Data Management Planning Workshop – Columbia, SC – November 14, 2005

Hosted by SEACOOS DMCC
Funding SURA/SCOOP through SECOORA

Aim: Kick-start discussions of regional data sharing and data quality to being development of crucial elements for a regional DM Plan.

Goal: Planning to plan. Plan content for a regional data management workshop in Spring 2006 (Chapel Hill).

Spring 2006 Workshop: Focus on data sharing and QA/QC of (but not limited to ocean currents. This blends everything: data sharing, aggregation, display, uncertainties, circulation models, fisheries, search and rescue.

After Spring 2006: Publish two white papers. One on data sharing and one on QA/QC that will help to define specifics towards a regional DM Plan.

Motives:
Network technical folks
Better mix of academic, government and industry
Hear summaries and updates from latest national workshops
Provide technical tutorials
Get everyone up to the same playing level

SEACOOS DMCC funded to participate in national workshops, host today's planning workshop, host regional DMAC workshop, and write two white papers.
Widening the SEACOOS DMCC circle
A chance to have input or impact on regional data management strategies

Meeting Agenda:

830 Introduction and goals – Sara Haines, UNC
845 Overview of SEACOOS – Dwayne Porter, USC
900 NERRS – Tammy, USC
915 SCDNR – George
930 OMS Tech – Steve Browdy
0945 FWC, Kathleen
1000 Break
1015 Summary from OOSTech 2005, Payne
1045 Summary from MMI, Vembu
1115 Summary from QARTOD III, Sara Haines
1145 Summary from Ocean.US , Dwayne Porter
1215 Box Lunch
1330 Planning the Spring 2006 Workshop, Sara Haines
1415 Technical Tutorial- XML Processing, Jeremy Cothran
1500 Break

1515 Technical Tutorial, Maps and Products, Jeremy Cothran
1600 Next Steps (all)
1650 Wrap up, Sara Haines
1700 Adjourn

Overview of SEACOOS
Dwayne Porter, USC, dporter@inlet.geol.sc.edu
www.seacoos.org

SEACOOS Goal: To significantly increase the quantity and quality of environmental information from the coastal ocean of the SE U.S. and make this readily available for a range of societal, scientific and educational applications.

SEACOOS components:
Observing subsystem, data management subsystem, products production and outreach

Data Management Coordinating Committee (DMCC)
SEACOOS DMCC achievements

1) Data commons – protocols and standards
2) Aggregation and display
3) Common data language (netCDF)
4) Data Dictionary – table registers known standards

Other ongoing SEACOOS issues

1) Aggregation formats
2) Storage
3) Normalization – time steps, remotely sensed data, area coverage
4) Visualization and graphics

Data aggregation/visualization

- central aggregation site as opposed to a distributed network
- link to software

Sara- comments - SEACOOS has balanced aggregation with leaving the original data at the service provider (OpenDap) – normalization (time) – being able to shift between different reference systems – provided by the netCDF format attributes – Dwayne – Documentation efforts have been very important

Data dissemination

OpenDAP
Data sharing with Open GIS Consortium (OGC)
Web mapping services (WMS) and Web Feature Services (WFS)

Obstacles/Issues

1. QA/QC of real time data and archives
2. Resource availability and allocation
3. Data from national providers

4. Engaging users/ getting feedback

NERRS - The National Estuarine Research Reserve System
Tammy Small
http://nerrs.noaa.gov/

NOAA NERR Centralized Data Management Office (CDOM)
Was established in 1995 in support of the National Esturarine Research Reserve a nework of 26 reserves in 21 states and territories.

