DIF shall provide Current data from shipboard ADCP profiles. GAP: Verify data source provider.	6.1.2 6.2.1
MTD001	Metadata Standards	Each DIF logical dataset or data service shall have at least one FGDC or ISO metadata record.	6.1.3 6.3.3 6.4.2
MTD002	Metadata Standards	DIF metadata standards shall adequately define data attribute details (e.g. unit of measure, reporting convention, precision, code definitions)	6.1.3 6.3.3 6.4.2
MTD015	Metadata, Public Access	The DIF shall encode all metadata in valid XML and make it available for public access.	6.1.3 6.3.3 6.4.2
MTD015	Metadata Reporting	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).	6.4.2
QC015	Provide data quality information in metadata	The DIF shall document quality procedures in the metadata record.	6.1.3 6.3.3 6.4.2
XPT010	Data Transport – Real Time	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.	6.1.2 6.2.1
XPT015	Data Transport – Delayed Mode	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.	6.1.2 6.2.1

FILT010	Aggregation	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.	6.1.2 6.2.1

7.2 Phase II Requirements

Table 2 – Phase II Requirements Traceability Matrix

ID	Subject	Requirement	Applicable Use Cases
GEN001	Core Variable – Temperature	The DIF shall retrieve, process, and deliver to end-users data containing the core variable of Temperature.	6.1.2 6.2.1 6.3.1
GEN005	Core Variable – Salinity	The DIF shall retrieve, process, and deliver to end-users data containing the core variable of Salinity.	6.1.2 6.2.1 6.3.1
GEN010	Core Variable – Currents	The DIF shall retrieve, process, and deliver to end-users data containing the core variable of Currents.	6.1.2 6.2.1 6.3.1
GEN015	Core Variable – Sea Level	The DIF shall retrieve, process, and deliver to end-users data containing the core variable of Sea Level.	6.1.2 6.2.1 6.3.1
GEN020	Core Variable – Ocean Color	The DIF shall retrieve, process, and deliver to end-users data containing the core variable of Ocean Color.	6.1.2 6.2.1 6.3.1
GEN025	Real-time data	The DIF shall support access and delivery (transport) of real time data, as applicable, for the data sets defined in Section 5.2.	6.1.2 6.2.1 6.3.1
GEN026	Delayed mode data	The DIF shall support access and delivery (transport) of delayed mode data, as applicable, for the data sets defined in Section 5.2.	6.1.2 6.2.1 6.3.1
GEN030	Data Type – Time Series	The DIF shall support data in the form of time series.	All
GEN035	Data Type – Profiles	The DIF shall support data in the form of profiles.	All
GEN040	Data Type – Gridded	The DIF shall support data in the form of gridded data.	All
RDS005	Salinity / C-MAN / NDBC	The DIF shall provide CTD offshore Salinity data from NDBC buoys.	6.1.2 6.2.1 6.3.1
RDS020	Ocean Color / SeaWiFS / CoastWatch	The DIF shall provide near real-time SeaWiFS Ocean Color data from CoastWatch servers.	6.1.2 6.2.1 6.3.1
RDS025	Ocean Color / MODIS Aqua (NASA) / CoastWatch	The DIF shall provide near real-time MODIS Aqua Ocean Color data from CoastWatch servers.	6.1.2 6.2.1 6.3.1
RDS045	Salinity (Estuarine) / NERRS / NERRS	The DIF shall provide estuarine Salinity data from NERRS.	6.1.2 6.2.1 6.3.1

