

Space Details

Key:	DAG
Name:	Data Aggregation
Description:	
Creator (Creation Date):	kgomes (Nov 13, 2006)
Last Modifier (Mod. Date):	kgomes (Nov 13, 2006)

Available Pages

- MBARI Data Aggregation (Home) 
 - Meeting_2006.11.10_Francisco
 - Meeting_2006.11.10_Technical
 - Meeting_2006.12.09_Francisco
 - Meeting_2007.01.12_Francisco
 - Meeting_2007.01.26
 - Meeting_2007.02.08
 - Meeting_2007.03.01
 - Meeting_2007.03.20
 - Meeting_2007.03.29
 - Meeting_2007.04.09
 - Meeting_2007_06_13

MBARI Data Aggregation (Home)

This page last changed on Apr 09, 2007 by [bermudez](#).

This is the home page for the MBARI Data Aggregation space.

Introduction

Several subgroups have been talking about data aggregation ideas at MBARI. The idea of this project is to instigated technical and requirements discussions to have a coherent strategy and move towards a one-stop-shopping center for MBARI data; not only for a portal design but also to have external data access by programatic mechanisms (e.g. web services). The discussion will center on what each subgroup is doing, fomenting subgroups collaboration and avoiding duplication of efforts.

Documents

[Latest strategic plan for data aggregation](#)
[Aggregation Strategies](#)

Meeting_2006.11.10_Francisco

This page last changed on Apr 06, 2007 by [bermudez](#).

Nov. 10, 2006

Attendees

- Francisco Chavez
- Seth Bushinsky
- Julien Barde
- John Graybeal
- Duane Edginton
- Luis Bermudez

Purpose

Present to Francisco Chavez a data aggregation strategy for MBARI. Get a use case and requirements.

Highlights

- Presented the [conclusions](#) from the data aggregation technical meeting.
- Presented an aggregation strategy and the following list of parameters (extracted from John's graybeal HTML query form)
 - backscatter
 - chlorophyll
 - current direction
 - current speed
 - co2
 - fluorescence
 - iron
 - nitrate
 - nutrients
 - phosphorous
 - pigments
 - salinity
 - temperature, air
 - temperature, sea surface
 - temperature, water
 - turbidity
 - wind direction
 - wind speed
 - wave height
 - wave interval
- A scientist should be able to:
 1. download all the data in ASCII from one platform (deployment). For example all data from M0
 2. For a property, get data from different platforms (e.g. AUV, ship and moorings) at a time when they were all collecting data
 3. Goal is to make accessible MBARI column's data from one location

Actions

- Seth will continue to do his work (addressing n.1)
- Julien will address n.2
- Luis will create a list of platforms and sensors and monitor the state of aggregation

Next meeting

Dec 8th 900 AM

Meeting_2006.11.10_Technical

This page last changed on Dec 09, 2006 by [bermudez](#).

2006.11.10

Attendees

- Julien Barde
- John Graybeal
- Duane Edgington
- Luis Bermudez
- Mike Godin
- Brian Schlining
- Kevin Gomes
- Mike McCann

Purpose of the meeting

Discuss in a technical group the data aggregation concept at MBARI to have a coherent strategy to propose to Francisco Chavez.

Minutes

- Luis B. presented these 2 documents to frame the discussion:
 - Data aggregation ideas: [MDA.pdf](#)
 - MBARI-Related Data and Infrastructure Projects (by John G.): [DataProjects.doc](#)
- Aggregation is a function of the scientific audience. It depends on the use case. Need to further investigate this topic with Francisco.
- Thredds has limitations in terms of aggregation. A dimension in all the data sets have to be compatible in order to aggregate by that dimension. Need to perform interpolation. Thredds uses the NectCDF aggregation library, so Thredds is really not necessary.
- When two heterogeneous data sets are put together like CODAR and AUV data it makes good impressions to scientists
- Aggregate by platform makes sense.
- Andrew Chase did previous work using WMS - which we could leverage
- Mike Godin is working on MoQua to not be depended on LAS.
- As a first step taking aggregating data using OPenDAP and NetCDF makes sense. This will be a baby step in the design of a more general aggregation framework.

