

Alaska Ocean Observing System Data Management Work Plan 2/5/10 draft

	Status as of Jan 1 2010	Planned Action Jan-December 2010
DMAC System		
Overall System	<p>The overall AOOS data system has been developed and includes these major functions: data warehouse, data access and discovery, data transport, and archive, and incorporates these as appropriate and available: national standards, metadata, QA/QC protocols, and IT security.</p> <p>All the code and scripts for the data system have been documented and archived.</p> <p>This plan updates the AOOS DMAC Plan for 2007-2011 which was reviewed and approved by DMAC Committee 12/06-1/07 and approved by AOOS Governance Committee 3/07. The original vision for AOOS DMAC was overly-optimistic in what it could achieve under realistic staffing levels. Although the observing portion of the system was downscaled appropriate for the funding received, the data management group did not downscale any of the requested tasks for implementation of the data management portion of the system.</p>	<p>The overall system is running and functional.</p> <p>Current objectives, no specific priority:</p> <ul style="list-style-type: none"> • Increase data holdings; capacity to manage embargoed data • Sustainment of existing assets • Implementation of national data management standards (IOOS) • Web portal revamped to target major audiences: <ul style="list-style-type: none"> ○ Query, discovery, access tool (formerly AMIS) ○ General public portal (operational product) ○ Special project portal of portals (PWS & other DM funded efforts) ○ IOOS specific portal <p>ACTION NEEDED: Prioritization of major DM functions: data ingestion, data system development vs. product development, levels of sustainment of formerly encumbered assets, MOUs and procedures for encumbering new tasks and initiatives as well as encumbering operations that require operational grade reliability and quality.</p>
Data Warehouse	<p>AOOS data is stored on several servers at UAF including the Arctic Region Supercomputing Center’s Tape Silo. There is in-house redundancy of the data. National standards require an additional offsite copy that should be accomplished by the Archive portion of the system.</p> <p>The original goal was to be a distributed system, accessing datasets from other sources. Instead, because most other systems do not meet IOOS interoperability standards, the system has become mostly an archive of data that is given to AOOS and then normalized to meet IOOS standards.</p> <p>Federal agencies such as NDBC have begun implementation of national data transport standards enabling a distributed data system.</p>	<p>Staff time is devoted to ingesting data sets, historical and legacy physical oceanographic data, near real time data streams from Prince William Sound and those relating to ArcOD, BSIERP and SIZOnet projects. There is currently no formal process in place for prioritizing data ingestion. The AOOS Board in the past has indicated priorities for real-time data statewide, the PWS demonstration project data, and other observational datasets that are collected on a routine basis. These guidelines are still too broad for the amount of resources allocated.</p> <p>ACTION NEEDED: Overall guidance from board and DMAC on priorities; user input from broad user community which includes research community on the order in which to ingest the vast amounts of data available in Alaska. MOUs are needed for provisioning of data and products at an operational level. Specific targets need to be set which are not wholly driven by deliverables required for meetings and workshops.</p>
Data Discovery and Access	<p>There are currently 2 home web pages. The entry point for the AOOS data warehouse is the AOOS web portal at ak.aos.org. There is a separate corporate page (www.aos.org) that houses items such as</p>	<p>Data security layer to properly handle embargoed data will be added to the AMIS and IOOS web services.</p>

