Executive Summary

The Integrated Ocean Observing System (IOOS) is designed as a national network of Regional Associations (RA) that provide information about the ocean environment to end-users. The regional association for the Pacific Islands, PacIOOS, is under development at the University of Hawaii (UH) School of Ocean and Earth Science and Technology (SOEST). PacIOOS has different components including modeling, water quality, estimating ocean state, *etc*. This document provides an overview of one of these components, the Data Management System (DMS). The present PacIOOS grant follows several other grants that initiated the insular Pacific IOOS and the Hawaii regional IOOS as separate entities (PacIOOS and HIOOS, respectively). These two were merged in early 2011, thus the data management plans for the insular Pacific and for Hawaii were parallel activities that are now integrated, and the current five-year grant runs through spring 2021. Due to fundamental differences in capacity and infrastructure in the two regions, the functionality will likely remain different for these two components (*e.g*., real-time and forecast data products for Hawaii and map-based GIS services for the insular Pacific); however, real-time data and forecast products will expand into the insular Pacific when feasible.

The fundamental function of the DMS is to aggregate the multiple data streams from the sensors and models that comprise PacIOOS as well as from independent data providers into a central storage and provide these data to users via standard services. More broadly however, the Data Management Group (DMG) is involved with all aspects of PacIOOS data flow including data storage, discovery, and transport. Specifically, the PacIOOS DMG will focus on two essential activities:

1. Store PacIOOS data, and
2. Provide data to end users via standard services as recommended by the IOOC and the IOOS Program Office (*e.g.*, OPeNDAP).

PacIOOS is aware that there is an evolution of IOOS in general and data management in particular. As such, PacIOOS has a process to accept new requirements, internally prioritize, and execute in a timely manner. In order to stay apprised of new requirements, PacIOOS remains in a continuous communication and implementation loop with the IOOC and the IOOS Program Office – from the perspectives of both program management and data management. Additionally, the two lead data management personnel participate in all IOOS level committees responsible for managing data and data management services.

Whenever guidance is provided by the IOOC or the IOOS Program Office on data management protocols, PacIOOS implements and responds as soon as possible. Typically, new requirements are prioritized by both PacIOOS management and the DMG. If there is any question on prioritization, PacIOOS management and the DMG discuss competing tasks in order to ensure base operations are maintained while new requirements are addressed. If a new protocol is not yet considered stable or ready by the DMG, existing tasks take

priority, and the DMG waits until the new protocol is deemed ready for implementation by the DMG and the larger data management community.

Once a system is stable and ready to be implemented, the data management lead tasks his team to respond in a reasonable and timely manner. Since PacIOOS owns its own servers, there is no administrative body between the data managers and the data services. This makes implementation of new requirements/protocols straight forward and efficient. In most cases, it involves the data management lead updating the underlying file format before the data systems administrator updates the PacIOOS server(s).

In addition to the core functions described above, the DMG is closely involved with data and data-service related activities. These include things such as web page development and maintenance, developing and serving data-derived products, and more general "information technology" like support for PacIOOS.

The DMG, including potential roles and responsibilities, is described in more detail in the following sections. All require substantial information technology support in the form of system administration of the various computers and networking between instruments, models and the data server system, and so this additional task is included in the plan. The following sections describe the functional role of the PacIOOS DMS, implementation of these, and a long-range work-plan for the group.

The DMS plan focuses on the management and delivery (*i.e.*, serving) of PacIOOS-related data. PacIOOS will implement recommended and standard practices as defined by the IOOS Data Management and Communications (DMAC) committee and more specifically those in the *Guide for IOOS Data Providers*, *version 1.0* [2006]. These practices apply to data archive, data discovery, data serving (web-based browsing), data transport (binary access to data), metadata, IT security and data QA/QC. Note that the *IOOS Guide* combines metadata and data discovery as one category, and data transport and data serving as one category called data transport and access. The PacIOOS plan separates these, and it does not specifically address IT security since PacIOOS falls under the School and University guidelines and requirements, nor data QA/QC[1], which is to be handled by individual research groups. Successful implementation of this plan permits efficient access to data for end-users as well as integration of the PacIOOS data streams to the IOOS national backbone. The philosophy is to design and maintain a system with minimal complexity to the user, but this typically carries higher complexity to the internal system (higher internal complexity but lower external complexity).

1. Data storage: The PacIOOS DMS provides storage for data streams on a local RAID[2] system. The long-term archive will be provided by NOAA NCEI[3] as appropriate. PacIOOS

[1] IOOS will continue to provide recommendations on real-time procedures via the QARTOD effort

[2] RAID is *Redundant Array of Independent Disks,* essentially a single unit with multiple hard drives, and data are stored redundantly across the disks, so in case of a hardware failure on a single disk data are preserved on another.

data formats are not limited, however most data are composed of NetCDF[4], flat IEEE binary, ASCII, HDF[5], GRIB[6], and GIS[7] formats. Three database systems are employed. MySQL is used for various web browsing tools (*e.g.*, LAS[8] and DChart), maps and geospatial data are stored and accessed using Postgres, and data from real-time sensors are handled with DataTurbine.

2. Metadata: The PacIOOS data sets conform to the Federal Geographic Data Committee (FGDC) and/or ISO standards. Additional IOOS standards will apply when these are fully developed. A PacIOOS web catalog service is maintained to provide access to the metadata.

3. Data discovery: Search utilities will be employed via a web-browser to browse the data stores. These are developed through scripts accessing tables, XML parsing and/or access via relational databases (RDBMS). A Catalog Web Server (CWS) is used to search for data across all data servers.

4. Data display/browse: PacIOOS maintains several web-based data browsing and display tools, including the locally-developed "Voyager" map viewer. In addition, the Environmental Research Division's Data Access Program (ERDDAP) is used for a variety of services, including display and browse. Finally, a Web Map Server (WMS) based on GeoServer, is used to serve maps and geospatial data (mainly for the insular Pacific beyond Hawaii). The THREDDS data server (described next) also provides in-line tools for data browse, such as Godiva-2 and in that respect can be viewed as an additional resource for data display/browse.

5. Data transport: Direct, binary access will be provided through OPeNDAP (Open-source Project for a Network Data Access Protocol), WFS (Web Feature Service), and WCS (Web Coverage Service). The main service is OPeNDAP, and the PacIOOS system is be built around the Thematic Real-time Environmental Distributed Data Services (THREDDS) DODS[9] Server (TDS). TDS is restricted to NetCDF formatted data (as input), but it provides many additional services including "best time-series", WCS, WMS, sub-setting, ISO metadata output and NcML. As stated above, PacIOOS also maintains an ERDDAP server that provides DAP and WMS capabilities. PacIOOS also can provide ftp services, but only via special request as this technology is no longer considered secure. Additionally, services such as "sensor observation service" (SOS) have been depreciated at the IOOS-level, but PacIOOS can provide SOS via TDS if requested.

---

[3] NCEI is National Centers for Environmental Information, a consolidation of NOAA's three national data centers: the National Climatic Data Center, the National Geophysical Data Center, and the National Oceanographic Data Center.
[4] Network Common Data Format
[5] Hierarchal Data Format, primarily used for satellite images
[6] GRIdded Binary format, primarily used in weather applications
[7] Geospatial Information Service
[8] Live Access Server
[9] Distributed Ocean Data System, was the precursor to OPeNDAP and was run through cgi-bin.