Actions

- Discuss next meeting with Francisco the concept of Aggregation - using NetCDF files and set a use case

Meeting_2006.12.09_Francisco

This page last changed on Jan 02, 2007 by [graybeal](#).

Attendees

- Mike McCann
- Francisco Chavez
- Reiko Michisaki
- Seth Bushinski
- Rich Schramp
- John Graybeal
- Duane Edginton
- Julien Barde
- Luis Bermudez

Overview of the Data Aggregation Strategy

(Not presented in the meeting, but added here for clarity)

[Aggregation strategy image](#)

Status BOG

Mike - Prototype by the end of December. Mike M. showed dapper/dchart, where 50% of BOG data is already available. Process consists of converting BOG projects into NetCDF files and uploading them to Dapper. There are some limitation of Dapper, such as presenting profiles at various depth.

Status Moorings to OGC / SWE

Luis - working on it - plan to finish end of December

Status list of devices and the status for data aggregation

Luis - in progress ...

Status Use case - Mooring AUV and Ship

Julien. Made a [presentation](#) containing abstraction strategies from the special use case. Julien presented his tool and how to catalog metadata to be searchable. The group discussed that use case cannot be solved using a metadata catalog system (e.g. SSDS), a data base that has spatial-query capabilities will serve this purpose better. Mike suggested that getting mooring data in a similar format of how the BOG files are being created and upload them to Dapper could address Francisco's use case. Dapper uses MySQL as a back-end to increase the performance on queries.

Status to-friendly-ASCII

Seth - almost done with mooring raw data. Francisco wants the plots to quality controlled them.

Status Big Picture

Getting data to NetCDF holds as the strategy. We will need more metadata about a NetCDF files to be able to wrap this file in other formats and metadata interfaces. Two possible ways: add more metadata in the NetCDF files or use SSDS to get more metadata information. For the later all the NetCDF files created need to be register in SSDS. Mike M. can add to header as needed. To have a one stop shopping center based on SSDS will take about 3 - 6 months of software engineer with Kevin's expertise.

Action items for the next meeting

- Luis will talk to Mike Godin about how did he handle NetCDF files in the MB06 project. If he was using SSDS to register the metadata or other approach.

- Julien and Mike will meet and discuss the Francisco's use case. This will illuminate us about dapper capabilities. Questions to answer for the next meeting: Is Dapper a good end user interface to provide a fast searchable access to MBARI column data ? What needs to be done (e.g. conversion of NetCDF files in special formats (sequence, grid, timeSeries ??) for this to happen ?)
- Seth will not working any more on producing ASCII files. He will work on producing plots to do quality control.
- Luis will continue working on extracting OGC Services from NetCDF files.
- Mike will add "device type" to the header of the NetCDF files. He can illuminate us about the standard way (or best practice) to do such.

Next steps

- Need to make available ROV/CTD files to outside of MBARI. Need NetCDF files from this data base.
- Discuss how a one stop shopping center will work for MBARI taking into account the different possibles functionalities. (Plot, download, etc..)

Meeting_2007.01.12_Francisco

This page last changed on Jan 12, 2007 by [graybeal](#).

