

Data Management Plan

Coastal Data Information Program (CDIP)

Jennifer McWhorter[‡], Darren Wright[‡], Julie Thomas[‡]

[‡] University of California, San Diego, United States of America

Corresponding author: Jennifer McWhorter (jmcwhorter@ucsd.edu),
Julie Thomas (jthomas@ucsd.edu)

Reviewable v1

Received: 12 Apr 2016 | Published: 15 Apr 2016

Citation: McWhorter J, Wright D, Thomas J (2016) Coastal Data Information Program (CDIP). Research Ideas and Outcomes 2: e8827. doi: [10.3897/rio.2.e8827](https://doi.org/10.3897/rio.2.e8827)

Abstract

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, with a focus on waves. CDIP has many partners including industry, federal and state agencies and academia.

New information

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 to interested parties.

Keywords

waves, data, quality control, real-time, long-term time series, sea-surface temperature

Types of data and information created

What data will you collect or create in the research?

Contextual statement describing what data are collected and relevant URL (IOOS Certification, f 1. ii)

What data types will you be creating or capturing?

The program captures wave, wind, and temperature data in real-time, updating every 30 minutes.

How will you capture or create the data?

Describe how the data are ingested (IOOS Certification, f 2.)

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.

Describe how data are managed (IOOS Certification, f 2.)

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 biannually to Amazon Cloud.

Describe the data quality control procedures that have been applied to the data. (IOOS Certification, f 3.)

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>

If you will be using existing data, state that fact and include where you got it. What is the relationship between the data you are collecting and the existing data?

N/A

Expected schedule for data sharing

Adheres to the NOAA Data Sharing Procedural Directive. The System is an operational system; therefore the RICE should strive to provide as much data as possible, in real-time or near real-time, to support the operation of the System. (IOOS Certification, f. 4.)

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)

How long will the original data collector/creator/principal investigator retain the right to use the data before opening it up to wider use?

N/A

How long do you expect to keep the data private before making it available? Explain if different data products will become available on different schedules (Ex: raw data vs processed data, observations vs models, etc.)

N/A

Explain details of any embargo periods for political/commercial/patent reasons? When will you make the data available?

N/A

Standards for format and content

Which file formats will you use for your data, and why?

How can the information be accessed? (IOOS Certification, f 1. ii)

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>

What file formats will be used for data sharing?

All of the above.

What metadata/ documentation will be 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 XML from the CDIP THREDDS catalog (<http://thredds.cdip.ucsd.edu>)

What contextual details (metadata) are needed to make the data you capture or collect meaningful?

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.

How will you create or capture these details?

CDIP's FGDC metadata is generated by querying our 'archive' MySQL database and passed through the US Geological Service's utility 'mp': <http://geology.usgs.gov/tools/metadata/tools/doc/mp.html> The mp program verifies that the metadata is FGDC-compliant, and then outputs it in the desired format, either html or xml.

CDIP's NetCDF files have ISO 19115 compliment metadata which are generated with custom FORTRAN scripts.

What form will the metadata describing/documenting your data take?

CDIP's data sets are described by detailed metadata in a number of formats:

- FGDC XML - cdip.ucsd.edu/data_access/metadata
- ISO 19115 XML - NetCDF <http://thredds.cdip.ucsd.edu>
- HTML - CDIP website has extensive documentation <http://cdip.ucsd.edu>

Which metadata standards will you use and why have you chosen them? (e.g. accepted domain-local standards, widespread usage)

FGDC and ISO 19115 metadata are both accepted standards and mandated by the US Federal Government.

Policies for stewardship and preservation

What is the long-term strategy for maintaining, curating and archiving the data?

Points of contact- Individuals responsible for the data management and coordination across the region (CV's attached); (IOOS Certification f 1. i)

Julie Thomas - Employee 39 years, Principal Investigator/Program Manager

Darren Wright - Employee 10 years, Programmer/Analyst

Jen McWhorter - Employee 1 year, Administrative Analyst

Identify the procedures used to evaluate the capability of the individual (s) identified in subsection 997.23(f)(1) to conduct the assigned duties responsibly. (IOOS Certification, f 1. iii)

The University of California has a process in place for personnel evaluation. These evaluations are on file with UC San Diego Human Resources. All personnel listed have received excellent evaluations.

Which archive/repository/database have you identified as a place to deposit data?

Documents of the RICE's data archiving process or describes how the RICE intends to archive data at the national archive center (e.g., NODC, NGDC, NCDC) in a manner that follows guidelines outlined by that center. Documentation shall be in the form of a Submission Agreement, Submission Information Form (SIF) or other, similar data producer-archive agreement (IOOS Certification, f 6.).

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>).

What procedures does your intended long-term data storage facility have in place for preservation and backup?

