NOAA

Integrated Ocean Observing System (IOOS) Program Office

Data Integration Framework (DIF)

DIF As-Is Baseline Systems Document

(DRAFT Version 0.6)

September 19, 2007

DIF-0807-IOOS-PLN-PP

Revisions

Version	Description of Version	Date Completed
Draft 0.1	Initial draft	mm/dd/yyyy
Draft 0.2		
Draft 0.3		
1.0		
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 Purpose and Scope

1.1 Purpose and Scope

The purpose of this document is to summarize the current "as-is" state of the systems and data products that feed the five core ocean variables to the proposed customers of the Integrated Ocean Observing System (IOOS) Data Integration Framework (DIF). This is intended to provide a baseline of the subject systems and data products including data sources, formats, contents, frequency, metadata, and transport and access methods currently in place that provide the core variables to the identified models for the Initial Operating Capability for the DIF. This document will serve as both a general reference and as a basis for the further development of the DIF system requirements and concept of operations.

The scope of this document is primarily the outputs of, and the methods of access to, the various collection and processing systems that provide the data products in use by the customer models. It includes summary descriptions of each system as well as more detailed information tables for each data product in use by the models that include information about the responsible Line Office, data product format, source data attributes, metadata information, product dissemination, IT security, and data integrity and archival properties.

This bulk of the remainder of this document presents a summary baseline, or "As-Is", description of the various systems that act as data sources for the identified customer models for the five core variables. A brief description is given of the data products and users, metadata policy, transport and access methods, data archiving, and quality control mechanisms. Following each summary are tables that list more details about each data product being used by the identified customer models.

The last two sections contain a listing of additional input data products and sources not yet incorporated into separate descriptive sections and a listing of definitions and acronyms.

2 Center for Operational Oceanographic Products and Services (CO-OPS)

As part of NOAA's National Ocean Service (NOS), CO-OPS collects, analyzes, and distributes historical and real-time observations and predictions of water levels, coastal currents, and other meteorological and oceanographic data. The CO-OPS provides data in use by the customer models via two systems: the National Water Level Observation Network (NWLON) and the Physical Oceanographic Real-Time System (PORTS). These systems are summarized below along with the relevant data products from each. CO-OPS is currently working on achieving interoperability with the IOOS and has completed the IOOS DMAC Interoperability Plan The Current State Assessment for Center for Operational Oceanographic Products and Services' (CO-OPS) National Water Level Observation Network (NWLON), dated May 22, 2006, and much of the content of this section comes from that document.

2.1 NWLON and PORTS System Description

The NWLON system is part of the National Water Level Program (NWLP) and is based on a network of continuously operating long-term water level stations in the U.S. coastal areas, U.S. possessions, and the Great Lakes, which provide tidal and vertical water level. The NWLP consists of networks of long-term and short-term water-level stations and is an "end-to-end" system of data collection, quality control, data management and product delivery.

The PORTS system measures and disseminates observations and predictions of water levels, currents, salinity, and meteorological parameters (e.g., winds, atmospheric pressure, air and water temperatures) primarily in the support of maritime activities in selected U.S. shipping ports that include the following:

- Narragansett Bay
- New Haven, CT
- New York/New Jersey Harbor
- Delaware Bay and River
- Chesapeake Bay
- Tampa Bay
- Houston/Galveston
- Los Angeles/Long Beach
- San Francisco Bay
- Lower Columbia River
- Tacoma, WA
- Anchorage
- Soo Locks, MI

2.2 NWLON and PORTS DIF End Users

The end users of initial interest to the DIF project are the following models:

1. Coastal Inundation Modeling: Sea, Lake, and Overland Surges from Hurricanes (SLOSH), Advanced Circulation Hydrodynamic Model (ADCIRC)

2. Hurricane Intensity Modeling: Hurricane Model and Hurricane Weather Research and Forecasting (HWRF) model

2.3 NWLON and PORTS Data Products

NWLON and PORTS provide core IOOS variables of sea level, currents, water temperature, and salinity with products including the following:

- 6 minute water levels, raw data
- 6 minute water levels, verified data
- Hourly water levels, verified data
- 6 minute High/Low water levels, verified data
- Daily High/Low water levels, verified data
- Monthly High/Low water levels, verified data
- 6 minute tide predictions
- 6 minute water temperature, raw data
- 6 minute salinity (conductivity), raw data

CO-OPS presently makes its data products available in the following formats:

- CREX bulletins (Character based Representation and Exchange) (WMO standard used on the GTS) available to NWS
- SHEF bulletins (Standard Hydrometeorological Exchange Format) available to NWS
- Space delimited ASCII files available to specified users on the CO-OPS FTP site
- Comma delimited OPeNDAP available through the NDBC DODS system
- XML formatted data available through Web Services.

2.3.1 NWLON and PORTS Data Discovery and Metadata

For NWLON, CO-OPS generates metadata in Federal Geographic Data Committee (FGDC) format that describes the major data parameters of water level trends, sea level trends and

currents. They also capture metadata for collection stations and sensors, but this metadata is not provided in accordance with any formal standards. ASCII data has self-describing column headings and OPeNDAP provides description boxes that contain metadata information about particular parameters.

Part of the effort underway to achieve interoperability is to meet basic compliance with FGDC standards for station and sensor level metadata. CO-OPS does not post any metadata or data to any clearinghouses.

CO-OPS also does not use a formal vocabulary. The current vocabulary for NWLON uses the names of the tables in the relational database to describe and access the station metadata.

2.3.2 NWLON and PORTS Data Transform and Assembly

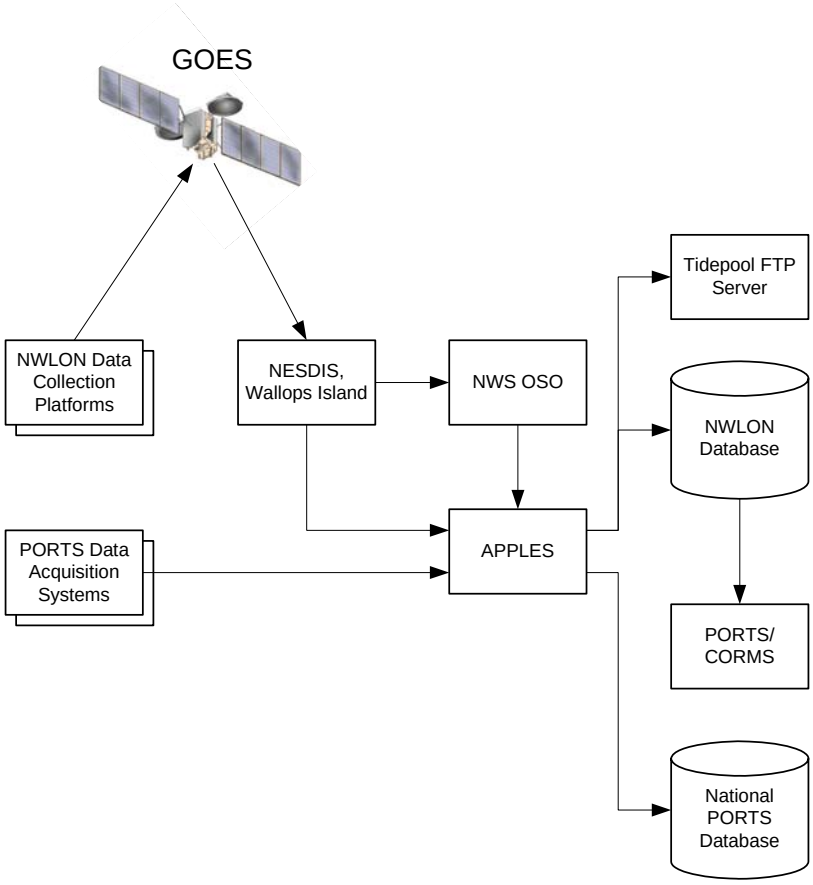
NWLON data are transmitted from the Data Collection Platforms (DCP), which are the water level gauges, to the GOES system and into National Environmental Satellite, Data, and Information Service (NESDIS). The data are then sent to the NWS Office of System Operations (OSO). The CO-OPS APPLES (Acquire, Prepare, Process and Load Environmental System) processing sub-system in Silver Spring polls NESDIS and the NWS OSO for the files and decodes them and performs quality control processing on the data and generates various output formats for inclusion in the NWLON database and Tidepool FTP server. CREX bulletins are generated and placed where NWS OSO can access them and download them for distribution to NWS field offices.

PORTS data are collected from numerous sensing stations in a particular location by a local PC-based Data Acquisition System (DAS). The DAS formats the information into PORTS Uniform Flat File Format (PUFFF) files and sends these files via FTP to the APPLES ingest system for QC, processing, and insertion into the PORTS and NWLON databases and are routed through the Continuous Operational Real-Time Monitoring System (CORMS) for access via the CO-OPS website.

The data are also formatted into CREX bulletins, ASCII files, and SHEF bulletins.

Figure 1 below depicts the simplified flow of data for NWLON and PORTS data from the collection systems through to the end-user access points.

Figure 1 – NWLON and PORTS Data Flow Diagram



2.3.3 NWLON and PORTS Data Transport and Access

Once the data are available in the CO-OPS NWLON database, primary access to the data by end users is through the Internet. ASCII-based observations are available on the CO-OPS website at http://tidesandcurrents.noaa.gov/ on a station-by-station basis for users to save or copy data. OPeNDAP data are available through the NDBC DODS system (described in a separate section).

CREX bulletins are posted to a site available to specified users, such as NWS, which forwards them through the NWS Telecommunications Gateway (NWSTG) for access by NWS field offices. ASCII files are posted to the Tidepool FTP server for interagency access, as are SHEF files for access by the Office of Hydrology. Data are also inserted into the NWLON database for access through the CO-OPS website.