Attendees

- Mike McCann
- Francisco Chavez
- Seth Bushinski
- Julien Barde
- Luis Bermudez
- Mike Godin
- Kevin Gomes

Review action items

- BOG in OpenDAP- Mike M.
Rich S. making a lot of the work. All the profile data is being made available via [DAPPER/DHART](#). Data was loaded in 2 ways: per project and all profiles aggregated. The later had problems that only a maximum of 1000 profiles are shown. Process was: 1) create NetCDF files from BOG. 2) load attributes to Dapper. DCHART, which is an AJAX application reads from DAPPER.
- Mooring data from NetCDF available in OGC / SOS service for OOSTETHYS - Luis B. (3 min)
Luis working from NetCDF files created by Mike M. He had some discussions with Mike G. about java objects use to store netcdf data/metadata and to pass to an SOS service
- THREDDS aggregation update - Julien B. (3 min)
Is working on it, problem is metadata that is stored and is overwritten every time the catalog is generated.
- Use case - [Mooring and AUV] - Julien B. (3 min)
Experiment with a PostgreSQL and spatial cartridge, that allows to perform spatial functions, such as: get near by. Using as source DAPPER.
- MOQUA plans for data aggregation at MBARI - Mike G. (3 min)
Is a king of DCHART, is going to be build on top of hypersonic, light database easy to put in a WAR file. Intention is to write a generic reader that can sit on top of Thredds, Dapper or an OGC catalog.
- Review/discussion DATA Aggregation Strategy (45 min)
Luis presented a diagram and general questions for discussion [Aggregation strategy issues for discussion](#)

Strategy discussion

Image comments

- DAPPER and DCHART should be two boxes
- MOQUA interfaces with DAPPER and not DCHART
- NETDF REPOSITORY goes directly to DAPPER, DAPPER doesn't interferes with THREDDS.
- OGC interfaces can go directly from DAPPER
- present diagram in UML

General comments

- SSDS, DAPPER and MOQUA are complementary and do not overlap. SSDS concentrates in provenance, while DAPPER being a relational data base, concentrates in the most efficient schema to support NetCDF (OpenDAP) queries. MOQUA is a user interface that tries to aggregate heterogeneous data. An interesting project (for outside funders as well as MBARI) is how to integrate these different technologies, to allow geospatial functionality and more standard serving capabilities (like serving ISO metadata and making available data via OGC interfaces)
- Francisco in charge of starting the thread about a proposal from CENCOOS to NOAA CSC RFP. - we can fold the previous ideas to work on the case 3 of the proposal to getting us started.
- Carbon and Nitrogen Initiative is in its planning phase, and due to the heterogeneity of data being involved, it can be a good test-bed for the technologies being discussed in this project.
- Is not true that the work to be done here is not funded, Julien and Seth are funded by CIMT, and it is also possible to have some SSDS time next year.
- This should start folding in to an internal project to be presented next June.
- Use case is tough due to the 3 dimensionality of the problem.

- DAPPER is an aggregator and seems to have a lot of the functionality that THREDDS has.

Action Items

- Start writing a proposal about the idea of integrating SSDS, DAPPER and MOQUA with OGC capabilities (it should be consistent with the CI proposal send by MBARI to ORION). - Luis
- Start building the architecture (updating the image) for data aggregation more formally - Luis , Kevin and Mikes
- Setup meeting next week with CENCOOS to propose a concrete idea for the proposal - Francisco
- Continue testing the spatial functionality in POSTGRESQL and then we can think about how to add the functionality to the technologies that we will use (e.g. DAPPER) - Julien
- Invite Luis to next Carbon Nitrogen discussion meeting - Francisco
- Luis meet with Julien to talk more about OGC and OPenDAP

Next Meeting

Friday - January 26th 9:00 AM - Ocean's View conference room

Minutes Data Aggregation Meeting Jan. 26, 2007, 9:30 AM Oceans View conference room.

Attendees

- Mike McCann
- Francisco Chavez
- Seth Bushinski
- Julien Barde
- Mike Godin
- Kevin Gomes
- Reiko Michisaki
- Luis Bermudez

Topics

CENCOOS and NOAA CSC proposal

Francisco explained the 3 focuses of the NOAA CSC proposal. We confirmed that we would like to move forward #3, which is related to the data aggregation idea but in a regional level. Luis will start a draft about the ideas to have something ready on Monday to send to other CENCOOS partners.

