EARTHCUBE END-USER DOMAIN WORKSHOP REPORT

Articulating Cyberinfrastructure Needs of the Ocean Ecosystem Dynamics Community

Woods Hole Oceanographic Institution
Woods Hole, MA
7-8 October 2013

November 29, 2013

Editors: Planning Committee Members

Table of Contents

WORKSHOP SUMMARY

An EarthCube Water Column Domain End-User Workshop hosted by the Biological and Chemical Oceanographic Data Management Office (BCO-DMO) was held October 7-8, 2013 at Woods Hole Oceanographic Institution. The goal of the workshop was to articulate cyberinfrastructure needs of the ocean ecosystem dynamics community with particular focus on the challenges presented by multi-disciplinary marine ecosystem research that requires investigations in four dimensions.

The workshop included 50 participants in the domain of oceanic ecosystem dynamics (established and early career researchers, teaching faculty, graduate students, postdocs, data and information managers and cyber-related researchers) to explore and document the community's cyberinfrastructure needs from the users' viewpoint. The participants self-identified into one or more interest areas: biological, chemical, and physical oceanography (68%), data and information managers (42%), computer science researchers (22%), education and social science specialists (14%), and modelers (8%).

Participants identified five high-priority science questions that will be the focus of interdisciplinary efforts during the next 5-15 years. These include aspects of ocean acidification; impacts of changing ocean circulation and warming on upper ocean ecosystem dynamics; impact of changing ice cover in the Arctic Ocean on other ocean regions as well as the impact on arctic ecosystems; how physical, chemical, and biological processes come together to create more complex emergent properties in ocean ecosystems; and how anthropogenic impacts other than climate change influence marine ecosystems.

A number of themes emerged as consistent challenges faced within and across water column ecosystem dynamics research. They include cultural issues (relating to the sharing of scientific data, and communicating science to the general public) in addition to aspects of data management, such as discovery, access, storage, sharing, data heterogeneity, and interoperability. Other data-related concerns focused on quality control and assurance, the availability of tools for automation, integration and visualization of data, education to increase data and computational literature, and insufficient funding for interdisciplinary science and international collaboration. To address the scientific problems and challenges, participants identified tools, enhancements to current databases, and required global support structures.

Use cases that defined a scientific question and the requirements for addressing it were used as a vehicle to identify tangible actions by EarthCube that would facilitate the research required. Five use cases were identified by the participants and developed on the second day of the workshop. Topics included: subpolar North Atlantic plankton dynamics; revisiting anthropogenic ocean carbon data/model intercomparisons; impacts of sea ice reduction; coastal ecosystem responses to climate change; and changes in coastal plankton communities over decades. Use cases gave rise to discussions on community next steps. Summarized below is a series of short and longer-term next steps, designed to provide EarthCube with tangible actions that would facilitate the community achieving its science:

Short term:

- Catalog of different data repositories and tools.
- Catalog of existing and dark tools[1] for discovering, analyzing, and visualizing data.
- Better search tools to discover data.
- Data and computational literacy curricula.
- To aid data submission, create a “help desk” to guide investigators to appropriate repositories.

Longer-term:

- Making databases and repositories interoperable.
- Develop a plan to identify and liberate dark data[2].
- Enhancing access to international databases.
- Centralized access to earth system and ocean model output data.

---

[1]For the purposes of this workshop, a dark tool was defined as a tool, not publicly discoverable or accessible.

[2]For the purposes of this workshop, dark data were defined as extant data, not publicly discoverable or accessible.

INTRODUCTION

EarthCube is an initiative to advance research in Earth sciences, such as ocean ecosystem dynamics, through the creation of a robust cyberinfrastructure enabling cross-disciplinary science. It is a collaboration between the National Science Foundation and Earth, atmosphere, ocean, computer, information, and social scientists, educators, data managers, and more. EarthCube's goal is to create a well-connected, transparent system for sharing data and knowledge that is both derived and governed by the community. It will connect the many facets of data and information management, from data and metadata discovery and access, to data curation and management, through interface building that enables analysis and knowledge creation.

The NSF has adopted a novel approach to developing this infrastructure as an emergent community process. For the last two years, the Earth Science community has been engaged in a series of workshops, charrettes, white paper and roadmap writing, and project development to explore and elucidate the EarthCube concept. A series of domain workshops have been held since October 2012 bringing together stakeholders from various Earth Science communities to define what EarthCube could do to increase their scientific productivity.

Ocean Ecosystem Dynamics encompasses a broad array of disciplines, which seeks to increase our understanding of the interplay between biological, chemical, and physical processes in the ocean. It is fundamentally an interdisciplinary science, incorporating the oceanographic sciences with atmospheric and geologic sciences. By its nature, this field produces highly diverse data types that pose unique challenges for management, integration, and analysis. The ability to discover, access and synthesize high quality data from various disciplines is crucial to ocean ecosystem sciences.