Water level data are served through the OPeNDAP server from the NWLON relational databases. Basic web services, using SOAP via JAVA, are now available for users to down

load these same data. The data are returned to the users in an XML schema with an XML style sheet allowing verification of the schema by users.

CO-OPS also accommodates special requests from certain users by providing user specific files containing subsets of ASCII data that are available for FTP download from the CO-OPS website.

Other access methods include FTP and model NetCDF files. Many routine requests by users are for historical data since they are not in the active databases.

2.3.4 NWLON and PORTS Data Storage and Archive

Metadata and data in the databases are routinely backed up to tapes for long-term storage and retrieval, but are not removed from the primary server. As of May 2006, CO-OPS was planning to create a Data Assembly Center (DAC) that will contain data and associated metadata older than two years rolled off from the primary database server.

CO-OPS data and metadata are not archived to a formal data archiving center. When FGDC metadata compliancy is reached, data and metadata will be sent to a formal archiving center. Once real-time data are being routinely archived, CO-OPS plans to assess archiving historical data.

CO-OPS backs up its primary databases with full backups occurring weekly and incremental backups occur daily. Tapes are rotated on a monthly basis to an off-site storage vault in SSMC-3 of the Silver Spring Campus as well as in a storage vault at the CO-OPS Pacific Regional Office in Seattle, WA.

2.3.5 NWLON and PORTS Data Quality Control

NWLON and PORTS data are run through an automated quality control (QC) routine running within the APPLES (Acquire, Prepare, Process and Load Environmental System) quality control and data processing sub-system.

The automated QC checks consist of min/max checks, rate of change analyses, and comparisons to predictions and alternate sensors. As of May 2006, CO-OPS was developing enhancements to these basic QC tools using rule and case based reasoning technology (CORMS AI – Continuous Operational Real-Time Monitoring System Artificial Intelligence). A prototype system was developed and tested using the National PORTS® database (npdb). A production application based on this prototype was under development and CO-OPS had plans to integrate CORMS AI into the next generation IQC sub-system to monitor both NWLON and PORTS data flows.

In addition to the automated QC, CO-OPS employs ‘watch standers’ to visually inspect the data plots for erroneous values. There is a watch stander on duty 24 hours a day, 7 days a week. If a sensor fails visual or automated QC checks then the sensor’s data stream is ‘turned off’ for real-time use until the error can be validated or the instrument is fixed or repaired.

2.4 NWLON and PORTS Data Product Information Sheets

The following data product information sheets comprise tables of relevant data about each product in use, or planned for use, by the initial DIF customer models.

Table 1 – NWLON and PORTS Data Product Information Sheets

System name: NWLON				
1	**System Information:**			
1.1	Responsible Program, Line Office (CO-OPS, NDBC, NODC, etc.)	Center for Operational Oceanographic Products and Services (CO-OPS)		
1.2	Program/Line Office POC			
1.3	Technical/Data POC	Andrea Hardy (Andrea.Hardy@noaa.gov)		
1.4	CASANOSA system name and IMS name	NWLON		
1.5	NOAA system name and ID			
1.6	OMB 300? Or which one do you fall under			
1.7	Variables (from list of 20 core variables) collected into system	Water Level, sea surface temperature, salinity, wind speed/direction, barometric pressure.		
1.8	System Function			
2	**Data Product Information:**			
2.1	Data Product title	1) NWLON	2) NWLON	3) NWLON
2.1.1	Related IOOS core variables	Sea (Water) Level	Sea Surface Temperature	Salinity
2.1.2	Data product format (comma delimited ASCII, XML, NetCDF, GRIB, BUFR, etc)	Space delimited ASCII, comma delimited ASCII, XML, CREX and SHEF bulletins	Space delimited ASCII, comma delimited ASCII, XML, CREX and SHEF bulletins	Space delimited ASCII, comma delimited ASCII, XML, CREX and SHEF bulletins
2.1.3	Data latency for each data product version	6-min raw: 6-min verified: Hourly verified: Daily verified: Mthly verified:	6-min raw: 6-min verified: Hourly verified: Daily verified: Mthly verified:	6-min raw: 6-min verified: Hourly verified: Daily verified: Mthly verified:
2.1.4	Data product frequency (e.g. once an hour, twice a day, etc.)	6-min, hourly, daily, monthly	6-min, hourly, daily, monthly	6-min, hourly, daily, monthly
2.2	Input/Source Data Attributes			
2.2.1	Sensor/platform name(s)	NWLON stations	NWLON stations	NWLON stations
2.2.2	Sampling rate / time frequency (How often are the data collected?)	Every 6 minutes	Every 6 minutes	Every 6 minutes
2.2.3	Time coverage (start and end date/time)			

2.2.4	Spatial resolution (50 km, etc)	N/A	N/A	N/A
2.2.5	Data type (Grids, Profiles, Time Series, Points Data, etc)	Time series	Time series	Time series
2.2.6	Data format (comma delimited ASCII, XML, GRIB, BUFR, etc)	Space delimited ASCII, comma delimited ASCII, XML, CREX and SHEF bulletins	Space delimited ASCII, comma delimited ASCII, XML, CREX and SHEF bulletins	Space delimited ASCII, comma delimited ASCII, XML, CREX and SHEF bulletins
2.2.7	Geographic location key words (GOM, etc)			
2.2.8	Boundaries of coverage			
2.2.9	Units of measure (F vs C, etc)	Feet (to two decimal places; referenced to local datum)	Deg F	psu (practical salinity units)
2.2.10	Original source (e.g., Univ of MS)	NOS/CO-OPS operated stations	NOS/CO-OPS operated stations	NOS/CO-OPS operated stations
2.2.11	Transport method into the system (T1, Internet, private network, etc)	GOES satellite	GOES satellite	GOES satellite
2.2.12	Reference to QC procedures for each datastream (e.g., latency: real time, 6 min values vs week later; calibration; etc)			
2.3	Metadata			
2.3.1	Metadata description available? (Yes / No)	Yes (FGDC format for water level trends)		
2.3.2	Metadata posted to clearing house? (Yes / No)	No	No	No
2.3.2.1	Name of clearinghouse (if applicable)			
2.3.2.2	Name of metadata record (if applicable)			
2.3.3	Using a standard parameter vocabulary? (Yes / No)	No	No	No
2.3.3.1	Which standard vocabulary is used?			
2.3.3.2	Reference to standard vocabulary			
2.4	Data Product Dissemination			
2.4.1	Transport method(s) from system to customer (T1, Internet, etc)	Public Internet	Public Internet	Public Internet
2.4.1.1	Output data format(s) for each transport method (e.g. ASCII, XML, GRIB, BUFR, etc.)	Space delimited ASCII, comma delimited ASCII, XML, CREX and SHEF	Space delimited ASCII, comma delimited ASCII, XML, CREX and SHEF	Space delimited ASCII, comma delimited ASCII, XML, CREX and SHEF

		bulletins	bulletins	bulletins
2.4.1.2	Data volume (e.g. MBytes per product)			
2.4.2	How dataset is made available (OpenDap, ftp, http, WMS, etc)	OPeNDAP, SOAP, HTML	OPeNDAP, SOAP, HTML	OPeNDAP, SOAP, HTML
2.4.3	Data product access URL (access point)	http://tidesandcurrents.noaa.gov	http://tidesandcurrents.noaa.gov	http://tidesandcurrents.noaa.gov
2.4.4	Reference documentation URL			
2.4.5	Other known processing locations?	NDBC	NDBC	NDBC
2.4.5.1	Secondary provider URL (if applicable)	http://www.ndbc.noaa.gov/	http://www.ndbc.noaa.gov/	http://www.ndbc.noaa.gov/
2.5	IT Security			
2.5.1	Are data available to public?	Yes	Yes	Yes
2.5.2	Is authorization required to view/release data? (e.g. user or subscription account is required)	For some data products	For some data products	For some data products
2.5.3	What are controls used to prevent malicious attacks?			
2.6	Data Integrity			
2.6.1	What are used to ensure transmitted data equal received data?			
2.7	Data Archive			
2.7.1	Is data product being archived?	Yes	Yes	Yes
2.7.2	Where is data product archived?	Tape (formal archive center when FGDC compliancy is reached)	Tape (formal archive center when FGDC compliancy is reached)	Tape (formal archive center when FGDC compliancy is reached)
2.7.3	What is the data archive access URL?			
2.7.4	What is the reference documentation URL?			

3 National Data Buoy Center (NDBC)

The NDBC collects, processes, and formats meteorological and oceanographic data from the NWS-Buoy and the Coastal-Marine Automated Network (C-MAN) systems. These systems are summarized below along with the relevant data products from each. NDBC is currently working on achieving interoperability with the IOOS and has completed the NDBC's IOOS DMAC Interoperability Plan, *National Weather Service Buoys and Coastal-Marine Automated Network (C-MAN)*, dated May 22, 2006, and much of the content of this section comes from that document.

3.1 NWS Buoy and C-MAN System Description

The primary functions of the buoy and C-MAN systems provide are to perform primary data assembly, quality control and distribution services for surface marine weather and ocean observations generated by NDBC platforms, other NOAA and federal partners, and IOOS partner regional observing systems. The buoys and C-MANS are the primary systems for observing in-situ, real-time weather and ocean observations in coastal waters.

3.2 NWS Buoy and C-MAN End Users

The end users of initial interest to the DIF project are the following models:

1. Hurricane Intensity Modeling: Hurricane Model and Hurricane Weather Research and Forecasting (HWRF) model –sea surface (and subsurface) temperature, salinity, sea level.