RDS050	Salinity (Estuarine) / PORTS® / CO-OPS	The DIF shall provide estuarine Salinity data from the CO-OPS PORTS® data sets.	6.1.2 6.2.1 6.3.1
RDS055	Salinity (Estuarine) / NEPs / EPA	The DIF shall provide estuarine Salinity data from the EPA National Estuary Program (NEP)	6.1.2 6.2.1 6.3.1
RDS180	Temperature/ /CO-OPS	The DIF shall provide Water Temperature data from CO-OPS systems.	6.1.2 6.2.1 6.3.1
RDS185	Temperature/ /NDBC	The DIF shall provide Water Temperature data from NDBC.	6.1.2 6.2.1 6.3.1
RDS180	Temperature/ /NERRS	The DIF shall provide Water Temperature data from NERRS.	6.1.2 6.2.1 6.3.1
RDS180	Temperature/ /NEP	The DIF shall provide Water Temperature data from NEP systems.	6.1.2 6.2.1 6.3.1
RDS185	Water Level/ /CO-OPS	The DIF shall provide Water Level data from CO-OPS systems.	6.1.2 6.2.1 6.3.1
RDS185	Water Level/ /NDBC	The DIF shall provide Water Level data from NDBC systems.	6.1.2 6.2.1 6.3.1
RDS190	Sea Level/Water Level Gauges/USGS (via NHC)	The DIF shall provide Sea Level data from USGS portable water level gauges via NHC has data.	6.1.2 6.2.1 6.3.1
VAL005	Remove duplicate records	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.	6.1.2 6.2.1 6.3.1
TRAN001	Data transformation	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.	All
TRAN020	Aggregation by data type (VDAC)	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.	6.1.2 6.2.1 6.3.1
MTD003	Metadata Standards	DIF metadata standards shall adequately convey data file formats or structures.	6.1.3 6.3.2 6.4.2
MTD004	Metadata Standards	DIF metadata standards shall adequately capture data quality information (e.g. QC tests applied, QC flags and flag definitions)	6.1.3 6.3.2 6.4.2
MTD005	Metadata Standards	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.	6.1.3 6.3.2 6.4.2
MTD010	Data-Metadata Linkage	The DIF shall ensure that the linkages between data and metadata are maintained with great reliability.	6.1.3 6.3.2 6.4.2

MTD020	Metadata Tools	The DIF shall provide tools to enable end-users and data providers to increase their capability in metadata generation and management.	6.1.3 6.3.2 6.4.2
MTD025	Consistent Terminology	The DIF shall establish a standard glossary for use by customers and data providers.	All
MTD030	Central Metadata Registry	The DIF shall provide a central metadata registry.	6.3.2 6.4.1
QC001	QC Partnership	The DIF shall ensure that quality control operations are a partnership among data observation/collection components, processors, analysts, other users, and the DMAC.	All
QC002	QC Standards	The DIF shall identify, adopt, and adapt existing standards for quality control.	All
QC005	Known/documented Quality	The DIF shall provide a mechanism for ensuring that data are of known and documented quality.	All
QC010	QC Documentation	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.	6.1.3 6.3.3 6.4.2
QC020	QC Transport	The DIF shall transport QC and error characteristics, flags, through from data provider to data consumer. **Gap:** will need to identify all error and QC data per integrated data set.	6.1.2 6.2.1 6.3.1
QC030	Data QC	The DIF shall accurately report lat/long and time of measurement.	6.1.2 6.2.1 6.3.1
QC035	Flag Missing Data	The DIF shall consistently flag missing data.	6.1.2 6.2.1 6.3.1
QC040	QC of Sea Level products	The DIF shall provide evaluated QC'd Sea Level products and elevation contours from data sources. **GAP:** Identify data sources.	6.1.2 6.2.1 6.3.1
QC050	Data quality flags	The DIF shall implement data quality flags consistently across the data community.	6.1.2 6.2.1 6.3.1
QC055	Data quality flags	The DIF shall provide raw data with data quality flags rather than cleaned data. **GAP:** Identify which variables and data sources.	6.1.2 6.2.1 6.3.1
QC060	Deliver Sea Level data of known quality	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.	6.1.2 6.2.1 6.3.1
QC065	Access and QC of ADCP Current data	The DIF shall provide QC of Current data from shipboard ADCP profiles.	6.1.2 6.2.1 6.3.1
DSA005	Contribution to archive	The DIF shall facilitate contribution of standards-based data to NODC archive from ecosystem community.	
XPT001	Transport, Machine-to-Machine	The DIF shall support transport protocol(s) to allow machine-to-machine data transport.	6.1.1 6.1.2
XPT005	Transport, End-User	The DIF shall support transport protocol(s) to allow data retrieval manually by a user.	6.1.1

