

Data Management Work Plan

Table of Contents

Data Management Work Plan.....	1
Table of Contents.....	1
Data Management Work Plan.....	2
Data collection	2
Data from the Field Experiment (collection).....	3
Data dissemination.....	3
FTP.....	3
HTTP.....	4
AOOS data portal.....	4
AMIS data portal.....	4
Web Services	4
Custom Web Services.....	5
OPeNDAP.....	5
General Visualization.....	5
Data Management Tasks and Datasets for PWS FE.....	6
Scott Pegau.....	6
Yi Chao	6
Vijay Panchang.....	7
Peter Olsson	7
Mark Johnson.....	8
Jennifer Ewald/Steve Okkonen.....	8
Mike Burdette	8
Rick McClure/Rob Cambell	8
John Oswald/Nathan Wardwell	9
John Whitney/Glen Watabaysh.....	9
AOOS Data Management (Rob Cermak, et al.)	9
Data Management Timeline.....	10
August 2008.....	10
September 2008	10
October 2008.....	10
November 2008.....	10
December 2008	10
January 2009	10
February 2009	10
March 2009	11
April 2009	11
May 2009	11
June 2009	11
July 2009.....	11
August 2009.....	11
September 2009	11
Change History	12

Data Management Work Plan

Data management will largely fill three roles:

1. Data collection
2. Data dissemination
3. General visualization
4. Media/Outreach

Data collection

Data management will need approximately three months to integrate all datasets identified by the field experiment participants prior to the “virtual” experiment or the actual field experiment. If the “virtual” experiment is slated for April 2009, then all required datasets need to be identified by November 21st, 2008.

A summary of requested datasets and formats can be found under **Data Management Tasks and Datasets for PWS FE**. This includes datasets that participants plan to send to data management for integration into the data management system. Requested datasets, formats and field experiment dataset examples need to reach Rob Cermak via email or mail by **November 21st, 2008**. Requests after this date are not guaranteed to be available for the field experiment in a timely manner. Information may be sent well before the deadline to allow for additional discussion and consideration.

After November 21st, 2008, the dataset and task list will be prioritized. A web page will be provided by data management to keep field experiment participants informed of progress. See **data management timeline** for details.

Rob Cermak

Email: cermak@sfos.uaf.edu

Phone: (907) 474-7948

Mailing Address:

Rob Cermak
AOOS/Suite 245
905 N. Koyukuk Dr.
Fairbanks, AK 99775

Data from the Field Experiment (collection)

Method: Data Push/Pull via FTP

Field experiment participants are free to use two available FTP logins for posting and fetching datasets. FTP (IP address: 137.229.40.79). Use one or both, but the general distinction is non-modeling type datasets in the pws account and modeling related datasets and model output are put into the pwsmp account.

Prince William Sound Field Experiment Participants

Username: pws

Password: pwsFE2007

Prince William Sound Field Experiment Modeling Participants

Username: pwsmp

Password: pws2007mp

Method: Pull via HTTP

Information and data that are planned to be regularly posted to a website can also be utilized by the data management system. The SWAN model forecasts and SnoTel/AMBCS station data are examples of data that data management obtains using this method of data acquisition. Datasets available in this manner will typically experience less lag time for assimilation into the data management system.

Data dissemination

Information is/will be available using several distribution methods. Data management anticipates that most participants will utilize the FTP, HTTP, AOOS data portal and AMIS data portal to access datasets. Additional access methods are described for completeness.

FTP

See “**Data from the field experiment**” for access details. This area provides the fastest method to collection information from field experiment participants in their native formats.

The data management system harvests datasets in raw form from the FTP site to produce datasets in other formats, archive and perform general visualization. Automated integration takes upwards to two hours.

Participants must notify Rob Cermak of any datasets placed on the FTP site that need to be incorporated into the data management system. New datasets of an unsupported format will take approximately two to four weeks to integrate end to end.

