


# Data Services at the IOOS Regional Association for the Pacific, PacIOOS

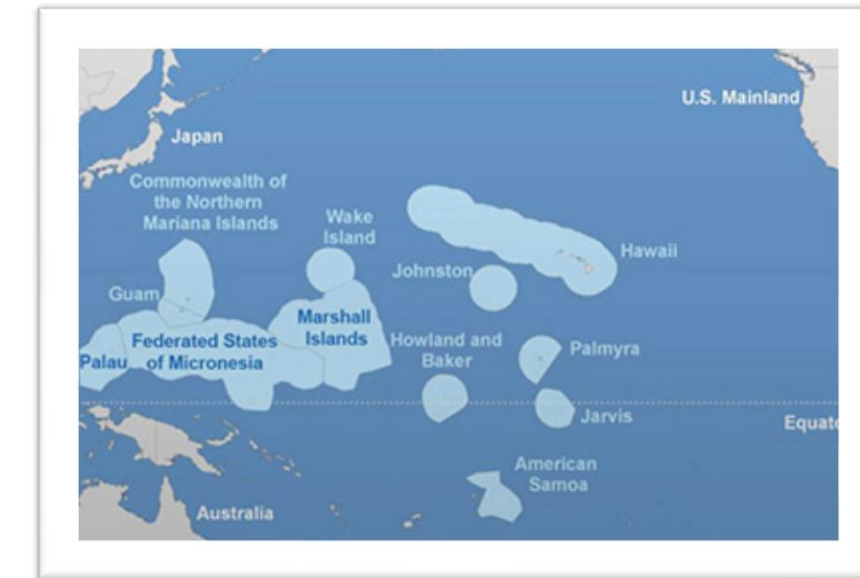
John Maurer (jmaurer@hawaii.edu), James T. Potemra (jimp@hawaii.edu), Richard (Abe) Coughlin (rac88@hawaii.edu). University of Hawai'i at Mānoa. AGU 2012 Poster #1498909.



## PacIOOS:


### Pacific Islands Ocean Observing System, <http://pacioos.org>

*by observing we learn*




PacIOOS provides timely, reliable, and accurate ocean information to support a safe, clean, productive ocean and resilient coastal zone in the U.S. Pacific Islands region.

As our eyes on the oceans, coasts, and Great Lakes, the United States Integrated Ocean Observing System (IOOS) is a tool for tracking, predicting, managing, and adapting to changes in our marine environment. IOOS delivers the data and information needed to increase our understanding of our oceans, coasts, and Great Lakes, so decision makers can take action to improve **safety**, enhance our **economy**, and protect our **environment**. PacIOOS is one of eleven regional associations that were created as a part of U.S. IOOS and spans the whole of U.S. interests in the insular Pacific region.

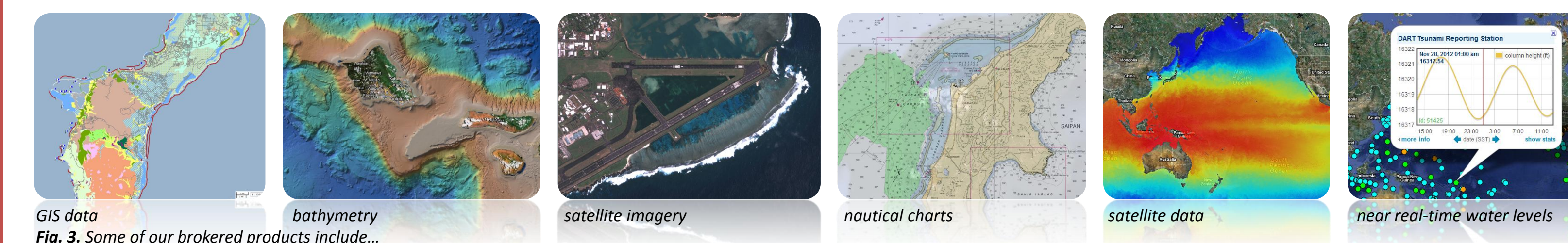


## data:

PacIOOS and its partners deploy and maintain a variety of *in situ* and *remote* oceanographic instruments, produce a suite of **forecasts** from regional numerical models, and collect information from tagged marine **species**:



In addition to these core assets, we also identify and broker important **third-party** data sets to facilitate their **dissemination**, enhance their **utility**, and provide **integration** with our existing product line:



While we support a large variety of **data formats** and provide numerous output formats for users to download, our primary data stores are contained in one of the following formats:

- time series** → **NetCDF: Network Common Data Form**  
An OGC standard, NetCDF is a self-describing data format that supports encoding of multidimensional geospatial information representing space/time-varying phenomena.
- GIS** → **PostGIS** / **OSGeo**  
An open source object-relational database management system that adds geographic support to PostgreSQL. Used in place of Shapefiles for querying and spatial operators.
- imagery** → **GeoTIFF** / **GeoTIFF**  
A public domain metadata standard that allows georeferencing information to be embedded within a TIFF file. Popular in the GIS community for storing raster imagery.