CDMO priorities

1) advancement of system-wide monitoring program data and information management program
2) to maintain the online data and information server http://cdmo.baruch.sc.edu
3) to provide technical support services via phone, e-mail and individual and group training to NERRS

Goal: promoting stewardship if the nation's estuaries through science and education using a system of protected areas

Presently, NERR is using Excel for data processing and database (EQWin) – generating metadata and submitting to clearinghouse – then uploaded in to a CDMO SQL database

Web interface has query -> station location, data type, date boundaries – graphic ability and statistics. Recently switched map to Google Maps presentation

Using ColdFusion (by MacroMedia) along with SQL for graphics presentations of queried data or real-time latest data.

http://www.macromedia.com/software/coldfusion/
http://www.houseoffusion.com/

CMS for SWMP staff - Hidden web site for data collections – macros, manuals, tools, information sharing through forums, calendar on the front page has even calendar

NERRS/CDMO IOOS activities

- NERRS was recently identified as a component of the IOOS national backbone
- In support of this effort the NERRS is undertaking two significant efforts:
- A metadata translation effort, implementation of GOES telemetry for improved data delivery

FGDC content compliant – but not FGDC format compliant. The Metadata translation effort uses Metadoor convert to XML – FGDC compliant metadata

Opensource issues with NERRS - being addressed – they do have QA/QC

documentation.

SCDNR Offshore Fisheries Research and Monitoring Data
George Sedberry, Jessica Stephen, Marine Resources Research Institute, Baruch
SC Department of Natural Resources
http://ekman.csc.noaa.gov/seageofish/viewer.html

Utilization of ArcGIS for mapping.

Online database covers original historical trawl surveys, plankton surveys and some historical hydrographic data from Cape Canaveral to North Carolina. It also includes spawning locations, drifter data, tagging data and vertical temperature profiles from vertical migration data.

NOAA Southeastern U.S. Deep-Sea Corals Initiative SEADESC: The working Model – Tim Birdsond, SEADESC Meeting Apr 28-29, 2005 UNC Wilmington, Center for Marine Science – Developing an online Atlas – interactive Map by region
NOAA Explorer – will include cruises with layers
Steve Ross (UNCW) –

Data Sharing and Observations Systems
Steve Browdy – OMS Tech - steveb@omstech.com

OMS Tech is a private company – scientific programming and analysis. Clients include: University of Miami, SeaKeepers International.

Main Points
Data Transmission and Control

- Inmarsat-C used by SEAKEEPERS, also looking at Iridium – because it is 2-way communications for instrument and project management
-

Archival, Retrieval and Storage

- relational database model is the most used, strong points: great for inserting data, weak points: getting data out (especially in cases of having so many data tables)
- post-relational database (sits between relational and object-oriented), based on a mathematical technology called sparse matrices. Minimal footprint in terms of storage – retrieval is very fast. Supports new data records based on old data records. So if a new sensor is added, you can define a new record based on the old field with the addition or change for the new sensor. "Inherited records".

Presently no opensource, although PostGres does have some level of inheritance – but the proprietary post-relational is much faster.

Proprietary -> Cache' - Intrasystems.com

In one example, data acquisition devices that were being deployed had one acquiring

module with many different sensors – transmission record was a solitary record, data record variations produced a need for other database records that kept track of changes.

Issues:

- Aside from researchers – who are the data use clients – how do you get the information about who the potential users and present users how and how effective the data dissemination is?

- Two types of data retrieval options should be offered on-demand and batch access. Using metadata – the on-demand versus batch request can be determined – an algorithm can determine the best delivery method depending upon the request.

- Data formats – open standards are nice, however the standard process can be so slow, it is better sometimes instead of waiting to have data standards

AJAX – Asynchronous Java Script and XML – can give a rich client feel without having to write a client sided application. It may play a big role in making data dissemination available better so web pages can be changed piecemeal without having to reload the full page.

Data QA/QC
System should provide an infrastructure so that everyone can "plug" into. The idea is to submit data to the infrastructure "server" which provides the following:

Exception handling, work-flow management, automated usage of models, metadata management, common usage of QA/QC flag definition. You could use a standard code or utilizing a user specified algorithm.