2. Integrated Ecosystem Assessments Modeling - sea surface (and subsurface) temperature, currents.

3. Harmful Algal Bloom Forecasting and Assessments – through NDBC - observed winds from selected NWLON platforms in the Gulf of Mexico

3.3 NWS Buoy and C-MAN Data Products

NWS buoys and the C-MAN produce many data based on a wide range of collected parameters. For the purposes of the DIF, the NDBC generates data sets and products that include four of the five core variables: currents, salinity, temperature, and water level.

The NDBC operates a total of 102 variously sized, moored buoys with some variation in deployed (and operational) data collection payloads and operates 56 C-MAN stations with varying payloads (mainly based on whether the station is deployed in contact with the ocean or not and can therefore report sea surface temperature and water levels).

The NDBC also aggregates and makes available data from numerous other observing platforms including NWLON, COMPS, NERRS, HF Radar, and others.

Individual sensor platform reports are real-time and are reported every six minutes.

Data products are available in three formats along with archived data as follows:

- WMO formats are used for data that are distributed through the NWS Telecommunications Gateway to various government and commercial users as well as wider distribution through the WMO Global Telecommunications System (GTS). In particular, temperature, salinity, and current data are encoded in accordance with the FM64 format described in WMO Manual on Codes, WMO No.-306.

- Standardized ASCII formats developed for the NDBC website are used to display real-time and historical data available on the website. The units of measure, parameter names, and processes are well defined on the NDBC website. Instruments and quality control procedures are defined in a general sense on the website.

- NetCDF formatted data are available from the NDBC OPeNDAP server using the Cooperative Ocean/Atmosphere Research Data Service (COARDS) convention and the Climate and Forecast (CF) NetCDF standard names are used for units of measure and parameter names.

- Archived data are encoded in the F291 format with well defined units of measure and parameter names.

3.3.1 NWS Buoy and C-MAN Data Discovery and Metadata

At the basic database level, NDBC does not adhere to any specific metadata standards nor does it post metadata to any clearinghouse. NDBC has its own form of metadata for data available on its website that originates in the NDBC database. These consist of short ASCII identifiers corresponding to an internal set of ASCII descriptors, which have been gradually developed when needed as new data fields are added. Definitions of the root internal identifiers are maintained within the database and the short identifiers are described online at http://www.ndbc.noaa.gov/measdes.shtml .

The “standard names” for variables within data files available via the NDBC OPeNDAP server adhere to the Climate and Forecast (CF) NetCDF convention. This metadata includes only station name, location, and unit of measure for each collected variable, however the NetCDF convention has the potential to convey a great deal more metadata.

Data that are forwarded to the National Centers for archive are encoded in the F291 format. Some basic metadata are included with within this format such as station location, water depth and sensor heights. For some, but not all variables, the sampling rates and sampling durations at which they were measured are included

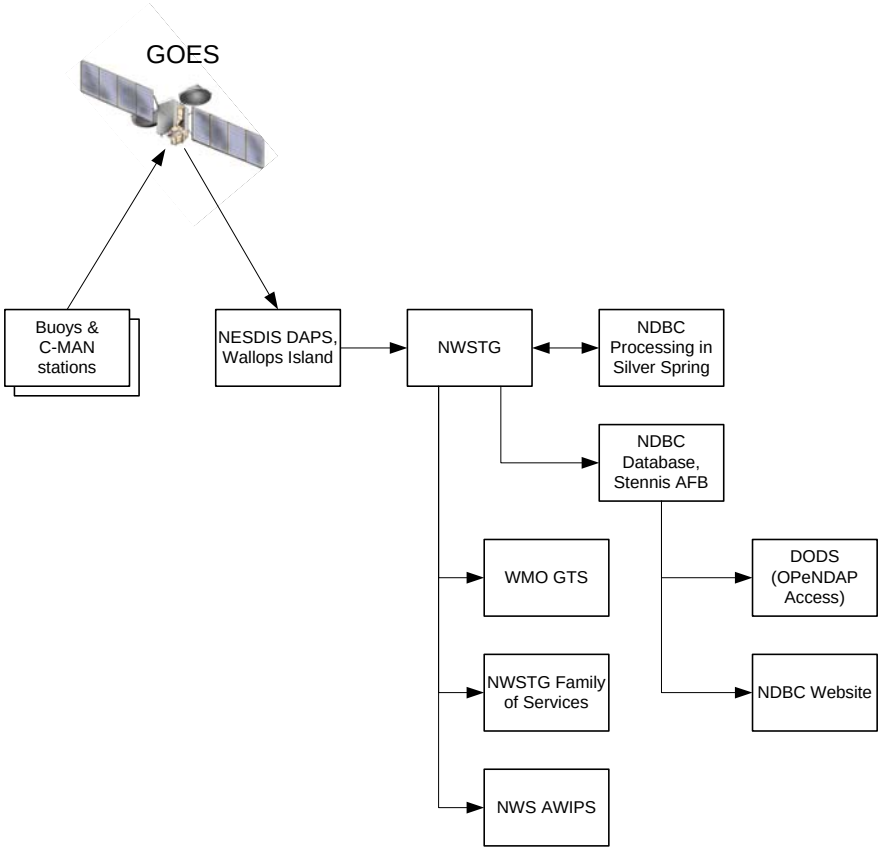
3.3.2 NWS Buoy and C-MAN Data Transform and Assembly

Buoy and C-MAN data are transmitted from the collection platforms via GOES to the NESDIS Data Acquisition and Processing System (DAPS) at Wallops Island, VA. The DAPS sends the data to the NWSTG where NDBC computer systems process the raw GOES data. These systems decode the data, perform automated QC checks, and generate reports in standard World Meteorological Organization (WMO) format. These real-time reports are released in collective bulletins to users via the NWS Family of Services (FOS) network, the Global Telecommunications System (GTS), and within the NWS through the Advanced Weather Interactive Processing System (AWIPS) via NOAAPort.

Private bulletins are generated at the NWSTG to update real-time observations on the NDBC website and to transmit the processed data and quality control flags to NDBC. These bulletins contain the complete set of data acquired by the payload and are used to update the NDBC Oracle database in real-time.

Figure 2 below depicts the basic data flow for NDBC buoys and C-MAN stations from the sensing platforms through to the end user access points.

Figure 2 – NDBC Buoy and C-MAN Data Flow Diagram



3.3.3 NWS Buoy and C-MAN Data Transport and Access

Data transport is through three basic methods already mentioned: through the NWSTG Family of Services, through the WMO GTS, from the NDBC public website using a web browser, and through the NDBC OPeNDAP server, i.e., the DODS, using an OPeNDAP client application.

NDBC reported that the WMO format is most commonly used by Government, commercial, and modeling communities while the ASCII data on the website are accessed by a wide range of users from the general public to meteorological researchers, whereas the OPeNDAP system is clearly the least common method of access.

3.3.4 NWS Buoy and C-MAN Data Storage and Archive

NDBC maintains its own data archive for all station data including raw measurements, derived data, system monitoring data, and station metadata. This archive is only available internally at the NDBC.

Web accessible historic files are maintained on NDBC servers in ASCII format and in NetCDF files that are accessible via the OpenDAP server.

NDBC archives data on a monthly basis at the National Archive Centers (NCDC and NODC) in F291 format. The NCDC archive consists only of atmospheric data.

3.3.5 NWS Buoy and C-MAN Data Quality Control

The NDBC employs fully automated and human-machine mixed quality control measures. The automated process is performed within the NWSTG to eliminate gross errors by checking transmission parity error, range limit, and time continuity. When such errors are identified, the data are not released in real-time or posted to the website. Less stringent automated checks are also performed to assist in identifying sensors with degraded performance.

Data quality analysts at NDBC access the database to note the occurrence of flagged data and to conduct further quality control. The human-machine mix of quality checks use techniques such as inspection of time series plots, analysis of spectral wave curves, and comparison of observations with numerical models. When a sensor or system degradation is detected in post-real-time, the affected data are removed before posting on the NDBC website or archive.

Indication of quality control checks is generally not used since the data are either verified and released, or did not pass QC and are not released or made available externally.

NDBC recognizes the need for quality indicators. Currently, ADCP data from some IOOS partners on the website include quality flags adhering to the QARTOD standard and the corresponding NetCDF files also contain quality flags as a variable.

3.4 NWS Buoy and C-MAN Data Product Information Sheets

The following data product information sheets comprise tables of relevant data about each product in use, or planned for use, by the initial DIF customer models.

