**Background**

Since its inception in November 1975, the Coastal Data Information Program (CDIP) collects near real-time physical environmental data mostly in the coastal US and South Pacific. CDIP has many partners including industry, federal and state agencies and academia. In all cases, these data are transmitted from the station location to CDIP at the Scripps Institution of Oceanography (SIO), La Jolla, CA where the data are processed and disseminated.

The program captures wave, wind, and temperature data.

CDIP Shares data in a variety of file formats.

1. FM 65 XML - Used for the real-time data push to the NDBC. FM 65 format is described here http://www.ndbc.noaa.gov/decode.shtml.
2. NetCDF - A self-describing, machine-independent data format that support the creation, access, and sharing of array-oriented scientific data, available from the CDIP site http://thredds.cdip.ucsd.edu.
3. ASCII - Text file that are easily read and parsed by people and programs via the web, available from the CDIP site, e.g., http://cdip.ucsd.edu/?nav=recent

**Data Ingestion**



The data are collected by several redundant pathways.

1. The majority of our buoy data are transmitted via iridium. The path is shown in the following link which depicts the offshore buoy transmitting the data to iridium satellite, then to the Department of Defense iridium gateway in Honolulu and back to SIO or Amazon Cloud as appropriate. (<http://cdip.ucsd.edu/?nav=documents&xitem=dacq#system>)
2. For a select number of pier or near-shore stations the data are transmitted via network to CDIP. (<http://cdip.ucsd.edu/themes/cdip?d2=p20&u3=tab:1:display:system_organization>)
3. An internal compact flash card stores the data, available upon recovery.

**Data Management**

The data are managed at the SIO/CDIP server. Once ingested, CDIP processes and quality controls these data. The data are stored on disk in ASCII, NetCDF, and SQL formats. Back-up occurs hourly locally, daily offsite at the UCSD Supercomputer Center and monthly to Amazon Glacier.

**Data Distribution**

CDIP Access to Data (<http://cdip.ucsd.edu/?nav=documents&xitem=product#access>)

1. THREDDS data are organized into Archived and Realtime folders:
	1. Archived - contains individual folders for all CDIP stations, both active and decommissioned. Each station’s individual Archived folder contains NetCDF files for each separate deployment (e.g. ‘d17.nc’) and an aggregate file (‘historic.nc’) of the full time-span of data for a buoy.
	2. Realtime - contains single NetCDF files (‘rt.nc’) for CDIP stations that are currently active and transmitting data.
		1. OPENDAP - provides URL that can be used in Python/Matlab to automatically grab NetCDF file of data from server. Also provides option to download user-specified variables/timeperiods as ASCII or Binary file.
		2. HTTPServer - option to download the whole NetCDF file.
		3. NCML (NetCDF Markup Language) - XML document used to define a CDM dataset, and to allow user to add/delete/change metadata and variables, or combine data from multiple CDM files.
		4. ISO - XML metadata record for each station.
		5. UDDC (Unidata Data Discovery Convention) - tool to determine how well file metadata conforms to list of recommended metadata attributes.
		6. SOS - web service interface which allows querying observations, sensor metadata, and representations of observed features. Defines means to register/remove sensors and insert new sensor observations.

NetCDF files for Archived and Realtime data contain identical buoy parameters and variables, with the exception that the ‘historic.nc’ Archived file and the ‘rt.nc’ Realtime file do not contain Directional Displacement (xyz) data.

1. CDIP Data Access Routine (DAR) <http://cdip.ucsd.edu/data_access/justdar.cdip>
Returns CDIP data for automatic web downloads.
2. CDIP Website <http://cdip.ucsd.edu>
3. CDIP FTP ftp://ftp.cdip.ucsd.edu
4. National Data Buoy Center (NDBC) for distribution on their website and dissemination via the Global Telecommunications Service (GTS).
5. Several federal, state and private companies access CDIP data for distribution using one of the access methods above.
6. Several federal, state and private companies access CDIP data for distribution using one of the access methods above

**Quality Control**

A sophisticated suite of automated and human quality control procedures are developed, as defined in the QARTOD manual (http://www.ioos.noaa.gov/qartod/waves/welcome.html). In addition, CDIP has also developed further instrument and site specific tests. The tests are summarized in the following table: <http://cdip.ucsd.edu/documents/index/product_docs/qc_summaries/waves/waves_table.php?&xtab=CDIP>

All errors causing an exception are handled by the following:

* logged in a daily errors file
* error exception emailed to the CDIP software team
* categorized by error type and station at the end of each month to provide an error summary table.
* flagged and annotated in the NetCDF file as appropriate

When there are critical errors involving a buoy offsite or a station that has not updated within 3 hours, the software team is not only notified via email but, a designated watch person is also paged.