Alaska Ocean Observing System Data Management Work Plan 2/5/10 draft

	<p>meetings, project reports, policies, committee members, local and national news and events and other information about the Alaska Ocean Observing System.</p> <p>The key to easy data discovery and access is accurate, complete metadata records and a query tool that provides appropriate results.</p> <p>Data is accessible via IOOS web services, custom web services and the Alaska Marine Information System (AMIS) and other visual web applications on the data management web portal. The additional visual web applications are developed on a case to case basis.</p> <p>On-line browse components are available via AMIS and other tools on the AOOS web portal.</p> <p>Internally, data is discovered and accessed through a variety of services:</p> <ul style="list-style-type: none"> • Latest Data (formerly Data Catalog Explorer) • Mapserver • Data Portals (formerly Your AOOS Pages) • AMIS • IOOS and custom web services • IOOS Catalog (visual access to DAP web services) <p>Externally, AOOS data can be accessed through:</p> <ul style="list-style-type: none"> • IOOS Obs Registry • Geospatial One Stop • OBIS • GBIF 	<p>Capabilities will be augmented by improvement to IOOS web services and implementation of the IOOS Portal's Registry, Catalog and Viewer (RCV) which will tie together all of the IOOS regional associations and will advantage of NSF's OOI cyber-infrastructure. The RCV will be initially built and beta tested by the end of September 2010. There will be another round of standardization and optimization of the web services attached to the RCV. The bulk of the work to take place during the next five years of funding by IOOS.</p> <p>OTHER ACTION NEEDED:</p> <ul style="list-style-type: none"> • Specific actions and priority of these actions needs to be established by a Users group. This group may need to be split into three particular audiences as identified by IOOS: General users (public), Power Users (researchers) and Marine Spatial Planning (policy, gap analysis). If resource limited, the requested actions from all three groups need to be merged and prioritized by AOOS DMAC and/or the Board balanced by other tasks in other data management areas. • The general discovery and query service continues to be developed based on user feedback from its BSIERP constituents. • Data management is responding to IOOS Portal needs to ensure the success of access and display of AOOS assets via the RCV. • Continued enhancement of metadata via web services to enable proper access and discovery of AOOS assets. • Experimental data, graphics and applications will be moved into respective data portals. • Public access will be limited to mainstream and operational grade products and materials. • All data and products will be reviewed and removed if users of such information are not identified.
<p>Data Transport</p>	<p>The IOOS Program Office now requires implementation of one or more web service protocols to perform data transport. These web services are SOS, WMS, WCS and DAP.</p> <p>AOOS is compliant with IOOS Program Office requirements as of September 2009.</p> <p>AMIS provides data in alternative formats via an interactive GUI. The available formats depending on the data type are NetCDF, HDF, CSV, GML and KML. These formats support a wide variety of third</p>	<p>Modify AOOS web services to meet FY2010 IOOS Program Goals:</p> <ul style="list-style-type: none"> • Transition DAP web service from Hyrax to THREDDS Data Server (TDS). • Make all publicly accessible AOOS data available using one or more IOOS data protocols. • The entire time series of each dataset should be made available. • Data for 7 IOOS variables will be available in consistent fashion.

Alaska Ocean Observing System Data Management Work Plan 2/5/10 draft

	party display and analysis packages from Google, Matlab to ESRI ArcMap.	The developed web services should be 100% interoperable with the IOOS Program's National Registry, Catalog and Viewer (RCV) once it is developed. Additional standardization across all regional data nodes needs to occur.
Archive/storage	<p>AOOS guidelines require two copies of data be stored: one at AOOS and one at a national center. This has not happened yet for all data in the AOOS warehouse. Standards have not been established for data transport into/out of national archives. For FY2010, archive implementation remains a LOW priority.</p> <p>When space is critical on the AOOS servers, data is moved over to the Arctic Region Supercomputer Center Tape Silo.</p> <p>AOOS-produced data sets may go through a QA/QC program or procedure to ensure data quality before passing it to archive if this procedure is supplied by the data provider. Otherwise, it is assumed the data provider is providing quality data to the system. Only a handful of small procedures and guidelines have been submitted for data. None have been fully implemented due to work being completed on other higher priority data management tasks.</p>	<p>Backups of datasets will continue locally and to ARSC tape silo.</p> <p>Moving data to ARSC tape silo is now semi-automatic. Additional programming is required to make this fully automated. Additional programming will be required to pull data out of the ARSC tape silo based on data quests to the system.</p> <p>A possible improvement is to establish a data feed to NODC to nationally archive relevant datasets. At one time AOOS was working with AEFF to establish PWS data mirror site. This task is still on the books, but has been supplanted by other tasks requested of data management.</p> <p>The IOOS Program Office is developing initial standards for the RAs to follow. AOOS is ready to assist with implementation.</p> <p>Cost-effective solution for archive/storage issue needs to be found or readdress how data management holds certain datasets such as model data which is the largest consumer of space. One option is to store data in a cloud computing area (for cost of \$18k a year per 10 Tb for storage only; additional I/O will incur additional costs to access the 10Tb). A specific problem with storage in the cloud is that it doesn't guarantee data integrity.</p>
Product Development	Numerous stakeholder workshops have identified general user needs for observations and information products. Numerous datasets and products were identified. It was agreed to get the data system and physical oceanographic data installed into the data warehouse. Products were developed opportunistically where possible.	<p>A list of past products, current products and future potential products needs to be developed and maintained describing the original user request, and some way of assessing actual usage.</p> <p>The demands of creating and/or sustaining product need to be balanced with overall system maintenance, infrastructure changes, special projects, and product development.</p> <p>Several proposals for products have been suggested. Many RAs have a user group type committee that addresses these issues.</p>
Standards	<p>AOOS participates in the national processes for developing standards. AOOS currently meets IOOS program requirements for data transport. No other data management components have defined standards.</p> <p>AOOS data management has developed its own internal components</p>	AOOS continues to participate in standards compliance, procedures and establishing new standards through various IOOS committees. IOOS DMAC is currently under restructuring. AOOS participates in bi-weekly conference calls with regional partners and appropriate workshop settings.