Table 2 – NWS Buoy and C-MAN Data Product Information Sheets

System name: NWS Buoys and C-MAN Stations				
1	**System Information:**			
1.1	Responsible Program, Line Office (CO-OPS, NDBC, NODC, etc.)	National Data Buoy Center		
1.2	Program/Line Office POC	Dan Henderson		
1.3	Technical/Data POC	Bill Burnett		
1.4	CASANOSA system name and IMS name	Buoys; C-MAN		
1.5	NOAA system name and ID			
1.6	OMB 300? Or which one do you fall under			
1.7	Variables (from list of 20 core variables) collected into system	Wind speed and direction, wind gusts, wave height, period, and direction, air pressure and pressure tendency, air temperature, water temperature, dew point, salinity, tide level. NOTE: not all variables observed and reported by each station.		
1.8	System Function			
2	**Data Product Information:**			
2.1	Data Product title	1) Buoy observations	2) C-MAN observations	3)Other Platforms
2.1.1	Related IOOS core variables	Water Temperature, Salinity	Water Temperature, Salinity, Water Level	Water temperature, water level, Currents
2.1.2	Data product format (comma delimited ASCII, XML, NetCDF, GRIB, BUFR, etc)	ASCII, WMO FM64 via GTS, NetCDF using Climate and Forecast (CF) naming.	ASCII, WMO FM64 via GTS, NetCDF using Climate and Forecast (CF) naming.	ASCII, WMO FM64 via GTS, NetCDF using Climate and Forecast (CF) naming.
2.1.3	Data latency for each data product version	Near real time	Near real time	Near real time
2.1.4	Data product frequency (e.g. once an hour, twice a day, etc.)	6 min, hourly	6 min, hourly	Varied
2.2	Input/Source Data Attributes			
2.2.1	Sensor/platform name(s)	Moored buoys	C-MAN	COMPS, NWLON, NERRS, HF radar, MMS ADCP, TGLO
2.2.2	Sampling rate / time frequency (How often are the data collected?)	Every 6 minutes	Every 6 minutes	
2.2.3	Time coverage (start and end date/time)	Continuous	Continuous	

2.2.4	Spatial resolution (50 km, etc)	Varied	Varied	
2.2.5	Data type (Grids, Profiles, Time Series, Points Data, etc)	Time series		
2.2.6	Data format (comma delimited ASCII, XML, GRIB, BUFR, etc)			
2.2.7	Geographic location key words (GOM, etc)			
2.2.8	Boundaries of coverage			
2.2.9	Units of measure (F vs C, etc)	°F, psu	°F, psu, ft above MLLW	°F, ft above MLLW, knots
2.2.10	Original source (e.g., Univ of MS)	NDBC	NDBC	NOS, USF, MMS, NERRS, Texas, Louisiana, various oil companies
2.2.11	Transport method into the system (T1, Internet, private network, etc)	GOES	GOES	Various
2.2.12	Reference to QC procedures for each datastream (e.g., latency: real time, 6 min values vs week later; calibration; etc)			
2.3	Metadata			
2.3.1	Metadata description available? (Yes / No)	Yes	Yes	
2.3.2	Metadata posted to clearing house? (Yes / No)	No	No	
2.3.2.1	Name of clearinghouse (if applicable)			
2.3.2.2	Name of metadata record (if applicable)			
2.3.3	Using a standard parameter vocabulary? (Yes / No)	No; NDBC developed; short naming convention for OPeNDAP accessible files adheres to CF NetCDF convention.	No; NDBC developed; short naming convention for OPeNDAP accessible files adheres to CF NetCDF convention.	
2.3.3.1	Which standard vocabulary is used?			
2.3.3.2	Reference to standard vocabulary	http://www.ndbc.noaa.gov/measdes.shtml	http://www.ndbc.noaa.gov/measdes.shtml	
2.4	Data Product Dissemination			
2.4.1	Transport method(s) from system to customer (T1, Internet, etc)	Public Internet, NWSTG Family of Services, WMO GTS	Public Internet, NWSTG Family of Services, WMO GTS	

2.4.1.1	Output data format(s) for each transport method (e.g. ASCII, XML, GRIB, BUFR, etc.)			
2.4.1.2	Data volume (e.g. MBytes per product)			
2.4.2	How dataset is made available (OpenDap, ftp, http, WMS, etc)	ASCII, OPeNDAP (CF NetCDF) through DODS	ASCII, OPeNDAP (CF NetCDF) through DODS	
2.4.3	Data product access URL (access point)	http://www.ndbc.noaa.gov/	http://www.ndbc.noaa.gov/	
2.4.4	Reference documentation URL			
2.4.5	Other known processing locations?			
2.4.5.1	Secondary provider URL (if applicable)			
2.5	IT Security			
2.5.1	Are data available to public?	Yes		
2.5.2	Is authorization required to view/release data? (e.g. user or subscription account is required)			
2.5.3	What are controls used to prevent malicious attacks?			
2.6	Data Integrity			
2.6.1	What are used to ensure transmitted data equal received data?			
2.7	Data Archive			
2.7.1	Is data product being archived?	Yes	Yes	
2.7.2	Where is data product archived?	NDBC and National Archive Centers (NCDC and NODC)	NDBC and National Archive Centers (NCDC and NODC)	
2.7.3	What is the data archive access URL?			
2.7.4	What is the reference documentation URL?			

4 Global Temperature and Salinity Profile Program (GTSPP)

The Global Temperature and Salinity Profile Program (GTSPP) is a joint international project to develop and maintain a global ocean temperature and salinity resource with high quality, up-to-date ocean temperature and salinity data. It is a joint project between the WMO and the Intergovernmental Oceanographic Commission (IOC). Functionally, GTSPP reports to the Joint Commission on Oceanography and Marine Meteorology (JCOMM), a body sponsored by WMO and IOC and to the IOC's International Oceanographic Data and Information Exchange committee (IODE). Canada's Marine Environmental Data Service (MEDS) leads the project and has the operational responsibility to gather and process real-time data.

The role of GTSPP is production of two types of data and information products. The first is a set of end-user products such as maps and inventories of temperature and salinity data, statistics on GTS data flow to assist the data collectors, and maps of the distribution of surface and sub-surface temperatures from ships of opportunity and moored and drifting buoys. The second type consists of higher quality, QC'ed data sets provided in a delayed mode.

4.1 GTSPP System Description

The GTSPP aggregates data from numerous collection platforms into its Continuously Managed Database (CMD) resident at the NODC. The MEDS accumulates real-time temperature and salinity data from several sources via the GTS and checks the data for several types of errors, removes duplicate observations, and passes the data to the US National Oceanographic Data Center (NODC) for inclusion in the CMD. Collection platforms include ships at sea, fixed platforms (buoys), and profiling floats. These data sets are transferred to the NODC three times a week.

The GTS carries real-time operational data from ships and buoys in support of the IOC/WMO Integrated Global Ocean Services System (IGOSS) and these data are available to any organization connected to the GTS. Data are typically available through this method within a few minutes to a few hours after collection. These data are not quality controlled as they do not pass through GTSPP.

The MEDS also receives delayed mode data that come from data collectors first through various national and international centers. These data are received by the MEDS within a few days to a month of collection. Data are quality controlled and duplicates are removed, then forwarded to the NODC CMD. The CMD maintains temperature and salinity data, replacing near real-time records with higher quality delayed-mode records as they are received.

4.2 GTSPP DIF End Users

The end users of initial interest to the DIF project are the following models:

1. Integrated Ecosystem Assessment Modeling

4.3 GTSPP Data Products

The GTSPP provides the core IOOS variables of temperature and salinity. Aside from the finished products produced by GTSPP such as temperature and salinity maps, a range of high-quality, delayed mode data sets are available to the modeling community through the NODC GTSPP website as well as real-time data. Data are available in both ASCII and NetCDF formats; however, real-time data are only available in ASCII format.

Data sources include the following:

- BOT (Bottle Station Data)
- CTD (Conductivity-Temperature-Depth)
- MBT (Mechanical Bathy Thermograph)
- XBT (Expendable Bathy Thermograph)
- BA (GTS BATHY Message Data: Real-Time XBT)
- TE (GTS TESAC Message Data: Real-Time CTD, Floats, Buoys)
- PROFILING FLOATS - ARGO
- PROFILING FLOATS - PRE-ARGO
- FIXED BUOYS - GOMOOS
- FIXED BUOYS - PIRATA
- FIXED BUOYS - TOGA-TAO
- FIXED BUOYS - TRITON FIXED BUOYS - OTHER

These sources can be individually selected when a user accesses data from the GTSPP website through the User-Defined Data Sets option, which also allows sub-setting by geographic region, LAT/LON boundary, and period of time. Access to Real-Time Data Sets and Best Copy Data Sets yields access to consolidated files containing all of the data. Real-time data allow sub-setting by period of time (back to Jan 2005) and by either station location plot or full file download. Best copy data sets allow sub-setting by period of time (back to Jan 1990), by ocean (Atl, Pac, or Ind), and by either station location plot or full file download.

User defined and real-time data are available in ASCII text or GZIP compressed file. Best copy data are available in ASCII and NetCDF formats.

4.3.1 GTSPP Data Discovery and Metadata

The MEDS formats data into ASCII files using a format developed in-house that makes use of several international codes to describe the data and how they were collected. GTSPP publishes the ASCII data format on its website, and this acts essentially as the metadata descriptor. It is not apparent that GTSPP publishes metadata elsewhere with a clearinghouse. The codes that form the basis of the ASCII format are listed in their entirety on the GTSPP website and include the following categories of codes:

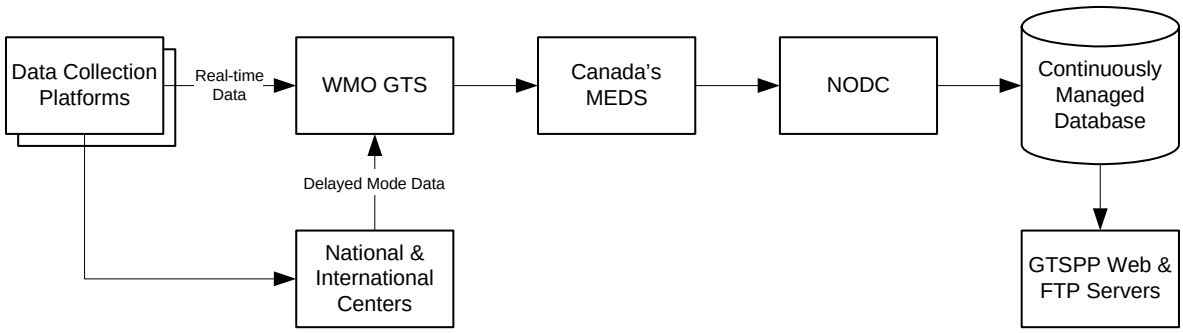
- GTSPP Codes: The data format makes use of a number of different codes to describe the data and how they were collected. Quality flags are described here and in the GTSPP quality control documentation.
- WMO Codes: WMO codes are also used to describe the data and how they were collected and for core variables include Code Table 1770: Ix Ix Ix - Instrument type for temperature, with fall rate equation coefficients for XBT.