As part of this effort, an EarthCube Water Column Domain End-User Workshop (NSF award 1338892) hosted by the Biological and Chemical Oceanographic Data Management Office (BCO-DMO) was held October 7-8, 2013 at Woods Hole Oceanographic Institution. The workshop goals were to create a set of data and cyber needs and challenges to inform the EarthCube effort, and to educate and engage scientists, data managers, and other members of this community to EarthCube's vision and goals. Specific objectives included:

1. Establishing and prioritizing science drivers and challenges facing water column ecosystem research, as well as those related to interdisciplinary research across other domains;
2. Identifying barriers preventing efficient use of data and tools, both within and external to the domain;
3. Identifying existing and critically needed data, products, and tools essential to forming a robust cyberinfrastructure for the community;

4. Developing use cases that demonstrate how the cyberinfrastructure will be used to address science drivers or provide critical resources enabling education and research; and
5. Generating a list of next steps to include short and long-term actionable items that would begin to fill the needs identified in the workshop.

A planning committee was created, consisting of the workshop project Co-PIs and six individuals possessing expertise in ocean ecosystem research and data management (Appendix B). Together, the team developed details of the workshop logistics, created a list of invitees, invited people with specific expertise to provide talks relevant to the workshop, created and distributed the workshop agenda (Appendix A.), distributed workshop-related tasks, and facilitated the event. The group leveraged the work and results of previous workshops by providing background information for workshop participants to review prior to the event.

The workshop included time split among plenary and breakout sessions, with time reserved for summary sessions each day. Plenary sessions began with invited talks that provided context for the breakout sessions. For one session following the plenary talk, a panel of representatives from data related infrastructures and repositories gave short presentations about their respective centers. The speakers, panelists, and their titles are included in Appendix B. An additional half-day was reserved for report preparation by the planning committee.

Results of this workshop will be synthesized with the output from other workshops, and distilled into a unified set of requirements, to be used by the NSF to help shape the development of EarthCube.

The workshop included 50 participants (Appendix C.) in the domain of oceanic ecosystem dynamics (established and early career researchers, teaching faculty, graduate students, postdocs, data and information managers and cyber-related researchers) to explore and document the community's cyberinfrastructure needs from the users' viewpoint. The participants self-identified into one or more interest areas: in biological, chemical, and physical oceanography (68%), data and information managers (42%), computer science researchers (22%), and education and social science specialists (14%), and modelers (8%).

This report summarizes the activities of the workshop and the outcomes of plenary and breakout sessions. An executive summary was generated and distributed to participants and NSF EarthCube program managers.

SCIENCE ISSUES AND CHALLENGES

On the first day, the morning plenary and breakout session focused on important science drivers and major challenges. In the afternoon, plenary and breakout sessions focused on barriers to research, and opportunities offered by future cyberinfrastructure developments. The results of these discussions are the following:

1. Important science drivers and challenges: Participants identified several high-priority science questions that will be the focus of interdisciplinary efforts during the next 5-15 years.

- How do ocean acidification, warming, and hypoxia affect ocean ecosystems? How do these affect the organisms? What species are most impacted by lowering pH and increasing CO_2, and what are the impacts on ecosystem structure and dynamics?
- As ocean circulation changes as the oceans warm and salinity varies, how will marine organisms respond to the changes? What will be the impacts on productivity, species distributions, and carbon flux in the ocean? How will such ecosystem changes affect human populations?
- How do changes in ice cover in the Arctic Ocean impact other regions and give rise to changing weather patterns, which in turn have bottom up effects on marine coastal and open ocean ecosystems? How are species distributions, life histories, and species interactions impacted by the changing conditions in the Arctic?
- How do different physical, chemical, and biological processes come together to create more complex emergent properties in ocean ecosystems? What are the feedback loops among these processes and how do they give rise to biological, chemical, and genetic diversity in the oceans?
- How are anthropogenic impacts other than climate change, such as eutrophication, overfishing, and coastal development influencing marine ecosystems? How does this influence our approach to ecosystem based management and conservation?

2. Current challenges to high-impact, interdisciplinary science:

A. During the discussion on science drivers, several themes emerged as consistent challenges faced within/across the involved discipline(s):

Cultural (Science)	Lack of a common data sharing expectation; lack of confidence that data will be cited.
Cultural (Public)	Lack of common language to communicate science to the public.
Databases	Difficult to keep on top of all current cyberinfrastructure efforts (DataONE, EarthCube, Data Conservancy); lack of interoperability and diversity of data structures among data centers; lack of a common database capable of storing, searching, and serving terabytes of images and videos; submission of the same data to multiple repositories;