HTTP

Raw and data management processed datasets will be available from several web pages:

1. Main field experiment page: <http://www.aos.org/fieldexp/index.html>
2. Field experiment 2004 and 2007 data: <http://ak.aos.org/pwsoos/fe2007/>
3. Field experiment 2009 data: <http://ak.aos.org/pwsoos/fe2009/>

Data management will provide general visualization needs to the main field experiment page as needed. The field experiment pages contain HTTP links to raw and/or processed information from field experiment participants. The processed data is in a delayed mode to allow for the automated system to correctly store and restructure information such that it can be made available for subsetting and in other formats.

AOOS data portal

Some field experiment data will become available through various tools and displays on the AOOS data portal. Information on the AOOS data portal is for general consumption by the public. In most cases, only a few datasets will be made available for real time display and download. Most of the data management effort will be for delivery of field experiment datasets by other means.

AMIS data portal

The Alaska Marine Information System will be a primary conduit for searching, subsetting, ordering and obtaining datasets in specific data formats. The AMIS system is currently slated to complete its first roll out by December 2008 or January 2009. Just in time to concentrate efforts on data integration for field experiment participants. Information and tools on the AMIS data portal are geared towards the savvy web user. All datasets requested for integration will be available through this system.

Participants will be able to search for datasets using spatial and temporal type searches. The search tool will lead participants to a data ordering page where datasets can be downloaded in formats of their choice. The data ordering system will notify the user via email when a dataset is ready for download will all its associated metadata.

Web Services

In most cases, integrated datasets will be available through various web services. This may include: SOS, WMS, and WFS.

The IOOS backbone has begun integration of datasets from national, state and regional associations using the Sensor Observation Service (SOS). At present, this service is slated to serve currents, salinity, water level, water temperature, waves and winds. Data management plans to implement this service as time permits while working on wrapping up AMIS and dataset integration for field experiment participants. Examples and more information on how this service works can be obtained from the National Data Buoy Center webpage about SOS: <http://sdf.ndbc.noaa.gov/sos/>

Data management will also provide datasets through OGC WMS and WFS web services. The Web Mapping Service (WMS) will allow participants to quickly obtain map images

of datasets. The Web Feature Service (WFS) will allow participants to obtain point information. The WFS service is very similar to the SOS service. Since these services are OGC compliant they quickly lend themselves to use in ESRI's ArcMap. However, since datasets are typically stored and accessed along the time dimension, it is often hard to view archived datasets at present using ArcMap.

The WMS, WFS and SOS data transport protocols are industry standard and facilitate machine to machine data transfer. The advantage of SOS and WFS is that users may create automated scripts to harvest information from the AOOS data management system. The SOS and WFS web services allow users request data for:

- A station at a particular time and parameter (point)
- A station and parameter for a range of time (time series)
- Any stations in a region using a latitude and longitude bounding box for a particular time and parameter (spatial)

In the majority of cases, these services are used internally by data management to provide graphics to the main project website, AOOS and AMIS data portals.

Custom Web Services

There are a few custom web services that operate much like SOS and WFS, but deliver datasets to a targeted audience. These services are also in use internally for data management. A few datasets below are using these custom web services.

OPeNDAP

The majority of processed gridded datasets are available via OPeNDAP. Some time series datasets are available via OPeNDAP. Data management just makes note of this in case some participants use software such as Ferret, Matlab, IDV or other OPeNDAP enabled package for data access or visualization.

General Visualization

It is anticipated that general visualization needs will be taken care of by three or more general website locations:

- AOOS
- AMIS
- One or more PWS FE experiment pages

The data management priority of support will be:

- Ensure proper collection and assimilation of new data into the data management system.
- Ensure proper dissemination and access methods for PWS FE PIs to obtain data needed to conduct their portion of the experiment.
- Provide general display of information as desired by PWS FE participants.