AJAX sites

http://java.sun.com/developer/technicalArticles/J2EE/AJAX/index.html?cid=59754

http://www.clearnova.com/ThinkCAP/servlet/LayoutMgr?Transaction=/dcont.htm&NAME=%2Fcn%2Fthinkcap%2Fmain_content&LAYOUT=%2Fcn%2Finside&subhead=thinkcap

Facilitating Data Management
Kathleen O'Keife, FWC

The Fish and Wildlife Research Institute (FWRI) (used to be the Marine Research Institute), Florida Fish and Wildlife Conservation

FWRI Projects are using GIS for:

- Monitoring & Mapping – Statewide benthic habitat, environmental sensitivity

index mapping , SEAMAP South Atlantic Bight Bottom Mapping
- Education
- Application Development

Internet Map Sites (visualization of data)
Data Standards Committee developed standards
Information Sharing – metadata
Data sharing tools – portals, IMS,SVG,XML web services
Data Integration tools – Oracle Discoverer, Metadata map server

Utilizing SMMS – Spatial Meta-Data Management System
GIS & Non-spatial scientific research projects require metadata at FWRI
Metadata Cooridinator & staff in each section submit FGDC-compliant metadata
Naming Convention – spent a lot of time figuring out how to name things so users can find it
Geographic extent, precision, spatial reference
Naming convention:
What(&scale)_where_who_project_when_topology

Archived Data – two oracle databases – just rastered data, the other is the rest
Archive ArcSDE – media storage, digital archives
Project Finder – Click anywhere on the map of Florida to see what projects we have ongoing in the area. Or slect project using query tool

Summary of OOSTECH Oct 24-26, 2005, Baltimore
"Web Services for Interoperable Ocean"
Payne Seal and Jeremy Cothran
http://twiki.sura.org/twiki/bin/view/Main/MinutesOOSTech2005

Dan Reed, gave the keynote speech.
Key points:
Benefits of standards, interoperability
Pre-requisites for web services
40% of e-bay calls are API calls

Phil Bogden
Key points:
There is ocean observing, there is a need, need to strengthen community
Goals: virtual organization new distributed labs, hazards planning, research-> observing systems.

Web Services – Machine to Machine interaction designed to maintain platform and language independence.
Technologies:
Mapping XML to Java – JaxP – allows parsing of XML documents via Java – also JaxB

– binding tool - Sun initiatives
XML doesn't care about display orientation like html – more concentrated on exchange of information

Trillion Grid – virtual data toolkit -

AJAX – covered in OOSTECH – google maps is a good demonstration of it

Two methods of securing messages – secure pipe or a secure message layer

Oostech Working Group to work with the DMAC steering committee

Video Services presentations were videotaped – both morning and afternoon sessions – implementation of Salinity_slim

Summary of Marine Metadata Initiative (MMI) Workshop
"Advancing Domain Vocabularies", August 2005, Bolder, Co
http://marinemetadata.org
Vembu Subramanian, USF, vembu@marine.usf.edu

The Marine Metadata Initiative
NSF and SURA funded
Initially one year project (Sep 2005)
Deliverables, community web site, interoperability demonstrations

Originally based on GCMD, BODC and CF data standards. OWL files have been generated for over 40 existing vocabularies

Workshop breakout groups initiated development of domain ontologies using VINE tool
Demonstration of interoperable data access using metadata and vocabulary mapping ontology

Tools:
Voc2OWL interface
VINE tool does the mapping

Projects:
Tethys

The Tethys project demonstrates semantic mediation by developing an end-to-end process for finding data in distributed and heterogeneous repositories and making that data accessible. MMI has worked for about six months with various systems (such as AOSN, SSDS, and SEACOOS), developing an architecture and an implementation code to facilitate interoperability based on metadata and semantics mediation. To improve the system (and our own skills) we invite you to be part of Tethys. You will benefit by

having hands-on experience in implementing SOAP web service, creating rich semantic metadata using RDF Dublin Core, and publishing the core vocabularies of your organization in ontologies in OWL. At the end we will have created an easily accessible portal to access distributed resources based on semantics.
Tethys URL:
http://marinemetadata.org/examples/mmihostedwork/demos/accessthrumetadata/index.html

Summary of QARTOD-III Meeting, San Diego, CA
Quality Assurance for Real-time Oceanographic Data

Sara Haines, Jeremy Cothran, Vembu Subramanian
QARTOD URL: http://nautilus.baruch.sc.edu/twiki/bin/view/Main/WebHome