Data Aggregation as an Engineering Contingency Project.

Luis is going to draft a proposal for Engineering to have this project as a contingency project. So some of us can charge to it, as well as organized ourselves better to write a successful internal proposal for MBARI. Need to create cons and pros of the technologies that we have and as they fit together.

Use case - Julien

Julien talked about some experimentation with different data bases - including Dapper. He will presenting a proposal to Duane and Francisco about what he is going to concentrate on in the future.

How does it looks a one stop shopping center for data at MBARI?

Answer was - a THREDDS server is the first step, and then other services could be build on top of them. - need a comparison of technologies.

Carbon Nitrogen Project

Francisco presented the overview of the project. In essence the project is going to capture data using different platforms (Floats, gliders and cruises) that need a data management component. Also he talked about the importance of being able to access national data sets like ARGO and satellite data. Our aggregation strategy should be able to facilitate the data aggregation for this project. Kevin has 10 days allocated and is participating in the meetings.

Action Items

- Luis - create draft internal MBARI proposal
- Luis - start product comparison in the wiki
- Luis - create mailing list
- Luis - draft ideas to respond to NOAA CSC
- Julien - develop and send proposal to Duane and Francisco

Minutes Data Aggregation Meeting Feb. 08, 2007, 10:45 AM Oceans View conference room.

Attendees

- Francisco Chavez
- Mike Godin
- Jim Bellingham
- Luis Bermudez

Topics

Comment: A lot of the participants were not available. A short discussion took place about MOQUA role in the Data Aggregation plan. Not all action Items were discussed.

CENCOOS and NOAA CSC proposal

Francisco send the proposal. Other CENCOOS system engineers agreed with what was being proposed.

Data Aggregation as an Engineering Contingency Project.

Luis has the month of February to sent this to Duane

MOQUA role in the Data Aggregation plan

It was suggested that MOQUA should be the user interface for the data being made available in the data aggregation project. A meeting was set for the 13th with Kevin G. to talk more about SSDS and general MOQUA interfaces.

New Action Item

- Luis, Mike and Kevin meet to talk about MOQUA interfaces

Action Items from previous meeting

- Luis - start product comparison in the wiki
- Luis - create mailing list
- Julien - develop and send proposal about his work at MBARI to Duane and Francisco

Next meeting Thursday March 1st 9:00 AM Oceans View Conf room.

Meeting_2007.03.01

This page last changed on Mar 01, 2007 by [bermudez](#).

h1 Minutes Data Aggregation Meeting March 1st 2007, 9:00AM Oceans View conference room.

Attendees

- Francisco Chavez
- Mike Godin
- Kevin Gomes
- Luis Bermudez

Topics

MBARI - Contingency Proposal

Luis presented a draft. Kevin, Mike G. and Luis said that they could assign 30 days this year for the development of the required interfaces. Mike Godin said his time could be cost share. Need to ask Mike M. Julien should have all his time allocated to this, meaning that what he is going to do should align very well with this strategy.

Julien Postdoc proposal.

Julien sent his proposal to Francisco and Duane. Julien said it was OK to share the proposal with the group. Francisco will send it.

NOAA CSC Proposal

Few days we will now about it, I f accepted a proposal will be due by April

Meeting SSDS and MOQuA

It was clear that we need to define interfaces between SSDS and MOQuA. Mike G. experiment with THREDDS and next week will experiment with SOS.

Others

- A dataaggregation list has been created in LYRIS list manager http://listserver.shore.mbari.org/read/all_forums
- Next meeting Tuesday March 20th 9:00 AM Oceans View Conf room.

Minutes Data Aggregation Meeting March 20th 2007, 9:00AM Tank View conference room.