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

How long will/should data be kept beyond the life of the project?

Data are indefinitely stored.

What data will be preserved for the long-term?

All data are publicly available and preserved.

What transformations will be necessary to prepare data for preservation / data sharing?

Raw data are decoded and formatted, analyzed and quality controlled.

What metadata/ documentation will be submitted alongside the data or created on deposit/ transformation in order to make the data reusable?

FGDC standard metadata are available per deposit and transformation. NetCDF files have complete metadata and quality control flags.

What related information will be deposited?

Time series and spectral files.

Procedures for providing access**What are your plans for providing access to your data? (on your website, available via ftp download, via e-mail, or another way)**

Describe how data are distributed including a description of the flow of data through the RICE data assembly center from the source to the public dissemination/access mechanism. (IOOS Certification, f. 2.)

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

1. THREDDS data are organized into Archived and Realtime folders:
 - a. 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.
 - b. Realtime - contains single NetCDF files ('rt.nc') for CDIP stations that are currently active and transmitting data.
 - i. 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.
 - ii. HTTPServer - option to download the whole NetCDF file.
 - iii. 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.
 - iv. ISO - XML metadata record for each station.
 - v. UDDC (Unidata Data Discovery Convention) - tool to determine how well file metadata conforms to list of recommended metadata attributes.
 - vi. 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.

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

Will any permission restrictions need to be placed on the data?

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 [CDIP](http://cdip.ucsd.edu/)
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."

With whom will you share the data, and under what conditions?

Data are publicly available.

Will a data sharing agreement be required?

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.

Are there ethical and privacy issues? If so, how will these be resolved?

N/A

Who will hold the intellectual property rights to the data and how might this affect data access?

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

Previous published data

Articles:

Storm wave induced mortality of giant kelp, *Macrocystis pyrifera*, in Southern California Seymour, R.J., M.J. Tegner, P.K. Dayton, and P.E. Parnell, Estuarine, Coastal and Shelf Science, Vol. 28, pp. 277-292, 1989

Unusual marine erosion in San Diego County from a single storm Dayton, P.K., R.J. Seymour, P.E. Parnell, and M.J. Tegner, Estuarine, Coastal and Shelf Science, Vol. 29, pp. 277-292, 1989

COASTAL FORUM: Unusual damage from a California storm Seymour, R.J., Shore & Beach, Vol. 57, No. 3, July 1989, p. 31, 1989

[Editorial] The great storm of January 1988 Seymour, R.J., Shore & Beach, Vol. 57, No. 4, October 1989, p. 2., 1989

A Comparison of Spectral Refraction and Refraction-Diffraction Wave Propagation Models William C. O'Reilly and R. T. Guza, J. Waterway, Port, Coastal and Ocean Eng., 117, (3), pp199-215, 1991.

A Comparison of Two Spectral Wave Models in the Southern California Bight William C. O'Reilly and R. T. Guza, Coastal Eng., 19, pp263-282, 1993.

Wave Monitoring in The Southern California Bight William C. O'Reilly, R. J. Seymour, R. T. Guza, D. Castel, Ocean Wave Measurement and Analysis, Proc. 2nd Int. Symp. July 25-28, 1993, pp448-457.

New Technology in Coastal Wave Monitoring Richard Seymour, David Castel, David McGehee, Julianna Thomas, and William O'Reilly, Ocean Wave Measurement and Analysis, Proc. 2nd Int. Symp. July 25-28, 1993, pp105-123.

Field Wave Gaging Program, Wave Data Analysis Standard Marshal D. Earle, David McGehee, and Michael Tubman, USACE Instruction Report CERC-95-1, March 1995.

Effects of Southern California Kelp Beds on Waves M. Hany S. Elwany, William C. O'Reilly, Members, ASCE, Robert T. Guza, and Reinhard E. Flick, J. Waterway, Port, Coastal and Ocean Eng., 121,(2), pp143-150, 1995.

A Comparison of Directional Buoy and Fixed Platform Measurements of Pacific Swell W. C. O'Reilly, T. H. C. Herbers, R. J. Seymour and R. T. Guza, J. Atmos. and Ocean. Technol., 13, (1), pp231-238, 1996.

Observations of Seiche Forcing and Amplification in Three Small Harbors Okihiro, M. and Guza, R., J. Waterway, Port, Coastal, Ocean Eng., 122(5), 232-223, 1996.

Effects of El Nino on the West Coast Wave Climate Seymour, R.J., Shore & Beach, Vol. 66(3): 3-6, 1998

Assimilating Coastal Wave Observations in Regional Swell Predictions. Part 1: Inverse Methods W. C. O'Reilly and R. T. Guza, J. Physical Oceanography, 28, (4), pp679-691, 1998.