Data Management Tasks and Datasets for PWS FE

It is very important that data management hears from PWS FE participants if:

- Any statement misrepresents participant involvement with data management.
- Datasets (and/or formats) the participant needs that is **not** listed.
- Datasets the participant will be providing data in a data format that are **not** listed.
- Datasets are listed as needed and are **not** needed.
- Is this dataset or task needed **before, during** or **after**?
 - April 'virtual' experiment
 - July/August field experiment
- Is this dataset or task required or optional?

All information sent well before the deadline will be compiled and revised through November 21st, 2008. After that date, tasks and datasets will be prioritized into a punch list for final implementation in preparation for the 'virtual' exercise in April 2009 and the actual field experiment in July and August 2009.

Scott Pegau

- Examples of drifter data will be needed prior to the experiment to facilitate data integration. May not be possible.
- Possibility of up to three different drifter sources?
- AOOS can pick up the data from a web server or process data sent to the AOOS FTP server.
- Once data is integrated into AOOS, what access/format requirements are required of PIs needing the drifter data?

Yi Chao

- Access to ROMS output TBD
 - Possibly by THREDDs/TDS? Related project with Rich Signal
 - April 2009?
- Collecting surface observations that include SnoTel stations via custom web service. This may be upgraded to the SOS or WFS service when available.
- Obtaining WRF model output in ASCII format provided by AEFf via AOOS FTP site.
- All raw data collected by PWS FE PIs are available online for immediate access while data management defines a storage and access standard. Access at present is via FTP, HTTP and OPeNDAP.

Datasets requested to date:

- ADCP
 - Available via HTTP in ASCII format (already post processed).
 - Plan to reformat into NetCDF using NDBC storage structure which will integrate into the data management system and become available via AMIS, SOS, and WFS.
- CTD

- Available via HTTP in ASCII format (already post processed).
- Plan to reformat into NetCDF using PMEL or AOOS storage structure which will integrate into the data management system and become available via AMIS, SOS and WFS.
- A few PMEL and existing IMS/SFOS CTD profiles are already integrated.
- Mooring (non ADCP data)
 - Available via HTTP in raw format.
 - No definite plans other than to possibly utilize PMEL NetCDF storage structure.
- HF Radar
 - Raw files are processed by AOOS and made available as ESRI Shape (SHP) files.
 - Storage plan needs to be reviewed.
 - Access is available by a custom web service. (Was in use by MMS when HF Radar was operating on the north shore of Alaska).
- SnoTel
 - Surface stations
 - Port San Juan buoy
 - Data is available via a custom web service.
 - Plan is to rework surface data database so that it is accessible via AMIS, SOS or WFS web services.

Vijay Panchang

- Formerly receiving RAMS model output via AEF/ AOOS. Last forecast was June 27, 2008 00Z.
- Receiving WRF model output via AEF/ AOOS via custom web service (REST). Lag time of WRF forecast to AOOS is approximately 8 hours. Processing time in AOOS data management is approximately 2 hours.
- Raw files from AEF may be obtained via FTP.
- SWAN model output is being received from TAMU on a daily basis and is available currently via OPeNDAP and HTTP.
 - OPeNDAP: <http://penguin.sfos.uaf.edu/opendap/forecasts/SWAN/PWS>
 - RAMS: TAMUG_RAM/nc
 - WRF: TAMUG_WRF/nc
 - HTTP: <http://penguin.sfos.uaf.edu/data/forecasts/SWAN/PWS>

Peter Olsson

- Delivered RAMS model output in ASCII format to AOOS via FTP. Last forecast was June 27, 2008 00Z.
 - Output is reformatted into NetCDF files and is available via OPeNDAP and HTTP.
- WRF model output is arriving daily for the 00Z run via FTP using the pwsmp account.
 - Raw WRF model output is available via FTP.

- Model output is processed into the data management system and made available via OPeNDAP, HTTP, custom web service, and graphics generated for the AOOS website.
- An additional ASCII file is uploaded for use by Yi Chao via FTP. No subsequent processing is done with these files.