Alaska Ocean Observing System Data Management Work Plan 2/5/10 draft

	for elements that remain generally defined by IOOS.	
Metadata	<p>AOOS uses the CSGDM metadata standard. We will use the new ISO FGDC standard when it is ready.</p> <p>AOOS publishes metadata for major data sets to one or more national metadata warehouses, primarily Geospatial One-Stop, although at the present time, GOS does not have any web services for direct access to metadata records. Currently, GOS is unable to harvest AOOS provided metadata. Some investigation is needed.</p> <p>Tight coupling of data to metadata to project information is in place.</p>	<p>We will continue to develop metadata as needed for datasets submitted to the data warehouse.</p> <p>Accurate metadata is needed for both data discovery and data transport. Most datasets acquired by AOOS do not have complete or accurate metadata, requiring additional work by the data team. During the next year, the responsibility for proper metadata will be placed back on the data provider.</p> <p>ACTION NEEDED: Promote use of complete and accurate metadata to all data providers. Failure to take action will result in data not being shown or being accessible via the data management system.</p>
QA/QC	AOOS will utilize QA/QC procedures and requirements presented by the data provider or data consumer. A few procedures have been submitted for data but have yet to be implemented.	In the future, AOOS will work with data providers and customers to establish needed QA/QC criterion for provisioning data. Any QA/QC procedures will need to accompany metadata records. Metadata typically includes data quality and quality control details as part of a research project.
IT security	<p>AOOS follows basic system security protocols for safeguarding systems, software and data through day to day monitoring of operating system and software patches.</p> <p>There is a pending request into the IOOS Program Office to make guidance available for securing systems to NOAA standards.</p>	No additional information has been supplied by the IOOS Program Office. For general consideration, there is the security of the machine itself and security and integrity of the data and information delivered to clients via the Internet.
Other Issues	Status as of Jan 1 2010	Planned Action Jan-December 2010
Performance metrics (including tracking of web hits)	Google Analytics and analysis of web server log files is performed to assess general usage of the data management system.	Additional programming will be needed to do a product analysis of specific web users of the web portal.
Staffing	AOOS staffing: 2007 DMAC Report indicates need for 8FTE. AOOS funding of \$550K covers 3.5FTEs in salary allocated, for example, for quarter time management, full time Data Manager, full time Web Developer/Programmer, full-time IT/Data Ingestion. 2FTE (IT/web/programmer) are required to maintain the system and provide an updated site. Equipment for servers is additional. UAF does not receive direct travel support in its budget.	As of May 2010, most effort is being expended to decommission our site pending the results of this unplanned transition. Our efforts are to insure the integrity of the data we have been asked to acquire and archive.
Space	Space at UAF SFOS provided through generated overhead.	Current space is sufficient.

Alaska Ocean Observing System Data Management Work Plan 2/5/10 draft

<p>Hardware/Software</p>	<p>Several CPUs performing data harvesting, processing, product production, web hosting and provisioning of data via web services.</p> <p>Software in use: CentOS, Apache, Tomcat, PHP, C, C++, Fortran, atlas, cJSON, digir, mx.DateTime, ferret, fpconst, gdal, geos, grib_api, hdf4, hd5, isite, jEdit, java, komodoEdit, ldm, libdap, libgd, libjasper, lxml, mapserver, matplotlib, meld, mod_jk, nco, netCDF, netCDF-perl, netCDF4-python, numpy, pdflib-lite, postGIS, proj4, pydap, pydap_grib2_plugin, pygrib, pyhdf, pyngl, pyproj, python, sciPy, scientificPython, setuptools, soapy, szip, thredds, udunits, udunit2, and xerces.</p> <p>Many custom scripts and programs to perform data ingestion, data processing and product production.</p>	<p>Need to provision for software to test standard IOOS web services in Matlab, IDL, ESRI ArcMap, R, ODV, uDig and Ferret.</p>
---------------------------------	---	--