Definitions of QA/QC:
QA – Quality assurance steps taken beforehand (calibration, maintenance, robust data communication
QA – quality assessment – characterize error and uncertainty
QC – steps you take after assessment to support the delivery of high quality data

Previous QARTOD goals accomplished:
Build work done at Qartod-II for waves, in situ currents, remote currents, add CTD and chlorophyll working groups
For waves: http://cdip.ucsd.edu/?nav=recent&sub=nowcast&xitem=socal_now

QARTOD III Goals
Identify and refine specific tests
Identify specific QC
Capture of relevant information – worry about FGDC

QARTOD III Group breakouts
1. In-Situ Currents (ADCP) group
2. CTD group
Built on Salinity workshop and early QARTOD temperature workshop
1 methods collection
2 directly measured
3 derived parameters
4 additional sensors
5 no brainer test
6 other tests
required and recommended tests for each measured parameter with specific definitions for climatology, range tests, etc.

3. Remote Currents (HF Radar) Group

Need more documentation from vendors – error flag meanings, signal to noise ratios etc.

4. Chlorophyll group
Identify and characterize different types of measurements: real-time measurements, in situ and discrete samples

Summary from OceanUS DMAC meeting
Dwayne Porter
DMAC Steering Team Meeting Nov 01-03, 2005
http://dmac.ocean.us/

Recently formed NFRA (National Federation of Regional Associations)
http://www.usnfra.org/

Will agree upon and map out efforts between national and regional DMAC communities

Identify core expert teams for the identified important technical issues:

- Metadata & Discovery
- Archive
- Transport & Access
- Standards Process

Mission Now

- Serving data – capture best practices, lessons learned, etc
- Collect/"publish" IOOS-wide data and instrumentation catalog(s)
- Conduct Pilot to connect an RA to an archive with automated data

SEACOOS Planning the Spring 2006 Workshop
Data Sharing and Data Quality focused on Ocean Current

Possible Dates:
Feb 13-14 (M-Tu)
Feb 20-21 (M-Tu)
Mar 1-2 (W-Th)
Mar 6-7 (M-Tu)
Mar 9-10 (Th-F) Preferred

What:
Data Sharing for an Operational System: ocean currents, one water quality parameter

Who: Send in names and suggested agency participants to Sara Haines by Dec 15
Possible Attendees outside of SEACOOS include:
National Park Service (NPS)
National Data Buoy Center (NDBC)
USGS

National Ocean Survey (NOS) [Ports and HF Radar]
South Florida Water Management District
Horizon Marine

Suggested Content:

- longitudinal integration, state water quality and estuarine programs, to offshore observational networks
- identify tools, formats and efforts needed to promote regional data sharing
- tools for data sharing
- smaller groups want to send data in the format that they are familiar
- on-the-fly translator
- web services (like soap)
- identification of formats and data types in the region presently not in SEACOOS commons
- identification of potential Users and their needs

Potential User base:

- Search and Rescue
- Fisheries
- Resource Managers
- Modelers
- Emergency Response (HAZMAT, oil spill, habitat)
- Forecasters and responders for natural disasters

Technical Tutorial – XML Processing
Jeremy Cothran, USC, jcothran@asg.sc.edu

Powerpoint presentation can be downloaded at:
http://nautilus.baruch.sc.edu/documents/presentations/XML_Tutorial_JCothran.ppt

Examples of XML processing and utility:

Salty Slim
http://twiki.sura.org/twiki/bin/view/Main/SalinityWorkshop
Carolinas Coast
http://nautilus.baruch.sc.edu/Carolinas/carolinascoast.php
SEACOOS XML services
http://nautilus.baruch.sc.edu/twiki_dmcc/bin/view/Main/CodeRepositorySeacoosXMLServices

Technical Tutorial
Maps and Products using MapServer, postGIS, ImageMagick and GMT

Jeremy Cothran, USC, jcothran@asg.sc.edu

Mapping Process:

- Data table creation
- Unique index (for duplicate rejection) definition
- Add PostGis 2 dimensional geospatial reference column type
- Add geometries
- Map
  - Map file defines (.map) defines connection between layer and the database query