Mark Johnson

- The HF Radar data files will be integrated into the AOOS data management system. At present, they are now being stored as ESRI Shape (SHP) files as a result of some prior work on the north slope with MMS.
- HF Radar operators will be onsite for a two week period during the peak of the experiment (approximately July 21 to August 3).
- Tentative plan is to make this information more accessible through other web services and via AMIS to obtain HF Radar data in GNOME format.

Jennifer Ewald/Steve Okkonen

- Hydrologic transects in post processed ASCII form are available via HTTP from one of the project data pages.
- Data plan:
 - CTD data will be converted to PMEL or IMS NetCDF structured files so they may be assimilated into the data management system.
 - ADCP data will be converted to NDBC NetCDF structured files so they may be assimilated into the data management system.

Mike Burdette

- NDBC is the authoritative source for meteorological, ocean and ocean current information for augmented buoys for the PWS field experiment.
 - NDBC is currently working on making their entire data holdings available via the IOOS backbone service (SOS). It is currently in testing and does not contain all the buoys for PWS.
 - Archived and near real time information is available in ASCII files and compressed ASCII files in various locations.
- AOOS will harvest this information and integrate it into the data management system. The information will be available through multiple services: AMIS, custom web services, SOS and WFS.

Rick McClure/Rob Cambell

- SnoTel data is available via a custom web service and displayed on various pages at AOOS. Information is not tied into the PWS FE project pages.
- Surface station data (SnoTel)
- Port San Juan buoy providing ocean information. Currently offline (Jan 2009).
- Data is currently gathering using a custom web service provided by AMBCS.
- Plan is to improve data processing of this dataset in conjunction with data from NDBC and make it available via AMIS, SOS, and WFS.

John Oswald/Nathan Wardwell

- Tide gauge information is currently available via three data sources
 - HADS: Hydrometeorological Automated Data System (IDD)
 - COOPS: Center for Oper. Oceanographic Products and Services (SOS)
 - NDBC: National Data Buoy Center (HTTP)
- Tide information is being delivered via the custom web service and on the AOOS website. Information is not tied into the PWS FE project pages.
- The data management plan is to provide this information via SOS and WFS web services and AMIS. Similar access to NDBC and SnoTel information.

John Whitney/Glen Watabaysh

- Need contact info for testing GNOME formatted data?
- Make Yi Chao's ROMS circulation forecast data available to NOAA Hazmat in real-time and in our GNOME format.
- Similarly for HF radar currents data in real-time and in our GNOME format.
- Make Peter Olsson's weather/wind forecast data available to NOAA Hazmat in real-time and in our GNOME format.
- Data management recently found information on how to construct GNOME formatted datasets. We would likely prefer the NetCDF structured files.
- GNOME formatted datasets will likely be available only through AMIS. As to allow specific ordering of datasets and their format.
 - The PWS WRF forecast is on a CURVILINEAR grid
 - HF Radar has a CURVALINEAR grid
 - Ocean circulation forecast data:
 - SWAN is on a REGULAR grid
 - ROMS model output is currently not available, grid type is unknown
- Data management does not anticipate any problem with providing GNOME formatted datasets in NetCDF (or ASCII). However, we urge testing well before the field experiment.
- Data management has a copy of GNOME downloaded as to attempt to do some local testing before making NetCDF or ASCII files available.

AOOS Data Management (Rob Cermak, et al.)

In order to provide the best possible support, data management urges PWS FE participants to review this document and subsequent revisions. Any questions or concerns need to be addressed prior to November 21st, 2008 with a subsequent review during the Alaska Marine Science Symposium January 19th, 2009.

As datasets are implemented into the data management system, participants will be required to test the system to ensure they can access the dataset which are needed. Integration and testing should be completed between January 2009 and March 2009 for participants needing information for the 'virtual' experiment. Integration and testing of datasets will be ongoing from January 2009 to June 2009.