XPT020	Metadata Transport	The DIF shall provide for the transport of metadata associated with all data transported.	6.1.2 6.1.3 6.2.1 6.3.1 6.3.3 6.4.2
XPT025	Data Transport Standards	Data transport standards applicable to the DIF shall be identified and adapted as needed to support DIF functional requirements.	All
XPT040	Automated Push Delivery	The DIF shall provide automated delivery, “push”, of data to a customer model data input.	6.3.1
XPT045	Manual Delivery	The DIF shall allow users to manually select data for delivery.	6.1.2
XPT050	Integrated Access Functions	The DIF data and metadata access functions shall be seamlessly integrated.	6.1.1 6.1.2 6.1.3 6.2.1 6.3.1
FILT001	Subsetting, Server side	The DIF shall allow for server-side subsetting of data for all relevant data.	6.1.2 6.2.1 6.3.1
FILT002	Subsetting, Data Field	The DIF shall allow subsetting of data based on any field in the relevant data set.	6.1.2 6.2.1 6.3.1
SEC001	NOAA Req’ts	The DIF shall implement applicable NOAA requirements for IT security.	All
SEC005	External Interfaces	The DIF shall ensure that interfaces to external systems are secure from unauthorized access.	All
SEC010	User and System Functional Separation	The DIF shall separate user functionality (including user interface services) from information system management functionality.	All
SEC015	Security Function Separation	The DIF shall isolate security functions from non-security functions.	All
SEC020	Protect Secure Information	The DIF shall protect the confidentiality of security-relevant system information and integrity of all transmitted information.	All
SEC025	Network Connection Termination	The DIF shall terminate network connections at the end of each session or after a configurable period of inactivity.	All
SEC030	Session Authenticity	The DIF shall provide mechanisms to protect the authenticity of communications sessions.	All
SEC035	Virus Protection	The DIF shall incorporate antivirus protections with automatic updates.	All
SEC040	Input Validation	The DIF shall check information inputs for accuracy, completeness, and validity.	All
SEC045	Source Authentication	The DIF shall authenticate that data sets received at client locations are from a valid DIF data source	6.1.2 6.2.1 6.3.1
ADM001	System Monitoring	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.	
ADM005	Fault Detection and Correction	The DIF shall provide a mechanism for detection of status of DIF data provider interfaces and report the status externally.	

ADM010	System Extensibility	The DIF shall provide a mechanism for system extensibility.	All
ADM015	Data Availability Policies	The DIF shall establish and publicize policies for data availability.	6.1.1 6.3.2 6.4.1
ADM020	User Feedback	The DIF shall provide a mechanism for soliciting and responding to user feedback.	
ADM025	International Linkages	The DIF shall establish and maintain international linkages.	
ADM030	Service Registry	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.	6.3.2 6.4.1