Only those data that pass all the QC tests are transmitted to the National Data Buoy Center (NDBC) & the National Weather Service (NWS).

The above quality control procedure can be monitored at: <http://cdip.ucsd.edu/diag>

Once data have been acquired, processed, and quality controlled, CDIP makes the complete data set available. (Near-real time, approximately 3 minutes after the data are transmitted)

*Metadata/ documentation are submitted alongside the data or created on deposit/ transformation in order to make the data reusable.*

All of CDIP's data sets are described by detailed metadata, which is continuously updated and available online in a number of formats. FGDC-compliant metadata are included, in both HTML and XML formats. The metadata for any specific data set are accessible from the station pages in the historic section of the website. In addition to the standard web pages, static XML metadata files are available for download or harvesting from a web-accessible folder (<http://cdip.ucsd.edu/data_access/metadata>). The NetCDF files also include metadata and are available in ISO 19115-2 metadata XML from the CDIP THREDDS catalog (<http://thredds.cdip.ucsd.edu>)

FGDC metadata consists of seven main sections, five of which do not need to be included if they do not apply to the data set in question. For CDIP metadata, two sections are omitted Spatial\_Reference\_Information and Spatial\_Data\_Organization\_Information - because they only apply to datasets that include spatial data. (Although CDIP's metadata contains spatial info - deployment positions - the data sets themselves do not.)

Thus CDIP metadata consists of five sections:

1. Identification\_Information
2. Data\_Quality\_Information
3. Entity\_and\_Attribute\_Information
4. Distribution\_Information
5. Metadata\_Reference\_Information

Many of the fields in the content standard are defined as free text, and can contain links to other resources. CDIP's metadata takes full advantage of this fact, linking to relevant documents and pages on the CDIP website wherever possible. This is the most efficient and effective approach because CDIP's online documentation is extensive and covers most of the topics addressed in the FGDC standard. By linking directly to CDIP's web resources redundancy is avoided and the metadata are ensured to be up-to-date. This same approach is used in defining CDIP's entity and attribute information.

**Archiving**

National Centers for Environmental Information (NCEI) is the federal archive repository. Historic data from CDIP stations are archived monthly and available at NCEI (<http://www.nodc.noaa.gov/access/index.html>). The archive process was established with the NCEI Submission Information Form (<https://goo.gl/AmX8F8>).

Local redundant HDD storage at the CDIP Lab, the UCSD Supercomputer center, Amazon Glacier and NCEI.

NCEI Submission Information Form: <http://sccoos.org/documents/home/archive/>

**Permission Restrictions**

CDIP data and products are freely available for public use. When referenced, please provide a link to the CDIP homepage.

Examples:

1) Standard html:

Data courtesy of <a href=http://cdip.ucsd.edu/>CDIP</a>

2) Offline references, choose the appropriate form from the recommended acknowledgements below.

* Short form (figure captions, etc.)
"... data from CDIP, Scripps Institution of Oceanography."
* Longer form (in text)
"...data were furnished by the Coastal Data Information Program, Integrative Oceanography Division, operated by the Scripps Institution of Oceanography."
* Full form (acknowledgements at conclusion of papers, etc.)
"...data were furnished by the Coastal Data Information Program (CDIP), Integrative Oceanography Division, operated by the Scripps Institution of Oceanography, under the sponsorship of the U.S. Army Corps of Engineers and the California Department of Parks and Recreation."

In general, a data sharing agreement will not be required. However, data should be properly acknowledged.

The one exception is with NOAA Physical Ocean Real Time System (PORTS). A Memorandum of Understanding (MOU) between NOAA PORTS and the US Army Corps, representing CDIP as the funding agency, is signed.

**Intellectual Property**

The funding agency & the University of California, San Diego through a contractual agreement.

**Publications:**

<http://cdip.ucsd.edu/go/publications>