The Relationship Between Incident Wave Energy and Seacliff Erosion Rates: San Diego County, California Bunumof, B.T., Storlazzi, C.D., Seymour, R.J., Griggs, G.B., California Journal Coastal Research, Vol. 16, No. 4, 1162-1178, 2000

Evidence for Changes to the Northeast Pacific Wave Climate Seymour, R.J., Journal of Coastal Research, Vol. 27, Issue 1: pp. 194-201, 2001

Rapid Erosion of a Southern California Beach Fill Seymour, R.J., R.T. Guza, W. O'Reilly and Steve Elgar, J. Coastal Engineering, 52, (2), pp151-158, 2004.

Application of Airborne LIDAR for Seacliff Volumetric Change and Beach-Sediment Budget Contributions Adam P. Young and Scott A. Ashford, J. Coastal Research, 22, (2), pp307-318, 2006.

Performance Evaluation of Seacliff Erosion Control Methods Adam P. Young and Scott A. Ashford, Shore and Beach, 74, (4), pp16-24, 2006.

Evolution of Surface Gravity Waves Over a Submarine Canyon R. Magne, K.A. Bellibassakis, T.H.C. Herbers, Fabrice Ardhuin, W.C. O'Reilly, and V. Rey, *J. Geophysical Research*, 112, C01002, pp1-12, 2007.

A Technique for Eliminating Water Returns from Lidar Beach Elevation Surveys Yates, M.L., R.T. Guza, R. Gutierrez, and R.J. Seymour, *J. Atmos. and Ocean. Technol.*, 25, pp1671-1682, 2008.

Seasonal Persistence of a Small Southern California Beach Fill M.L. Yates, R.T. Guza, W.C. O'Reilly, R.J. Seymour, *J. Coastal Engineering*, 56, pp559-564, 2009.

Overview of seasonal sand level changes on southern California beaches Yates, M.L., R.T. Guza, W.C. O'Reilly, and R.J. Seymour, *Shore & Beach*, 77(1), pp39-46, 2009.

Comparison of short-term seacliff retreat measurement methods in Del Mar, California Adam P. Young, R.E. Flick, R. Gutierrez, and R.T. Guza, *Geomorphology*, 2009

Coarse Sediment Yields from Seacliff Erosion in the Oceanside Littoral Cell Adam P. Young, J.H. Raymond, J. Sorenson, E.A. Johnstone, N.W. Driscoll, R.E. Flick, and R.T. Guza, *Journal of Coastal Research*, Vol 26, No. 3, pp. 580-585, May 2010.

Rain, Waves, & Short-Term Evolution of Composite Seacliffs in Southern California Adam P. Young, R.T. Guza, R.E. Flick, W.C. O'Reilly, and R. Gutierrez, *Marine Geology*, 2009

Comparison of Airborne and Terrestrial LIDAR Estimates of Seacliff Erosion in Southern California Adam P. Young, M.J. Olsen, N. Driscoll, R.E. Flick, R. Gutierrez, R.T. Guza, E. Johnstone, and F. Kuester, *Photogrammetric Engineering & Remote Sensing*, April 2010.

A Portable Airborne Scanning Lidar System for Ocean and Coastal Applications Benjamin D. Reineman, Luc Lenain, David Castel, and W. Kendall Melville, 2009

Equilibrium shoreline response: Observations and modeling Yates, M.L., R.T. Guza, and W.C. O'Reilly, *J. Geophys Res*, 114, C09014, doi:10.1029/2009JC005359, 2009

Short-term coastal cliff retreat statistics at Sunset Cliffs - Point Loma, California, USA Young, A.P., R.T. Guza, W.C. O'Reilly, R.E. Flick, and R. Gutierrez, *Natural Hazards and Earth System Sciences*, 11, 205-217, Jan 2011

The Effect of Temporal Wave Averaging on the Performance of an Empirical Shoreline Evolution Model M.A. Davidson, M.A., I.L. Turner and R.T. Guza, *Coastal Engineering*, 58, 802-805, 2011

Equilibrium Shoreline Response of a High Wave Energy Beach Yates, M.L., R.T. Guza, W.C. O'Reilly, P. Barnad and J. Hansen, *J. Geophys. Res.*, 116, C04014, doi:10.1029/2010JC006681, 2011

Coastal cliff ground motions from local ocean swell and ifragravity waves in southern California Adam P. Young, P.N. Adams, W.C. O'Reilly, R.E. Flick, R.T. Guza, J. Geophys Res., 116, C09007, doi:10.1029/2011JC007175, Sep 2011

Project

National Oceanic and Atmospheric Administration (NOAA) Template

Public Data Management Plan created with the DMPTool: <https://dmptool.org/plans/15485.pdf>