Data Management Timeline

August 2008

29th: Deadline for PWS FE 2009 work plans. The data management plan will continue to be revised until late December 2008 to address any changes to task and dataset priorities when released on December 8th, 2008.

September 2008

Finish and test project database entry portion of AMIS.

October 2008

Finish and test science portion of project database entry of AMIS.

November 2008

Finish implementation and testing of search portion of AMIS.

21st: Final deadline for identification of data management needs and datasets. Information may be sent before this deadline to allow for additional rounds of discussion and consideration. After this date, datasets and tasks will be prioritized by data management.

December 2008

Finish implementation and testing of data ordering and data download portion of AMIS.

8th: Data management work plan released with a prioritized list of tasks and datasets. Comments due December 12th, 2008. A final data management work plan will be available before Christmas break.

12th: Comments due on data management work plan. Propose to revise comments due to January 30th, 2009.

January 2009

AOOS will demonstrate the AMIS system at an information booth at Alaska Marine Science Symposium.

26th: PWS FE data management tasks and dataset integration is in full swing. Begin systematic testing of datasets and data access methods with PWS FE participants. As datasets become available participants will need to test data access methods to ensure they will be able to obtain the data they need for the April 'virtual' experiment and/or the live experiment in July and August.

February 2009

Continued work for PWS FE.

March 2009

Continued work for PWS FE. Final tests for datasets and participants in the April 'virtual' experiment.

April 2009

The PWS 'virtual' experiment. Weaknesses in the data management system will be analyzed and folded into the task and dataset priority list.

Initial access to ROMS PWS model output?

May 2009

Continued work for PWS FE.

June 2009

Final preparations for field work in PWS.

July 2009

21st: Near real time field experiment begins with live data flows from research field into AOOS data management system and disseminated as needed for subsequent analysis and model runs by PWS Field Experiment PIs.

August 2009

3rd: Prince William Sound Field Experiment ends.

4th to 21st: Data management wrap up of PWS FE datasets and metadata.

Data management will continue to work down the task and dataset list if items still remain throughout this period. Otherwise, data management will switch to support and management of existing services and assist with general preparation of graphics for the PWS FE data portal.

September 2009

Remaining tasks for PWS FE will be folded into overall tasks and datasets defined by AOOS DMAC.

Change History

September 2, 2008:

- Scott Pegau correctly points out that the ADCP and CTD data we provide has been post-processed before we obtained it. It is not the raw ADCP and CTD files. These will be converted to NetCDF using PMEL's storage structure. Modified Yi Chao, Jennifer Ewald/Steve Okkonen discussions.

September 3, 2008:

- Added information on Port San Juan buoy that is being provided by Richard (SnoTel/AMBCS/NRCS).

September 22, 2008:

- Conference call: Molly, Carl, Yi, Rob and Nancy.
- Operational field experiment time is approximately two weeks for the HF Radar. The field experiment will be highly concentrated in and around the operation of the HF Radar. July 20-24, 2009 (logistics); July 27-31, 2009 (experiment).
- Virtual command center in Cordova, AK. One day will be an open house with presentations and handouts of various parts of the experiment.
- Mid winter conference with PIs.
- Fall 2009, 2010 AMSS: highlight the results of the experiment. Closeout of the five year period.
- Drifter data pathway?
- NetCDF model output: sample model data. Sample 2004 dataset from Yi.

November 4, 2008:

- Offer an option to run model simulations at ARSC under specified guidance.
- Question about drifter deployments (data flow, QC, etc)

January 14, 2009:

- Updated timetable.
- Updates to tide gage information.
- Updates to Yi Chao (coordinated work with IOOS WSDE and Rich Signell), access to ROMS PWS output possibly in April 2009?
- Blackbox modeling at ARSC offering moved to AOOS Operations and Standards Manual
- Request and correspondence for PI work plans moved to a separate document.