7.3 Phase III Requirements

Table 3 – Phase III Requirements Traceability Matrix

ID	Subject	Requirement	Applicable Use Cases
RDS010	Real-time Winds / C-MAN / NDBC	The DIF shall provide C-MAN real-time Wind Speed and Direction data from NDBC.	6.1.2 6.2.1 6.3.1
RDS011	Real-time Winds / NWLON / NDBC	The DIF shall provide NWLON real-time Wind Speed and Direction data from NDBC.	6.1.2 6.2.1 6.3.1
RDS012	Real-time Winds / Buoys / NDBC	The DIF shall provide NDBC Buoy real-time Wind Speed and Direction data from NDBC.	6.1.2 6.2.1 6.3.1
RDS030	Currents / TGLO 3-D / Tx A&M (NDBC?)	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.	6.1.2 6.2.1 6.3.1
RDS060	Winds (Forecasted) / NAM / NCEP	The DIF shall provide NAM model Forecasted Winds data from NCEP sites .	6.1.2 6.2.1 6.3.1
RDS065	Winds (Forecasted) / NWS / NWSTG – GTS	The DIF shall provide Forecasted Winds data from NWS regional marine text forecasts. GAP: Verify source provider of forecast data.	6.1.2 6.2.1 6.3.1
RDS070	Cell Counts / Field surveys / State & local partners	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.	6.1.2 6.2.1 6.3.1
RDS075	Bathymetry / Coastal Relief Model / NGDC	The DIF shall provide Bathymetry data from the NGDC digital elevation model (DEM) Coastal Relief Model.	6.1.2 6.2.1 6.3.1
RDS080	Bathymetry / Hydrographic Surveys / NOS	The DIF shall provide Bathymetry data from NOS Hydrographic Surveys.	6.1.2 6.2.1 6.3.1
RDS085	RESERVED		6.1.2 6.2.1 6.3.1

ID	Subject	Requirement	Applicable Use Cases
RDS090	Unknown / Unknown / USGS	The DIF shall provide USGS data to replace current web scraping. GAP: Determine variables, platforms, sources and formats.	6.1.2 6.2.1 6.3.1
RDS095	Currents (Offshore) / AUVs / Unknown	The DIF shall provide offshore Currents and other data from existing AUVs. GAP: Verify data source provider.	6.1.2 6.2.1 6.3.1
RDS105	Topography / National Elevation Database / USGS	The DIF shall provide Topography data from the USGS National Elevation Database.	6.1.2 6.2.1 6.3.1
RDS110	Topography / LiDAR Data / Unknown	The DIF shall provide Topography data LiDAR. GAP: Verify data source providers, formats.	6.1.2 6.2.1 6.3.1
RDS120	Water Pressure / Unknown / Unknown	The DIF shall provide subsurface Pressure data from available sources. GAP: Identify platforms, data source provider(s), formats.	6.1.2 6.2.1 6.3.1
RDS125	Salinity / regional observing systems / Unknown	The DIF shall provide sub-surface and near-shore Salinity data from regional observing systems. GAP: Verify data source providers, access methods, formats.	6.1.2 6.2.1 6.3.1
RDS130	Currents / TBD / TBD	Increase spatial coverage of current data, or access to in situ currents. GAP: Identify specific sources.	
RDS140	Temperature / TBD / TBD	Obtain additional surface and subsurface Temperature data at various resolutions. GAP: Determine data availability and sources.	6.1.2 6.2.1 6.3.1
RDS145	Temperature / TBD / TBD	The DIF shall provide access to Temperature interpolations for Sea Surface and Subsurface Temperatures from models. GAP: identify models and source provider.	6.1.2 6.2.1 6.3.1
RDS150	TBD / TBD / GTS	The DIF shall provide access to historical data for all time series data provided through the GTS. GAP: Identify variables and observation platforms.	6.1.2 6.2.1 6.3.1
RDS155	Temperature / TBD / TBD	The DIF shall provide Subsurface Temperature data to address the Gulf of Mexico data gaps. GAP: Identify observation platforms and data source providers.	6.1.2 6.2.1 6.3.1
RDS160	Salinity / TBD / TBD	The DIF shall provide Salinity data to address Gulf of Mexico data gaps. GAP: Identify observation platforms and data source providers.	6.1.2 6.2.1 6.3.1
RDS165	Winds (Forecasted) / TBD / Regional sources	The DIF shall provide Forecasted Winds from regional sources. GAP: Need to identify specific sources.	6.1.2 6.2.1 6.3.1
RDS170	High Water Mark / Storm gauges / USGS	The DIF shall provide post-storm High Water Mark data from USGS storm gauges.	6.1.2 6.2.1 6.3.1
RDS175	Ocean Heat / TBD / Tropical Prediction Center	The DIF shall provide the Ocean Heat content product produced at Tropical Predication Center.	6.1.2 6.2.1 6.3.1

8 Summary of Impacts

This section describes the anticipated operational and organization impacts of the proposed system. It also describes the anticipated impacts during integration and transition.