- Ship call signs: This table gives information linking ship call signs (used when sending real-time data) to ship names and NODC ship codes.
- Parameter Codes: Listing of Parameter Code Meaning in the ASCII and NetCDF format files

4.3.2 GTSPP Data Transform and Assembly

Raw data are formatted for transfer through the GTS by the data collection systems for transport to various national and international scientific centers for delayed data and over the GTS to the MEDS for near real-time data. Real-time data collected at the MEDS are quality controlled for gross errors and then forwarded to the NODC in Silver Spring three times per week. The NODC maintains the data in its Continuously Managed Database (CMD). Delayed mode data are sent from the various centers to the NODC for inclusion in the CMD. During the merging of the data into the CMD, any duplicates occurring between near real-time and delayed mode data sources will be identified with the highest resolution duplicate being retained as the active CMD copy. Figure 3 below depicts the simplified data flow for GTSPP data.

Figure 3 –GTSPP Data Flow Diagram



4.3.3 GTSPP Data Transport and Access

Data transport and access by end users is through the Internet from the NODC GTSPP website and FTP site. Users can access data through the GTSPP website with conventional browsers where they can select various data sets in near real-time as ASCII downloads and delayed mode data (best copy data) as ASCII or NetCDF downloads. These data sets are also available through FTP transfer using web browsers or other specific FTP clients, which may facilitate automated retrieval.

4.3.4 GTSPP Data Storage and Archive

The GTSPP Continuously Managed Database (CMD) acts as the GTSPP data storage and archive repository. From the NODC GTSPP website, near real-time data are available back through 2005 and best-copy data are available back through 1990.

4.3.5 GTSPP Data Quality Control

Data quality control is performed in accordance with GTSPP Manuals and Guides #22, GTSPP Real-Time Quality Control Manual, and every day the profile has a number attached indicating the version of the Quality Control Manual describing tests performed.

The Quality Control Manual describes the parts of the quality control process including pre- and post-processing and the extensive set of actual quality control tests. Pre-processing consists of duplicate identification and removal and the algorithm for this process is presented in Annex A of the Quality Control Manual. Post-processing consists of various scientific assessments of the data quality performed by various centers as well as more sophisticated statistical tests.

The bulk of quality control processing is presented in Annex B of the Quality Control Manual and consists of the following five stages:

Stage 1: Location and Identification Tests. Determines that the position, the time, and the identification of the profile are sensible.

Stage 2: Profile Tests. Resolves and possible values for variables.

Stage 3: Climatology Tests. Examines the consistency of the incoming data with respect to references such as climatologies.

Stage 4: Profile Consistency Tests. Examines the internal consistency within a data set.

Stage 5: Visual Inspection. Manually inspects data after all other tests to ensure no questionable data values pass through the suite of tests without being detected.

The GTSPP QC process makes extensive use of flags to indicate data quality. GTSPP participants agreed to make data available to users on selectable basis using quality flags, as well as other criteria. Quality flags are always included with any data transfers that take place. GTSPP participants have also agreed to retain copies of original data and make these data available to users if requested. Therefore, a user can expect GTSPP participants to disseminate data at any stage of processing. Once a data set as successfully passed the first levels of checks (which consist of relatively simple data structure, format, and range-related tests), and has been transferred into the CMD, it will not be removed or deleted from the CMD except in the case of exact duplicates. Upon further quality control, the same data set, may be added to the CMD, however as a new version with an updated quality control flag.

As data are acquired in both near real-time and delayed mode, they are added to the database. Delayed mode data have a higher resolution and are calibrated and quality controlled by the originator. Thus, the delayed mode data represent a “better” version of the data for all purposes. This “better” version replaces the data obtained in near real-time. This “better” version of the data arrives later than the real-time version. The Continuously Managed Database therefore holds the most current and highest quality dataset available at any given time. It will be continuously refined as additional quality checks are undertaken. The term

"replace" here means replace as the active copy of the observation in the database. Observations that have passed quality control and entered the database are not removed. They are flagged to indicate that a higher quality version of the observation exists in the database.

4.4 GTSPP Data Product Information Sheets

The following data product information sheets comprise tables of relevant data about each product in use, or planned for use, by the initial DIF customer models.

Table 3 – GTSPP Data Product Information Sheets

System name: Global Temperature and Salinity Profile Program			
1	**System Information:**		
1.1	Responsible Program, Line Office (CO-OPS, NDBC, NODC, etc.)	National Oceanographic Data Center	
1.2	Program/Line Office POC	Charles Sun	
1.3	Technical/Data POC		
1.4	CASANOSA system name and IMS name	GTSPP	
1.5	NOAA system name and ID		
1.6	OMB 300? Or which one do you fall under		
1.7	Variables (from list of 20 core variables) collected into system	Sea surface temperature, salinity	
1.8	System Function		
2	**Data Product Information:**		
2.1	Data Product title	1) GTSPP	2) GTSPP
2.1.1	Related IOOS core variables	Sea Surface Temperature	Salinity
2.1.2	Data product format (comma delimited ASCII, XML, NetCDF, GRIB, BUFR, etc)	ASCII, NetCDF	ASCII, NetCDF
2.1.3	Data latency for each data product version	Days to months	Days to months
2.1.4	Data product frequency (e.g. once an hour, twice a day, etc.)	3 times wkly, monthly as available	3 times wkly, monthly as available
2.2	Input/Source Data Attributes		
2.2.1	Sensor/platform name(s)	Various ships, buoys, and profilers	Various ships, buoys, and profilers
2.2.2	Sampling rate / time frequency (How often are the data collected?)		
2.2.3	Time coverage (start and end date/time)		
2.2.4	Spatial resolution (50 km, etc)		
2.2.5	Data type (Grids, Profiles, Time Series, Points Data, etc)		

2.2.6	Data format (comma delimited ASCII, XML, GRIB, BUFR, etc)	ASCII iaw MEDS formatting	ASCII iaw MEDS formatting
2.2.7	Geographic location key words (GOM, etc)		
2.2.8	Boundaries of coverage		
2.2.9	Units of measure (F vs C, etc)		
2.2.10	Original source (e.g., Univ of MS)	MEDS, various national and international scientific centers through IODE and IGOSS	MEDS, various national and international scientific centers through IODE and IGOSS
2.2.11	Transport method into the system (T1, Internet, private network, etc)	GTS	GTS
2.2.12	Reference to QC procedures for each datastream (e.g., latency: real time, 6 min values vs week later; calibration; etc)		
2.3	Metadata		
2.3.1	Metadata description available? (Yes / No)	Yes	Yes
2.3.2	Metadata posted to clearing house? (Yes / No)	No	No
2.3.2.1	Name of clearinghouse (if applicable)		
2.3.2.2	Name of metadata record (if applicable)		
2.3.3	Using a standard parameter vocabulary? (Yes / No)	No	No
2.3.3.1	Which standard vocabulary is used?	MEDS developed in-house	MEDS developed in-house
2.3.3.2	Reference to standard vocabulary		
2.4	Data Product Dissemination		
2.4.1	Transport method(s) from system to customer (T1, Internet, etc)	Internet (web and FTP)	Internet (web and FTP)
2.4.1.1	Output data format(s) for each transport method (e.g. ASCII, XML, GRIB, BUFR, etc.)	ASCII, NetCDF	ASCII, NetCDF
2.4.1.2	Data volume (e.g. MBytes per product)	Up to ~16 MB	Up to ~16 MB
2.4.2	How dataset is made available (OpenDap, ftp, http, WMS, etc)	HTTP, FTP	HTTP, FTP
2.4.3	Data product access URL (access point)	http://www.nodc.noaa.gov/GTSPP/access_data/index.html ftp://ftp.nodc.noaa.gov/pub/gtspp/	http://www.nodc.noaa.gov/GTSPP/access_data/index.html ftp://ftp.nodc.noaa.gov/pub/gtspp/
2.4.4	Reference documentation URL		

2.4.5	Other known processing locations?		
2.4.5.1	Secondary provider URL (if applicable)		
2.5	IT Security		
2.5.1	Are data available to public?	Yes	Yes
2.5.2	Is authorization required to view/release data? (e.g. user or subscription account is required)	Not apparent	Not apparent
2.5.3	What are controls used to prevent malicious attacks?		
2.6	Data Integrity		
2.6.1	What are used to ensure transmitted data equal received data?	TCP/IP and FTP protocol-inherent delivery assurance methods	TCP/IP and FTP protocol-inherent delivery assurance methods
2.7	Data Archive		
2.7.1	Is data product being archived?	Yes	Yes
2.7.2	Where is data product archived?	Apparently in GTSPP CMD itself	Apparently in GTSPP CMD itself
2.7.3	What is the data archive access URL?		
2.7.4	What is the reference documentation URL?		

5 MODIS Aqua from NESDIS CoastWatch (OKEANOS)

The CoastWatch mission is to provide Federal, State, and local marine scientists, coastal resource managers, and the general public timely access to near real-time satellite data. These data consist of water-leaving radiance data collected by NASA's Aqua satellite. These data are processed by the OKEANOS computer system to provide ocean color data products including MODIS Aqua, which comprises a prime source of this core variable for IOOS customers.

Through collection of data from various other satellites including GOES and POES, CoastWatch also produces and distributes sea surface temperature and ocean surface winds data products

The CoastWatch Program consists of two components: Central Operations and Regional Nodes. Central Operations, managed by NESDIS, coordinates the processing, delivery, quality control, and storage of data products. The six regional nodes are made up of other NOAA line offices that participate in the CoastWatch Program. They are located around the country, hosting equipment and personnel to provide near real-time data distribution and regional scientific expertise to the local user community.