Attendees

- Francisco Chavez (Remotely)
- Mike Godin
- Kevin Gomes
- Luis Bermudez
- Julien Barde

Topics

Overview of data aggregation framework, roles and problems

We need to figure out:

- Minimum good metadata (need to add to SSDS, NetCDF or THREDDS). Mike G. and Luis B. will meet
- Build a Thredds server (Mike Godin will work on it. He will ping Mike McCann). Best serv is Elvis.

Proposal NOAA CSC - IOOS

Panel rated the LOI as excellent, good and very good. So we can move forward with the proposal. We need to generate about 7 pages of technicalities, plus respond to the comments.

Summary of Comments:

- We have to express more benefits to partners
- What will the data consumers gain ?
- Partners have to express that they are committed to the standards we are describing. Need to provide details of stronger participation by CenCOOS partners.
- Provide duration and total funds
- Suggestion of collaboration with SEACOOS.

TODOs:

- Francisco C. will work on the partners to see if they can finish the proposal to get more engagement from there side. He will meet with with Luis B to outline to response
- Luis will contact SEACOOS
- Only impact on Eng. resources will be Luis. TBD.

Contingency Time Proposal

Luis met with Duane. Need to specify more the details of the outcomes and time use by Luis. Need to specify how the work being done by others relate to what we are doing.

Others

Need for storage - MB06 has about a terabyte due to the numerical model output. He will talk with Todd R.

Next Meeting

March 29, 9:00 AM Ocean's View Conf Room

Meeting_2007.03.29

This page last changed on Mar 29, 2007 by [bermudez](#).

Meeting_2007.03.29

Minutes Data Aggregation Meeting March 29th 2007, 9:00AM Ocean View conference room.

Attendees

- Francisco Chavez
- Mike Godin
- Luis Bermudez
- Julien Barde

Topics

General progress

THREDDS server - not know

OceanObs

- run by Josh Pederson (Tom Wadsworth)
- Jim Boyed - using Ocean obs for the national registry
- Heather said proposal makes sense and there is no conflict with other CENCOOS proposals

NOAA CSC Proposal

- we have commitments
- Ask each of the partners for about one month/person of work. PFEL and MBARI get 4-5 months funding.
- Overall budget was increased to \$45 Million

TODOs

- [Mike M.] check with Mike G. about Thredds
- [Mike M.] Next meeting Mike M. prepare talk about Hyrax
- [Francisco] will send budget note (read above)
- [Luis] finish contingency proposal

Next Meeting

April 9th, 2007

Comments

Mike McCann, not Mike Godin.

Posted by [mccann](#) at Mar 29, 2007.

This page last changed on Jun 13, 2007 by [bermudez](#).

Meeting_2007.04.09

Minutes Data Aggregation Meeting April 4th 2007, 1:30 PM Ships View conference room.

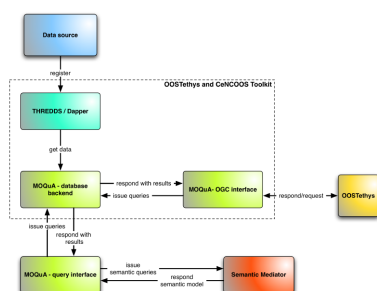
Attendees

- Edgington Duane
- Francisco Chavez
- Jim Bellingham
- Mike Godin
- Luis Bermudez
- Julien Barde
- Kevin Gomes