Initially the DIF will consist of a series of initial integration projects established at selected data providers and customers. At data provider locations, these projects will be used to establish the tools and processes needed to make integrated data readily available. Integration projects within customer analysis/decision-support tools will be used to access and receive the integrated data and measure the value of that data to those tools.

At the conclusion of these initial projects, an assessment and analysis of alternatives will be conducted to determine the way forward. It is anticipated that the analysis of alternatives for each project will produce one of the following recommendations:

- Transition to operational system – if the project is deemed successful (according to established performance benchmarks) and operationally feasible, the project will transition from an experimental system into a data provider's or customer's operational environment.
- Continuation/expansion of the project – if the initial project is deemed successful, but warrants additional development or analysis, the project can be maintained or expanded as an experimental system to allow for further test and evaluation.
- Discontinue projects – if it is determined that there is limited benefit in maintaining/expanding the initial project or transitioning it to an operational system, the project may be discontinued with no follow-on activity other than communication of "lessons-learned" to the larger data integration community.

Correspondingly, DIF components will typically be in one of three states: initial, operational, or discontinued. The impacts will be described in the sections below in the context of these states.

8.1 Operational Impacts

This subsection describes the anticipated operational impacts on the user, development, support, and maintenance agency or agencies during operation of the proposed system.

Operational Impacts of DIF Component at DATA PROVIDERS			
	Initial State	**Operational State**	**Discontinued**
User Impacts	N/A. User impacts do not apply at Data Provider sites.	N/A. User impacts do not apply at Data Provider sites.	N/A. User impacts do not apply at Data Provider sites.
Development	Significant development resources will be required to implement data provider design implementations. This will impact the data provider's development resources. Even if contract resources are used, some in-house resources will be	Once the system becomes operational, impacts on development resources are minimal and limited to occasional fixes of software defects.	No impact on development resources if a data provider design implementation project is discontinued, other than opportunity cost of sunk

Operational Impacts of DIF Component at DATA PROVIDERS			
	Initial State	**Operational State**	**Discontinued**
	required.		development funds.
Support / Maintenance	Support resources, including operations, maintenance, repair, and troubleshooting, at a data provider location will be required to support the DIF implementation. This support is in addition to existing operational systems which must be maintained during the design implementation phase.	After transition to operations, the additional support infrastructure required for the design implementation will need to continue as long as the data provider maintains their existing (legacy) operational systems.	No impact on support resources if a data provider design implementation project is discontinued; review to pre-implementation state.

Operational Impacts of DIF Component at CUSTOMER			
	Initial State	**Operational State**	**Discontinued**
User Impacts	Users may be required to perform their analysis multiple times; once using their legacy operational analysis/decision-support tool (as usual), and once using the analysis/decision-support tool which makes use of DIF components. This impacts user workload and required resources.	Once the DIF components at a customer location become operational, it is anticipated that workload is reduced compared with the initial state (assumes use of the legacy system no longer required). It is also envisioned that user workload is reduced when compared with pre-implementation activities; use of DIF integrated data and access methods could streamline workflow.	If the customer design implementation is discontinued, any modifications to user decision-support tool that were made during the initial phase will no longer be applicable; revert to legacy tool.
Development	Development resources will be required to implement customer design implementations. This will impact the customer's development resources. Even if contract resources are used, some in-house resources will be required.	Once the system becomes operational, impacts on development resources are minimal and limited to occasional fixes of software defects or enhancements to take advantage of additional sources of data or data types.	No impact on development resources if a customer design implementation project is discontinued.
Support / Maintenance	The support infrastructure, including operations, maintenance, repair, and troubleshooting, at a customer DIF	Once the design implementation becomes operational, the additional support resources required during the initial state should no	Reverts to pre-implementation state.