5.1 MODIS Aqua System Description

The NASA Aqua satellite continuously collects water-leaving radiance data using MODIS (Moderate Resolution Imaging Spectroradiometer. Along with all the data from other instruments on board the Terra spacecraft and Aqua Spacecraft, MODIS data are transferred to ground stations in White Sands, New Mexico, via the Tracking and Data Relay Satellite System (TDRSS). The data are then sent to the EOS Data and Operations System (EDOS) at the Goddard Space Flight Center. After Level 0 processing at EDOS, the Goddard Space Flight Center Earth Sciences Distributed Active Archive Center (GES DAAC) produces the Level 1A, Level 1B, geolocation and cloud mask products. The data are then transferred to the OKEANOS system, which produces higher level products. These are made available from the OKEANOS FTP servers and are forwarded to the CoastWatch web server.

5.2 MODIS Aqua DIF End Users

The end users of initial interest to the DIF project are the following models:

1. Hurricane Intensity Modeling: Hurricane Model and Hurricane Weather Research and Forecasting (HWRF) model

2. Harmful Algal Bloom Forecasting and Assessments

Currently, the HABs customer uses SeaWiFS data for production of the HAB bulletins and access MODIS data as a backup source. However, with uncertain access in the future to near real-time SeaWiFS data, MODIS data has become more important to the HAB bulletin process and the analysis tool has been updated to accommodate MODIS data. The new version of the analysis tool is still being tweaked to optimize performance.

5.3 MODIS Aqua Data Products and Formats

NOAA produces CoastWatch MODIS ocean color products in near-real time from global data collected by the NASA Goddard Space Flight Center (GSFC). NOAA obtains Level-0 data from NASA GSFC in 5-minute granules. These Level-0 datasets are processed to geolocated, calibrated radiances (Level 1b) and derived MODIS data products (Level 2) using NASA SeaDAS 'msl12' software. These products are then mapped to the CoastWatch geographic regions.

Through OKEANOS, CoastWatch provides twenty-one individual ocean color products falling into seven ocean color product categories including: normalized water-leaving radiances (Nlw), daily mean chlorophyll-a concentration, 61 day mean chlorophyll-a concentration, chlorophyll-a concentration anomaly, daily mean remote sensing reflectance (Rrs), 61 day mean remote sensing reflectance and remote sensing reflectance anomalies. One set of these products is generated using data from the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Aqua satellite and are covered in this section. Two sets of these products are generated using Sea-viewing Wide Field-of-view Sensor (SeaWiFS) data, one using a NOAA-unique atmospheric correction and the other using a standard atmospheric correction provided by the NASA SeaWiFS Data Analysis (SeaDAS) software modules. SeaWiFS is covered in Section 6.

CoastWatch provides the core IOOS variable of MODIS Aqua ocean color products including the following:

- Normalized water-leaving radiances (1)
- MODIS/Aqua nLw
- Chlorophyll_a mean concentrations (daily and multi-day composites) and anomalies (3)
- MODIS/Aqua Daily mean Chlorophyll_a concentration (OC3M)
- MODIS/Aqua 61-day mean Chlorophyll_a concentration (OC3M)
- MODIS/Aqua Chlorophyll_a concentration anomaly (OC3M)
- Remote Sensing Reflectances (3)
- MODIS/Aqua Daily mean Rrs667
- MODIS/Aqua 61-day mean Rrs667
- MODIS/Aqua Rrs667 anomaly

The primary data format for MODIS products is Hierarchical Data Format (HDF). Other online formats are PNG and GeoTIFF

5.3.1 MODIS Aqua Data Discovery and Metadata

CoastWatch has developed a metadata specification for its HDF products that is well described in Appendix B – *CoastWatch HDF Metadata Specification* of the *CoastWatch Software Library and Utilities: User's Guide*. The specification includes Global Metadata, Earth Location Metadata, and Variable Metadata.

5.3.2 MODIS Aqua Data Transform and Assembly

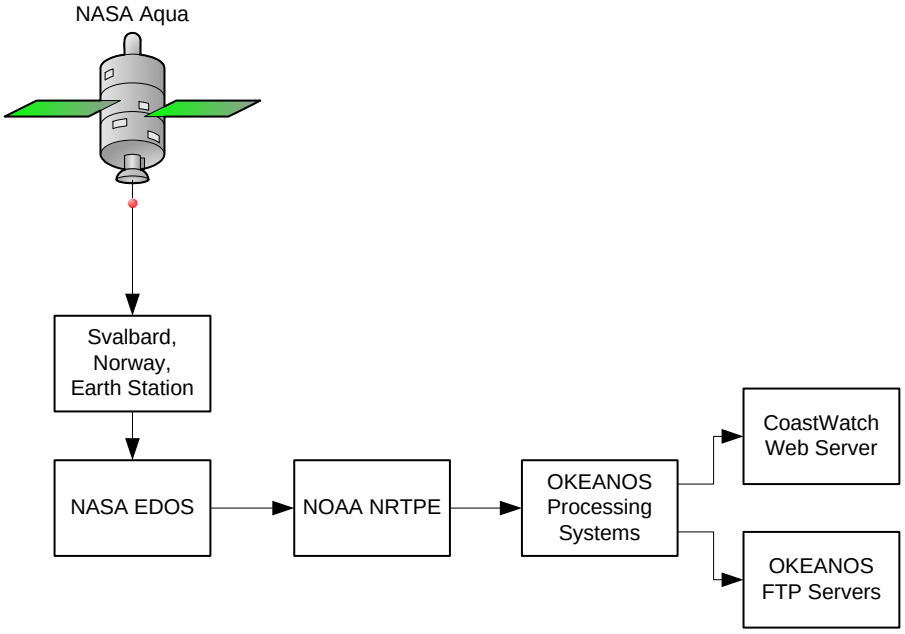
Data flow and assembly for MODIS Aqua data begins with reception of water-leaving radiance data from the NASA Aqua at an earth station in Svalbard, Norway. These data are transmitted to the NASA Earth Observing System (EOS) Data and Operations Center (EDOS) at the Goddard Space Flight Center (GSFC) in 5-minute granules. EDOS systems

transform the data into Rate Buffered Data (RBD). The RBD are sent from NASA to NOAA via FTP to the Near Real-Time Processing Effort (NRTPE) at the NOAA Satellite Operations Facility (NSOF) in Suitland, MD.

The NRTPE systems convert RBD to L0 data, then convert L0 data to L1B and L2 data. L1 and L2 data are then transferred to the OKEANOS system, also in Suitland. The OKEANOS system converts L1B data into L2, L3, and L4 data sets. The OKEANOS system FTP servers for direct access by regular customers, and also transfers data to the CoastWatch web server for public access.

Figure 5 below depicts the basic flow of data for MODIS Aqua.

Figure 4 – MODIS Aqua Data Flow Diagram



NASA also performs further processing of L1 data (not shown) and maintains the Ocean Color website with both MODIS Aqua and SeaWiFS data available.

5.3.3 MODIS Aqua Data Transport and Access

MODIS Aqua ocean color data and products are available for general customer access on the CoastWatch website via HTTP transfer over the public Internet at http://coastwatch.noaa.gov/interface/interface.html . Color data and images can also be accessed via FTP from OKEANOS servers. FTP access can be used at ftp://ftpoceanwatch.noaa.gov/pub/data_products/ as well.

5.3.4 MODIS Aqua Data Storage and Archive

To be supplied.

5.3.5 MODIS Aqua Data Quality Control

To be supplied.

5.4 MODIS Aqua Data Product Information Sheets

The following data product information sheets comprise tables of relevant data about each product in use, or planned for use, by the initial DIF customer models.

Table 4 – MODIS Aqua Data Product Information Sheets

System name: MODIS Aqua			
1	**System Information:**		
1.1	Responsible Program, Line Office (CO-OPS, NDBC, NODC, etc.)	Goddard Space Flight Center; NESDIS CoastWatch program	
1.2	Program/Line Office POC		
1.3	Technical/Data POC		
1.4	CASANOSA system name and IMS name		
1.5	NOAA system name and ID		
1.6	OMB 300? Or which one do you fall under		
1.7	Variables (from list of 20 core variables) collected into system		
1.8	System Function		
2	**Data Product Information:**		
2.1	Data Product title	1) MODIS Aqua	2)
2.1.1	Related IOOS core variables	Ocean color; chlorophyll-a concentration	
2.1.2	Data product format (comma delimited ASCII, XML, NetCDF, GRIB, BUFR, etc)	HDF	
2.1.3	Data latency for each data product version		
2.1.4	Data product frequency (e.g. once an hour, twice a day, etc.)	Level 2 – daily; level 3 – daily, 8-day, monthly, annually	
2.2	Input/Source Data Attributes		
2.2.1	Sensor/platform name(s)	MODIS (an NASA Aqua satellite)	
2.2.2	Sampling rate / time frequency (How often are the data collected?)	Continuously	
2.2.3	Time coverage (start and end date/time)		
2.2.4	Spatial resolution (50 km, etc)	1 km, 4.6 km, 36 km	
2.2.5	Data type (Grids, Profiles, Time Series, Points Data, etc)		
2.2.6	Data format (comma delimited ASCII, XML, GRIB, BUFR, etc)	HDF (primary), PNG, GeoTIFF	
2.2.7	Geographic location key words (GOM, etc)		