## servers:

PacIOOS manages a variety of data servers to provide easy and **interoperable** access to our data holdings. While overlapping in some regards, each of these **open source** applications handles a different swath of data types and capabilities:

- TDS: THREDDS Data Server: Thematic Realtime Environmental Distributed Data Services**  
<http://oos.soest.hawaii.edu/thredds/>  
Serves out an **aggregate** data set from a collection of NetCDF files in our archive, thereby allowing the end user to **subset** an arbitrary spatiotemporal range of their choosing. **NCML** allows dynamic data set modifications and metadata editing.
- ERDDAP: Environmental Research Division's Data Access Program**  
<http://oos.soest.hawaii.edu/erddap/>  
Supports a large variety of output formats for both gridded and point-based data sets, including **RESTful** access. Useful as both a data server and a data viewer. Like TDS, also provides data set aggregation and subsetting.
- LAS: Live Access Server**  
<http://oos.soest.hawaii.edu/las/>  
Serves data via a Ferret plug-in to TDS called **F-TDS**. Ferret is a data analysis package developed at NOAA PMEL. Provides many options for producing quality maps and plots of gridded NetCDF data.
- Dapper**  
<http://oos.soest.hawaii.edu/dapper/>, <http://oos.soest.hawaii.edu/dchart/>  
Dapper is an **OpenDAP** server for providing access to NetCDF time series data at point locations. Interactive time series plots of point-based data sets and maps of gridded data can be produced via the associated **DChart** application.
- OOSTethys: Ocean Observing System Tethys (Greek sea goddess)**  
<http://oos.soest.hawaii.edu/oostethys/>  
Provides Sensor Observation Service (**SOS**) access to point-based time series data. OOSTethys is an orphaned project, so we are looking to provide SOS support via TDS when **ncSOS** goes operational in 2013.
- GeoServer**  
<http://pacioos.org/geoserver/>  
Serves **GIS layers** for the insular Pacific. Supports several output formats and OGC data services. While our other data servers focus primarily on providing real-time observations and forecast models, GeoServer provides static maps.
- MapServer**  
MapServer also serves static **GIS layers** as OGC data services. While GeoServer is a Java-based GUI-driven application, MapServer is a **CGI-based** text-driven application that can be automated or dynamically modified via Python **MapScript**.

## PacIOOS data servers metadata services viewers you



## metadata: <http://pacioos.org/metadata/>

PacIOOS supports several community-based **standards** for representing geospatial metadata. In many cases, these **XML** documents get produced by our servers on-the-fly, while our brokered products may provide static **ISO** or **FGDC** metadata files in varying degrees of completion from their host institutions.

- THREDDS Dataset Inventory Catalog Specification**  
Presents metadata overview on the TDS catalog page for a dataset.
- CF: NetCDF Climate and Forecast Metadata Convention**  
Required global attributes, standard variable names, file organization, etc.
- NetCDF Attribute Convention for Dataset Discovery**  
Various global and variable metadata attributes to include in NetCDF files.
- TDS: THREDDS Data Server**  
Aggregates and subsets NetCDF, OpenDAP, HDF5, and other CDM files.
- NCML: NetCDF Markup Language**  
Allows TDS to dynamically modify data and edit metadata attributes.
- NCISO: NetCDF ISO**  
A plug-in for TDS that pulls together all of these elements into ISO format.

**ISO 19115-2: Geographic Information – Metadata**

**HTML** / **XSLT: Extensible Stylesheet Language Transformations:** Using XSLT, we can convert XML to/from ISO and FGDC as well as provide HTML and plain text outputs.

**Plain Text** / **FGDC: Federal Geographic Data Committee** / **CSGDM: Content Standard for Digital Geospatial Metadata**

**XML**  
XML also saved in Web Accessible Folders (**WAFs**) to enable other organizations to harvest our metadata holdings in externally-hosted data catalogs and mapping portals.