Operational Impacts of DIF Component at CUSTOMER			
	Initial State	**Operational State**	**Discontinued**
	implementation location will be required to support the design implementation project in addition to existing operational systems.	longer be required. Support requirements should revert to pre-implementation state.	

Operational Impacts of DIF Component for Management Function(s)			
	Intial State	**Operational State**	**Discontinued**
User Impacts	Users will require training in order to exploit the benefits of the functionality provided by the DIF management system. Workload could be streamlined by using the DIF management system as a single point of information, instead of sourcing information from numerous disparate locations. However, during the intial state users may have to continue with legacy processes to support operational systems, in which case use of DIF management functions is additive to their existing workload.	Users will require training in order to exploit the benefits of the functionality provided by the DIF management system. User workload could be streamlined by using the DIF management system as a single point of information, instead of sourcing information from numerous disparate locations.	No impact other than opportunity cost of sunk development funds; revert to pre-implementation state.
Development	Development resources will be required to implement DIF management functions. Depending on where and how the system(s) is/are implemented, impacts will vary.	Once the system becomes operational, it is likely that development resources will be required from time to time to add functionality to support DIF customers and data providers. In addition, development support will be required for troubleshooting and fixes for software defects.	No impact on development resources if the design implementation project is discontinued.
Support / Maintenance	A support infrastructure, including operations, maintenance, repair, and troubleshooting, will be required to support the design implementation	Once the system becomes operational, support resources will be required on an on-going basis. Because this functionality does not currently exist, the support structure will	Reverts to pre-implementation state.

Operational Impacts of DIF Component for Management Function(s)			
	Intial State	**Operational State**	**Discontinued**
	project. If the DIF management functions are co-located at a data provider or other customer, these impacts could be minimal compared with a non co-located facility.	need to be identified and funded. Again, if the DIF management functions are co-located at a data provider or other customer, these impacts could be minimal compared with a non co-located facility.	

8.2 Organizational Impacts

The DIF components are in many ways enhancements to current operational practices between data providers and data consumers, and fit within existing organizational structures. However, short-term, initial, inconveniences to technical and operations organizations will likely occur. These inconveniences include the additional technical and management resources associated with implementation and testing of DIF standards. Technical staff involved in the development of information management systems will need to ensure that their systems conform to the interoperability standards set by DIF.

In addition, the implementation and operation of the DIF requires a high degree of on-going coordination within and between the numerous participating organizations. The level of coordination required to ensure interoperability of all DIF components will have an impact on the internal and external processes and resources of all participating organizations.

The DIF management function described in this document implies the need for an over-arching organization that manages the implementation and operation of such a management system. This functionality is not specific to either a data provider or data consumer; there is currently no cross-cutting organization that is equipped to develop, operate, and maintain such a system. An existing organization would need to take responsibility for this function and associated systems.

8.3 Impacts During Transition to Proposed System

There are numerous impacts during the transition from DIF integration projects to operational system. Specific impacts are as follows:

1. Data provider integration projects – there are several impacts while the data provider project is undergoing test and validation:
    - Operational Impacts - The data providers will need to maintain their existing operational systems and processes so that they can continue to support existing customers. This will require parallel operation of the existing systems and the DIF projects. This parallel operation primarily impacts the data provider, necessitating additional resources to maintain two operational systems.
    - Development/Support Impacts – Data provider technical personnel will be required to support test and validation of the integration project before it can transition to an operational system. Appropriate IT security controls will need to be integrated during the development process.
    - Customer Impacts – In anticipation of a transition of the DIF integration project to an operational system, the data provider will need to notify customers of the timing, impacts, and benefits of the transition. This is particularly true if legacy services will be merged with the new services in a short period after DIF implementation.
    - Training Impacts – Training will be required on all aspects of the new system prior to transition. Training will be required for operators, maintenance personnel, and users. Training materials, user manuals, maintenance manuals, and troubleshooting procedures must be developed and distributed.