2.2.8	Boundaries of coverage	Global	
2.2.9	Units of measure (F vs C, etc)		
2.2.10	Original source (e.g., Univ of MS)		
2.2.11	Transport method into the system (T1, Internet, private network, etc)	Satellite, FTP	
2.2.12	Reference to QC procedures for each datastream (e.g., latency: real time, 6 min values vs week later; calibration; etc)		
2.3	Metadata		
2.3.1	Metadata description available? (Yes / No)	Yes	
2.3.2	Metadata posted to clearing house? (Yes / No)		
2.3.2.1	Name of clearinghouse (if applicable)		
2.3.2.2	Name of metadata record (if applicable)	CoastWatch HDF Metadata (http://coastwatch.noaa.gov/cw_formats.html)	
2.3.3	Using a standard parameter vocabulary? (Yes / No)		
2.3.3.1	Which standard vocabulary is used?		
2.3.3.2	Reference to standard vocabulary		
2.4	Data Product Dissemination		
2.4.1	Transport method(s) from system to customer (T1, Internet, etc)	Internet; FTP	
2.4.1.1	Output data format(s) for each transport method (e.g. ASCII, XML, GRIB, BUFR, etc.)	HDF, PNG, GeoTIFF	
2.4.1.2	Data volume (e.g. MBytes per product)		
2.4.2	How dataset is made available (OpenDap, ftp, http, WMS, etc)	HTTP; FTP	
2.4.3	Data product access URL (access point)	http://coastwatch.noaa.gov/interface/interface.html , ftp://ftpoceanwatch.noaa.gov/pub/data_products/	
2.4.4	Reference documentation URL		
2.4.5	Other known processing locations?	Yes, NASA GSFC	
2.4.5.1	Secondary provider URL (if applicable)	http://oceancolor.gsfc.nasa.gov	
2.5	IT Security		
2.5.1	Are data available to public?	Yes	

2.5.2	Is authorization required to view/release data? (e.g. user or subscription account is required)		
2.5.3	What are controls used to prevent malicious attacks?		
2.6	Data Integrity		
2.6.1	What are used to ensure transmitted data equal received data?		
2.7	Data Archive		
2.7.1	Is data product being archived?	Yes	
2.7.2	Where is data product archived?		
2.7.3	What is the data archive access URL?		
2.7.4	What is the reference documentation URL?		

6 SeaWiFS

SeaWiFS (Sea Viewing Wide Field of View Sensor) is a program operated by GeoEye (formerly ORBIMAGE, formerly Orbital Sciences Corporation) under contract with NASA to provide ocean color data. GeoEye operates the OrbView-2 satellite carrying the optical ocean color sensor in a sun-synchronous orbit so that the satellite crosses the equator at noon $\pm$ 20 minutes with an orbital period of 99 minutes.

The program is administered by the NASA Goddard Space Flight Center (GSFC) where raw data are processed and made available to end users. Data distribution has conditions, such as a two week delay in access to non-research users, due to the fact that ORBIMAGE also offers the data and related products commercially. Otherwise, if data are to be used for research purposes only, an organization can gain real-time access to data through Ocean Color Web site or CoastWatch upon approval of an associated application.

This SeaWiFS "data buy" program is under a five year contract that expires in 2007, therefore if NASA does not renew this agreement, data after the expiration date will not be available in real- or near real-time through GFSC or NESDIS CoastWatch.

6.1 SeaWiFS System Description

The OrbView-2 satellite collects color data continuously and transmits them twice a day to the GSFC High Resolution Picture Transmission station as well as to other receiving stations such as the Wallops Flight Facility, GeoEye, and other authorized receiving stations worldwide. Data are routed to the SeaWiFS Data Processing System (SDPS) operated at GSFC by the Ocean Biology Processing Group (OBPG) who is responsible for the production and distribution of the ocean color data products from SeaWiFS. The SDPS processes three types of raw data from the satellite into several end user product discussed in Section 6.3. Products are then made available via the web and FTP.

6.2 SeaWiFS DIF End Users

The end users of initial interest to the DIF project are the following models:

1. Hurricane Intensity Modeling: Hurricane Model and Hurricane Weather Research and Forecasting (HWRF) model

2. Integrated Ecosystem Assessments Modeling

3. Harmful Algal Bloom Forecasting and Assessments

6.3 SeaWiFS Data Products and Formats

CoastWatch provides the core IOOS variable of SeaWiFS ocean color data including the following products:

- SeaWiFS nLw (NOAA atmospheric corrections)
- SeaWiFS nLw (NASA atmospheric corrections)
- SeaWiFS Daily mean Chlorophyll_a concentration (NOAA atmospheric corrections, OC2.RS)

- SeaWiFS Daily mean Chlorophyll_a concentration (NASA atmospheric corrections, OC4v4)
- SeaWiFS 61-day mean Chlorophyll_a concentration (NOAA atmospheric corrections, OC2.RS)
- SeaWiFS 61-day mean Chlorophyll_a concentration (NASA atmospheric corrections, OC4v4)
- SeaWiFS Chlorophyll_a concentration anomaly (NOAA atmospheric corrections, OC2.RS)
- SeaWiFS Chlorophyll_a concentration anomaly (NASA atmospheric corrections, OC4v4)
- SeaWiFS Daily mean Rrs670 (NOAA atmospheric corrections)
- SeaWiFS Daily mean Rrs670 (NASA atmospheric corrections)
- SeaWiFS 61-day mean Rrs670 (NOAA atmospheric corrections)
- SeaWiFS 61-day mean Rrs670 (NASA atmospheric corrections)
- SeaWiFS Rrs670 anomaly (NOAA atmospheric corrections)
- SeaWiFS Rrs670 anomaly (NASA atmospheric corrections)

There are three raw SeaWiFS data types from the OrbView-2 satellite that are processed by the SeaWiFS Data Processing System (SDPS) at Goddard into products at different processing levels described below.

Processing of SeaWiFS data within OBPG begins with Level 0 (raw data) and culminates in Level 3 Binned and Mapped Browse data which are available for download from CoastWatch and the Ocean Color Web sites and archived in an FTP repository. Interim data sets include Level 1 and Level 2. All data sets are described below:

Level 1A data sets are generated by processing Level 0 data by appending calibration data, navigation data, instrument telemetry information, and selected spacecraft telemetry information to form a Level 1A Hierarchical Data Format (HDF) file that contains raw radiance values for each of the eight SeaWiFS optical sensor wavelength bands.

The input Level 0 data used to produce the Level 1A can be of two spatial resolutions - either LAC (Local Area Coverage), which is 1.1 km at nadir (directly below the spacecraft), or GAC (Global Area Coverage), which is 1.1 km resolution but with pixels spaced at 4.4km intervals.

Four types of Level 1 SeaWiFS data are processed by OBPG: GAC, LAC, HRPT, and MLAC.

- GAC (Global Area Coverage) data are sub-sampled and recorded onboard the spacecraft and subsequently downloaded twice a day at Wallops and NASA/Goddard. This data have effective resolution of about 4.5 kilometers along the center of the swath. For GAC data, individual products are generated from each Level-0 GAC recording period (the Earth data collection portion of an orbit). Each such GAC product thus constitutes one scene. GAC data are continuously collected and represent the primary data source from the onboard sensor.

- LAC (Local Area Coverage) data are recorded at full 1.1 kilometer resolution for selected parts of the world and downloaded with the GAC data. LAC data are broadcast to HRPT (high resolution picture transmission) sites around the world operating as ground-receiving stations.

- HRPT (High Resolution Picture Transmission) Direct Broadcast data have the same basic format and resolution as the LAC data, but they are collected by ground stations within range of the OrbView-2 spacecraft transmitter. For HRPT data, each scene is comprised of one satellite pass.

- MLAC data (merged LAC) contain all available SeaWiFS HRPT and LAC data for a given orbit. This is done by consolidating all SeaWiFS 1-km-resolution data which have been collected by various HRPT stations, as well as the LAC data recorded onboard the spacecraft into Level 1A files on a per orbit basis. Level-1A or Level-2 MLAC data are identical in format to SeaWiFS LAC and HRPT, except that the scan times may be discontinuous due to gaps in the HRPT/LAC geographic coverage. For MLAC data, each product contains the best available full-resolution data for a single orbit, without duplication.

Of the Level 1 and 2 SeaWiFS data, only GAC and MLAC data are distributed via the Ocean Color Web site.

Each **Level 2** scene corresponds to a level 1A scene. Level-2 processing is performed using the Multi-Sensor Level-1 to Level-2 (MSL12) code, which is developed and maintained by the OBPG. MSL12 is used for the standard processing of all ocean products distributed through the ocean color web browsers and ftp sites. This software is capable of retrieving oceanic optical properties and various derived products from the observed top-of-atmosphere (TOA) radiances collected by a variety of ocean remote sensing radiometers, including SeaWiFS, MODIS, OCTS, MOS, OSMI, POLDER, and CZCS. Full documentation, updated code source, sensor files, and output product descriptions are available.

Before computing Level 2 data, pixels are eliminated if they contain clouds, sun glint, or other abnormalities. For pixels that pass these screens, an atmospheric correction is applied to subtract the atmospheric scattering components from the total radiance to obtain the water-leaving radiances for bands 1-5.

Level 3 data processing involves both temporal and spatial binning and utilizes Level 2 GAC data as input in a process called l3bin. The Level 2 data are spatially binned before they can be temporally binned. Pixels containing valid Level 2 data are mapped to a fixed spatial grid composed of 9 x 9 km2 resolution elements (bins). The Level 2 GAC data are sub sampled by a factor of two to produce the Daily Level 3 product. This processing takes every valid measurement of water-leaving radiance falling within the lat/lon boundaries of a given grid square and compiles it within that bin. During the space binning step, a pixel from a parent Level-2 product can be excluded if certain flags are set in the Level 2 data.

Spatially binned Level 2 data are then temporally binned into single day, 8-day, monthly, and annual Level 3 products.

The Level 3 binned data set consists of geophysical parameters binned to a 9x9 km (81 km2) global, equal-area grid at daily, 8-day, monthly, and annual intervals. The parameters are the normalized water-leaving radiances (radiance data corrected for atmospheric light scattering and sun angles differing from nadir), and seven geophysical parameters derived from the radiance data. Each Level-3 binned data product will be stored in multiple HDF files. Each multi-file product includes a main file containing all product-level metadata and data for each bin that are common to all the binned geophysical parameters. In addition, each product includes 12 subordinate files, each of which contains data of one binned geophysical parameter for all bins. Subordinate files must be read in conjunction with the associated main file.