Improves readability. Provides hypertext links and interactive **Google Map** that shows the data set's spatial bounds.

Improves readability. Maintains 80 character page width. Good format for printing and saving.

## services:

Several services are made possible through configuration of our data servers and appropriate metadata, giving users **flexible, on-demand, interoperable** data access in a variety of geospatial **transfer protocols**. Some are understood by popular software applications for scientific computing and GIS analysis, thereby directly importing **data streams** and preventing the need for users to download, subset, and reformat files themselves.

- NCSS: NetCDF Subset Service**  
Subsets gridded data in **TDS** via lat/lon, time, and/or depth using a **RESTful** (URL-based) API or online Web form. Using **CF-1.6** discrete geometry conventions, TDS will soon extend NCSS to non-gridded data (points, profiles, trajectories) in a future release (TDS 4.3).  
**output formats: ASCII, XML, NetCDF**
- OpenDAP: Open-source Project for a Network Data Access Protocol**  
Subsets data via array **index ranges** (e.g. [start:stride:stop], [0:5:50]). While an online Web form can be used to tailor output, its strength lies in directly importing data to OpenDAP-enabled software clients such as MATLAB, GrADS, IDL, Ferret, ArcGIS (via Environmental Data Connector), etc.  
**output formats: DODS binary, ASCII**
- WMS: Web Map Service**  
Produces map images of gridded data using a **RESTful** (URL-based) API for specifying map projection, spatial bounds, and other options. The WMS behind TDS (**ncWMS**) allows further customization of color (palette, opacity, bands, min, max, logscale) and vectors (arrow styles).  
**output formats: PNG, JPG, GIF, KML**
- WCS: Web Coverage Service**  
While WMS renders map images, WCS produces the actual **gridded data** itself, again with flexibility in specifying a particular map projection, spatial bounds, variable, and time period of interest. Not multidimensional (as NCSS and OpenDAP are) so a single slice in time/space is required.  
**output formats: GeoTIFF, NetCDF**
- WFS: Web Feature Service**  
Like WCS, WFS is used to access data as opposed to rendered images. While WCS is used for gridded data, WFS provides access to **vector data** (points, lines, and polygons). Primarily for static GIS layers, it does not handle time series data, for which SOS was developed (see below).  
**output formats: Geography Markup Language (GML)**
- SOS: Sensor Observation Service**  
Flexible, RESTful (URL-based) access to real-time **sensor data** and other **time series of discrete geometries** (stations, profiles, trajectories). Comprises descriptions of sensors themselves, which are encoded in Sensor Model Language (**SensorML**).  
**output formats: Observations and Measurements (O&M) XML**
- REST: Representational State Transfer**  
All of the above standards are **RESTful**: results can be tailored through the specification of numerous input parameters in the access **URL** and are obtained over **HTTP**. **ERDDAP** provides its own custom REST API for both gridded and point data. PacIOOS has also developed a custom REST CGI (**get\_data.py**) to provide an enhanced mechanism for accessing our own assets as well as some externally-hosted, third-party sensor data.  
**output formats: ASCII, XML, JSON, NetCDF, others**

## viewers:

Building on the data services available, PacIOOS provides several web applications for visualizing data, including **maps, plots** (time series and depth profiles), and time series **animations**.

**data integration:** These viewers can overlay multiple data sets simultaneously in a single online mapping application.

- Voyager**  
Google Maps API that PacIOOS developed for viewing and accessing time series (near real-time sensors, model forecasts, satellite data), static maps, and more.  
<http://pacioos.org/voyager/>
- GeoExplorer**  
GeoExplorer that PacIOOS installed for viewing and accessing static GIS layers in the insular Pacific from **GeoServer**.  
<http://pacioos.org/geoexplorer/>

**one thing at a time, please:** While these viewers cannot overlay multiple data sets, they provide specialized utilities.

- LAS: Live Access Server**  
<http://oos.soest.hawaii.edu/las/>
- Dapper DChart**  
<http://oos.soest.hawaii.edu/dchart/>  
Interactive **time series plots** or depth profiles of point-based data sets, though it can also be used to produce maps of gridded data.
- ERDDAP**  
<http://oos.soest.hawaii.edu/erddap/>  
While ERDDAP's strength is in serving data in multiple formats, it can also produce simple maps and plots of either gridded or point data sets.
- ncWMS Godiva2**  
View interactive, customizable maps of gridded data via **ncWMS** using links on THREDDS Data Server cataloging pages. Query data at a clicked point.