2. Customer integration projects – similarly, there are several impacts while the integration project is undergoing test and validation:
    - Operational Impacts - The customer will likely need to maintain their existing operational analysis/decision-support tools and processes until the new system is validated. This will require parallel operation of the existing systems and the DIF integration project. This parallel operation primarily impacts the operator of the analysis/decision-support tool, necessitating additional resources to operate two systems.
    - Development/Support Impacts – Customer technical personnel will be required to support test and validation of the integration project before it can transition to an operational system. Appropriate IT security controls will need to be integrated during the development process.
    - Training Impacts – Training will be required on all aspects of the new system prior to transition. Training will be required for operators, maintenance personnel, and users. Training materials, user manuals, maintenance manuals, and troubleshooting procedures must be developed and distributed.

9 Analysis of Proposed System

This section provides a high-level analysis of the benefits, limitations, advantages, disadvantages, and alternatives and trade-offs considered for the proposed system.

9.1 Summary of Improvements

This subsection provides a summary of the benefits to be provided by the DIF.

1. **Standards Adoption** – The DIF establishes a defined list of standards and protocols for data content, encoding, transport and metadata (including QC) that govern the DIF architecture, and are broadly applicable to a range of data types and parameters. With these standards and associated conventions, a common data sharing infrastructure can be established and expanded which will allow data providers the tools and information they need to make their data available to a broader community of consumers using DIF-approved conventions. Consumers will enjoy easier and more efficient access and use of a wide range of oceanographic and ocean-based meteorological data from multiple providers. DIF will formalize a standards-based common data sharing infrastructure that is expected to facilitate and improve data integration of ocean variables across NOAA Line Offices, Regional Associations, and other data sources.
2. **Integrated Data and Standardized Access** – By applying DIF-adopted data content, transport, and metadata standards the DIF improves dissemination of mission critical ocean related data which increases the value and effectiveness of the data available for use in decision-support tools. It provides major advances in data consistency and facilitates interoperability between multiple independent data sources, and has the flexibility to be extensible to include additional functionality, enhancing the value of individual data sources beyond their original purpose.

   Specific benefits to decision-support tools that make use of the integrated data can include:

   - more timely forecasts,
   - improved decision-support tool accuracy,
   - greater geographic coverage and resolution,
   - improved efficiency in time and other resources spent in tool execution (e.g. reduced development time, reduced time from data acquisition to product dissemination), reduced operating costs,
   - broader application of data sources beyond what was used prior to the implementation of the DIF functionality, due to enhanced access to, and consistent formats of, previously "unavailable" (e.g. difficult to use/ access) streams of relevant data.

3. **DIF Management Functions** – The DIF management functions provides a cohesive layer and single source for information and functionality to assist both data providers and data consumers. Specific functionality and benefits include:

   - **Registry** – Through the DIF central registry, users can access key discovery metadata for data available within the DIF, including methods for access, etc. This provides a more efficient method of browsing available data than visiting individual data provider sites or using non domain-specific search facilities (Google, etc).
   - **Metadata management** – Through the DIF central metadata registry, users can access metadata for all data available within the DIF. This provides a more efficient method of browsing metadata than visiting individual data provider sites or using non domain-specific search facilities (Google, etc).

- **System Monitoring** – The DIF will provide basic operational monitoring of all DIF components, such as data provider services. The DIF will periodically poll the components to determine if services are currently available or not, and will maintain up-to-date status on each component. Status information can be broadcast to parties that have registered interest, or users can visit the DIF management interface to view which services are currently available. This monitoring function can provide users with an overview of which services are available at any time.

9.2 Disadvantages and Limitations

The following is a summary of the disadvantages and/or limitations of the DIF.