6.3.1 SeaWiFS Data Discovery and Metadata

CoastWatch has developed a metadata specification for its HDF products that is well described in Appendix B – *CoastWatch HDF Metadata Specification* of the *CoastWatch Software Library and Utilities: User's Guide*. The specification includes Global Metadata, Earth Location Metadata, and Variable Metadata.

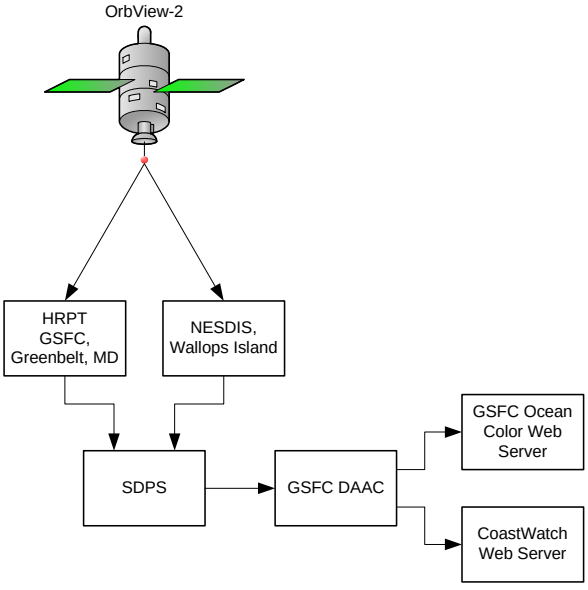
6.3.2 SeaWiFS Data Transform and Assembly

Data transform and assembly are essentially described above in Section 6.3 SeaWiFS Data Products and Formats in the discussion regarding processing of Level 0 raw data from the satellite into various Level 1, 2, and 3 data sets.

The data flow for SeaWiFS data starts with transmission from the OrbView-2 satellite to the High Resolution Picture Transmission (HRPT) station at the Goddard Space Flight Center (GSFC) in Greenbelt, MD, and to the NESDIS station at Wallops Island. Data are then sent to the Satellite Data Processing Center (SDPS), which comprises a number of elements including: SeaWiFS data processing, calibration, validation and data quality; data capture; and GSFC mission operations. The data processing element receives raw spacecraft data and generates standard global ocean color data products. These products are sent to the GSFC EOS Data and Information System (EOSDIS) Distributed Active Archive Center (DAAC), which is responsible for archiving and distributing SeaWiFS data.

Figure 5 below depicts the basic flow of data for SeaWiFS.

Figure 5 – SeaWiFS Data Flow Diagram



6.3.3 SeaWiFS Data Transport and Access

SeaWiFS data products are transported and accessed via the Internet from the Ocean Color Web site. Users can access delayed data using web browsers from HTTP web sites that include http://coastwatch.noaa.gov/ and http://oceancolor.gsfc.nasa.gov . Near real-time data are accessed from the CoastWatch web site that requires a login account, namely http://coastwatch.noaa.gov/interface_protected/interface_protected.shtml. This is the access portal used by the HAB forecasting and assessment office to access SeaWiFS data.

6.3.4 SeaWiFS Data Storage and Archive

To be supplied.

6.3.5 SeaWiFS Data Quality Control

In geographic regions with a high concentration of HRPT stations, the data collected from multiple stations will frequently overlap. Any duplicate scans resulting from that overlap are evaluated through a series of quality tests, and only the "best quality" scan is stored in the MLAC file. The quality tests look at pixel-to-pixel variance between duplicate scans, and bit-error counts derived from sync words. This process significantly reduces the transmission noise inherent in HRPT data stream, and elimination of duplicate observations makes it possible to bin the MLAC data with the same algorithms and software used for SeaWiFS GAC or MODIS.

6.4 SeaWiFS Data Product Information Sheets

The following data product information sheets comprise tables of relevant data about each product in use, or planned for use, by the initial DIF customer models.

Table 5 – SeaWiFS Data Product Information Sheets

System name: MODIS Aqua			
1	**System Information:**		
1.1	Responsible Program, Line Office (CO-OPS, NDBC, NODC, etc.)	Goddard Space Flight Center; NESDIS CoastWatch program	
1.2	Program/Line Office POC		
1.3	Technical/Data POC		
1.4	CASANOSA system name and IMS name		
1.5	NOAA system name and ID		
1.6	OMB 300? Or which one do you fall under		
1.7	Variables (from list of 20 core variables) collected into system		
1.8	System Function		
2	**Data Product Information:**		
2.1	Data Product title	1) SeaWiFS	2)
2.1.1	Related IOOS core variables	Ocean color; chlorophyll-a concentration	
2.1.2	Data product format (comma delimited ASCII, XML, NetCDF, GRIB, BUFR, etc)	HDF (primary), PNG, GeoTIFF	
2.1.3	Data latency for each data product version		
2.1.4	Data product frequency (e.g. once an hour, twice a day, etc.)		
2.2	Input/Source Data Attributes		
2.2.1	Sensor/platform name(s)	OrbView-2 satellite	
2.2.2	Sampling rate / time frequency (How often are the data collected?)	Continuously	
2.2.3	Time coverage (start and end date/time)	Continuous	
2.2.4	Spatial resolution (50 km, etc)	1.1 km directly below satellite	
2.2.5	Data type (Grids, Profiles, Time Series, Points Data, etc)		
2.2.6	Data format (comma delimited ASCII, XML, GRIB, BUFR, etc)	HDF	

2.2.7	Geographic location key words (GOM, etc)		
2.2.8	Boundaries of coverage	2,800 km swaths, global	
2.2.9	Units of measure (F vs C, etc)		
2.2.10	Original source (e.g., Univ of MS)	GeoEye (formerly ORBIMAGE)	
2.2.11	Transport method into the system (T1, Internet, private network, etc)	Satellite, FTP	
2.2.12	Reference to QC procedures for each datastream (e.g., latency: real time, 6 min values vs week later; calibration; etc)		
2.3	Metadata		
2.3.1	Metadata description available? (Yes / No)	Yes	
2.3.2	Metadata posted to clearing house? (Yes / No)		
2.3.2.1	Name of clearinghouse (if applicable)		
2.3.2.2	Name of metadata record (if applicable)	CoastWatch HDF Metadata (http://coastwatch.noaa.gov/cw_formats.html)	
2.3.3	Using a standard parameter vocabulary? (Yes / No)		
2.3.3.1	Which standard vocabulary is used?		
2.3.3.2	Reference to standard vocabulary		
2.4	Data Product Dissemination		
2.4.1	Transport method(s) from system to customer (T1, Internet, etc)	Internet; FTP	
2.4.1.1	Output data format(s) for each transport method (e.g. ASCII, XML, GRIB, BUFR, etc.)	HDF, PNG, GeoTIFF	
2.4.1.2	Data volume (e.g. MBytes per product)		
2.4.2	How dataset is made available (OpenDap, ftp, http, WMS, etc)	HTTP; FTP	
2.4.3	Data product access URL (access point)	ftp://ftpcoastwatch.noaa.gov/pub/data/products/	
2.4.4	Reference documentation URL		
2.4.5	Other known processing locations?	Yes	
2.4.5.1	Secondary provider URL (if applicable)	http://oceancolor.gsfc.nasa.gov	
2.5	IT Security		
2.5.1	Are data available to public?	Yes	

2.5.2	Is authorization required to view/release data? (e.g. user or subscription account is required)	Yes, for near real-time data	
2.5.3	What are controls used to prevent malicious attacks?		
2.6	Data Integrity		
2.6.1	What are used to ensure transmitted data equal received data?		
2.7	Data Archive		
2.7.1	Is data product being archived?	Yes	
2.7.2	Where is data product archived?		
2.7.3	What is the data archive access URL?		
2.7.4	What is the reference documentation URL?		

7 Additional Input Data Products and Sources

7.1 Additional Input and Data Products and Sources

As additional data sources and products are identified and researched for integration with the DIF, it is expected that this document will be updated with information about each. The expected list of these systems and products is as follows:

1. HF Radar
2. NWSTG / WMO GTS
3. CalCOFI
4. NERRS CDMO
5. US GODAE GODAC
6. ARGO
7. AOML xbts
8. SWFSC
9. NCDC NOMADS
10. NCEP NOMADS
11. NGDC IMS
12. GFDL Models
13. CREIOS
14. Tide Gauges
15. POES
16. USGS River Stage Gauges

8 Definitions and Acronyms

ADCIRC	Advanced Circulation Hydrodynamic Model
ADCP	Acoustic Doppler Current Profiler
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
CODAR	Coastal Ocean Dynamics Applications Radar
COMPS	Coastal Ocean Monitoring and Prediction System
CO-OPS	Center for Operational Oceanographic Products and Services
CSC	Coastal Services Center
CSDL	Coast Survey Development Lab
CTD	Conductivity, Temperature, Depth
DIF	Data Integration Framework
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
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
IOC	Initial Operating Capability
IOC	Intergovernmental Oceanographic Commission
IODE	International Oceanographic Data and Information Exchange committee
IOOS	Integrated Ocean Observing System
IT	Information Technology
JCOMM	Joint Commission on Oceanography and Marine Meteorology
MEDS	Canada's Marine Environmental Data Service
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
NIST	National Institute of Standards and Technology
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
POM	Princeton Ocean Model
PORTS	Physical Oceanographic Real-Time System
QA	Quality Assurance
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
USF	University of South Florida
USGS	United States Geological Survey
WMO	World Meteorological Organization