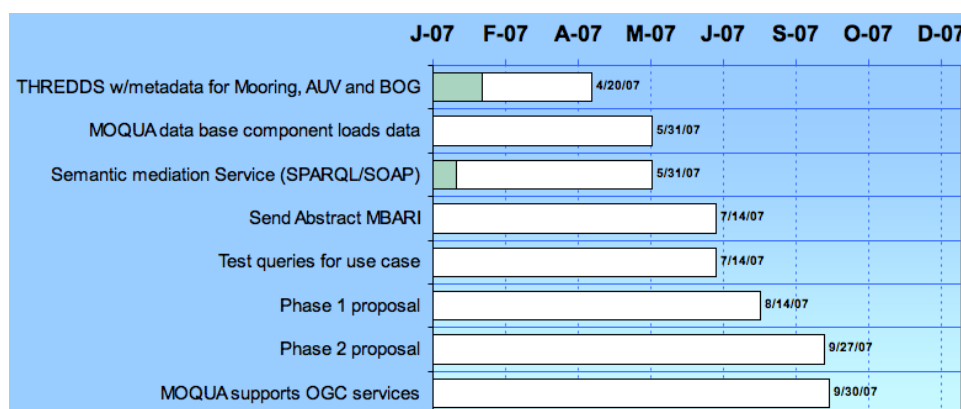
Topics

Discussions centered around the figures presented bellow. Participants agreed on the proposed strategies for short term road map, contingency proposal and NOAA CSC proposal.

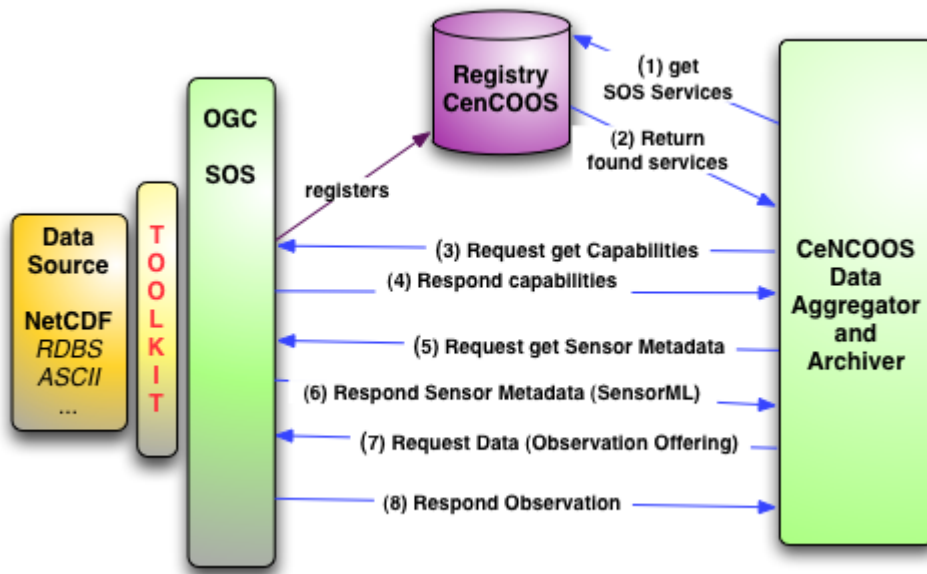
Road map



Contingency Proposal



NOAA CSC Proposal



Meeting_2007_06_13

This page last changed on Jun 20, 2007 by [bermudez](#).

Meeting_2007.06.13

Minutes Data Aggregation Meeting June 13th 2007, 2:00 PM Ocean View conference room.

Attendees

- Francisco Chavez
- Jim Bellingham
- Luis Bermudez
- Julien Barde
- Kevin Gomes
- Mike McCann

Topics

- review milestones that were detailed in the last meeting:
<http://oceana.shore.mbari.org:8081/x/WwAJ>
- Mike G. he has a database which is being populated with different types of data, purpose to perform fast searches. Not sure how much data is being stored. Need MOQUA to get data.
- MBARI Proposal plan (abstract due 7/14)
Show progress and what does it needs to access all MBARI data from one source (technology implementation and resources). Need to show more examples. An enclosed image was created by Francisco. <http://oceana.shore.mbari.org:8081/download/attachments/1179719/image-Jun13-2007-proposal.jpg?version=2>
- Status NOAA proposal
No news
- Others
participants should present progress in their work which related to this activity

Next Meeting

Wednesday June 20 2 PM

Comments

Update image - missing a line

Posted by [bermudez](#) at Jun 20, 2007.