1. **Data Discovery** – The DIF as planned does not provide for extensive data discovery. During the early planning stages of the DIF it was determined that data discovery was a topic too complex to address in the scope of the DIF. Limited data discovery capability will be supported through the search capability of the central registry.
2. **Archive** – The DIF will not implement a data archive. However, the DIF will provide the capability for the delivery of DIF data sets to archive centers, which may be done by having an archive center register with the DIF and get set up as a customer with a client infrastructure implementation.

9.3 Alternatives and Trade-offs Considered

This subsection describes major alternatives considered, the trade-offs among them, and rationale for the decisions reached.

1. **Distributed Architecture** – As shown in Section 6, the DIF architecture is a highly distributed, framework approach with components at data providers, client locations, software gateways, and possibly a central management location. An alternative to the framework approach is a highly centralized, enterprise systems approach.

   A framework establishes and coordinates people, processes, standards, workflows and can support many new or modified systems. This approach recognizes that the implementation environment is highly variable, and not in direct control of the framework designers. The framework is comprised of a series of specific design implementations that together make up the framework. Over time, additional implementations are added as resources and budget at independent entities allow.

   Alternatively, the DIF could be a well defined "enterprise" system consisting of a consolidated system of storage, hosts, services, networks and interfaces developed and managed by a central entity. This approach assumes a massive investment in hardware, software, and personnel and a strong central data management that is not currently in place or planned, and does not fit the overall DMAC concept.

   At its core, the DIF is an interoperability plan that must bring together diverse parts of NOAA, and non-NOAA players, to incrementally move toward a data integration framework. To design, build and implement the DIF, the project will leverage existing capacity and expertise resident in NOAA and Regional Associations. As needed, and subject to budget availability, resources will be provided to support these contributors as they help develop the DIF. Project teams and affiliated working groups composed of cross Line Office and Goal Team representatives will design, carry out, or direct the technical work and building of DIF components, and will be involved in the testing and evaluation of the DIF. This approach lends itself much more to a framework approach.

2. **DIF Management Functions** – The primary DIF data transform and delivery components reside at data providers, customer facilities, and aggregation service hosts. The primary operational concept of the DIF provides for a customer decision-support tool’s interaction with one or more data providers and/or aggregation services to request data or metadata. In this context, there was significant consideration as to the need for cross-cutting functionality that is not specific to an individual data provider or customer decision-support tool. The core DIF objectives of data transformation and delivery could be met without such cross-cutting functionality, but it was determined that the utility of such a central management functionality is worth the additional development and operational costs. This functionality includes a central data registry, central metadata directory and system monitoring of the various DIF components to provide status to users.

3. **Data Sources** – The initial focus of the DIF is to integrate data from all NOAA sources. However, through the requirements analysis process it was recognized that data from non-NOAA sources, such as Regional Associations and other IOOS partners, could provide additional benefit to the decision-support tools and should be included as part of the DIF.

10 Definitions and Acronyms

ADCIRC	Advanced Circulation Hydrodynamic Model
ADCP	Acoustic Doppler Current Profiler
API	Application Programming Interface
AUV	Autonomous Underwater Vehicles
AWIPS	Advanced Weather Interactive Processing System
CalCOFI	California Cooperative Oceanic Fisheries Investigations
CDMO	Centralized Data Management Office
CI	Coastal Inundation
C-MAN	Coastal-Marine Automated Network
COMPS	Coastal Ocean Monitoring and Prediction System
CONOPS	Concept of Operations
CO-OPS	Center for Operational Oceanographic Products and Services
CRC	Cyclical Redundancy Check
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
GSFC	Goddard Space Flight Center
GTS	WMO Global Telecommunications System
GTSPP	Global Temperature and Salinity Profile Program
HAB	Harmful Algal Bloom
HDF	Hierarchical Data Format
HF	High frequency (radar)
HI	Hurricane Intensity
HMI	Human Machine Interface
HTTP	Hyper Text Transfer Protocol
HWRF	Hurricane Weather Research and Forecasting
HYCOM	Hybrid Coordinate Ocean Model
IDS	Input Data Source
IEA	Integrated Environmental 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

NAM	North American Mesoscale
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