	difficulty knowing where to submit data.
Access	Lacking clearinghouses for (a) software/tools/code (b) databases (c) models, and (d) vocabularies/ontologies; difficult to find, access, and work with data outside of your area of expertise; difficult to access data from international databases; difficult if not impossible to access data from publications.
Dark data	No access to dark data (data on local computers or in file cabinets that are not in any public repository and not discoverable); challenges include lack of metadata, changing personnel, and lack of time and funding.
Data types	Some data are difficult to put into repositories (due to size, structure, format).
Metadata	Lack of rich, standardized metadata.
Vocabulary	Varying vocabularies within and between different domains.
Quality	No standardized way to assess data quality; inadequate quality control assessments.
Tools	Lacking tools for automation, integration of a variety of data, and visualization in 4-D; a challenge to combine different data types and to deal with new and evolving data types; difficult to analyze heterogeneous and gappy data; a lack of systems for documenting provenance.
Hardware	Increasing need for data storage, processing capabilities, and bandwidth.
Education	Lack of data and computational literacy; need for data literacy courses; lack of online training tools for discovering, accessing, and using data.
Effort	Substantial time and cost in the effort of making data available.
Agencies	Insufficient funding for cutting edge science; current funding and proposal review structures hinder cross-discipline and international collaboration**;** lack of inter-agency coordination.

B. Continuing on day one, the second breakout session discussed the current barriers to research. These included some of the challenges noted above. The participants recognized three different categories of barriers: cultural, institutional, and technical.

Cultural Barriers:

1. Lack of data and computational literacy, training/tutorials, etc.
2. Accessing dark data (also recognized as a technical barrier).
3. Reluctance to share data; such behavior might be changed by providing ‘carrots’, such as persistent identifiers for data recognition.
4. Differences between domains such as vocabularies and data sharing expectations.

5. Lack of ability to communicate to the public. Lack of communicator / technical solutions that translate science to lay people (also recognized as a technical barrier with regard to vocabularies for communicating to the public).
6. Fear of 'showing the warts' on the data. Need mechanism to 'encourage' submitters to include metadata on sensor calibration (traceability), known issues with data, etc. that will improve data quality.

Institutional Barriers:

1. NSF funding stovepipes hinder cross-discipline collaboration.
2. Lack of Intra/Inter-agency collaboration on data management (NOAA, NSF, NASA, etc.) (e.g., within NSF: DataONE, EarthCube, Data Conservancy, IEDA, BCO-DMO).
3. Lack of support for International collaboration - little incentive for international collaboration. There are institutional barriers to international collaboration. The barriers exist in both directions (e.g., U.S. investigators can not get money from E.U. collaborative projects and vice versa and there are no good mechanisms to support jointly funded international collaborations).
4. Interdisciplinary proposals are difficult for reviewers and difficult to get funded.
5. Lack of adaptive funding systems that are compatible with current research/needs (temporally responsive and sustainable in long term).
6. Tenure/promotion committees do not reward effort for data organization and submission to databases.
7. Lack of support for routine monitoring and observations.

Technical Barriers:

1. Lack of one resource for discovery, access, analysis scripts/code/tools for automation and visualization.
2. Difficulty working with very large datasets ('big data').
3. Lack of rich metadata and provenance.
4. Lack of quality control assessments. Existing systems do not meet all needs.
5. Difficulty with combining heterogeneous data and data types from different streams (format, scale, vocabularies, and standards).
6. Difficulty dealing with evolving data types: E.g. "Omics" data and data from new/improved instruments.
7. Lack of seductive technological solutions that are so effective you want to use them to share data, etc. (no need for contrived "carrots" or "sticks"; e.g. Google, smartphones, etc.).

TECHNICAL INFORMATION/ISSUES/CHALLENGES

On day two, a plenary talk was given describing the evolution of hardware and software in support of data management over the last fifteen years. This talk provided the context for thinking about data resources, products and tools needed for the future. Following the talk were presentations by panelists representing data related infrastructures and repositories. The breakout groups identified 63 sources of known and readily used online data resources and tools (appendix D.) A Google Earth file containing the location of these data resources was created shortly after the workshop by a participant and is currently hosted by UCSD (http://hdo.ucsd.edu/explore_kml/porthub). This website provides useful information for all oceanographic researchers and is a resource EarthCube could help promote and advertise. The groups then identified the following list of critically needed tools, repositories and support structures. The extracted notes from these sessions giving rise to this summary are provided in Appendix E.

1. Tools:

- Visualization tools for interactive data analysis.
- Tools to track and control data versions.
- Tools to foster data quality assurance and quality control.
- Tools to enhance data discovery and searches.
- A community-level interface and facility to share tools and programming code.
- Technology to assimilate metadata produced by smart sensors.
- A tool to translate from different format types to a standard format.

2. Repositories and Databases:

- Data repositories are needed for specific subdomains such as organic chemistry and metabolomics.
- A sharable system for handling massive amounts of data including images and video.
- Production of a wider range of more sophisticated data products and derived calculations.
- Automatic incorporation of new data into databases/repositories.

3. Global Support Structure:

- Guidance on where to submit data including any restrictions and guidance for repository use.
- Centralized forum providing information about models, scripts, software, and documentation.
- Graduate-level curricula for training the next-generation of scientists to be able to find and work with data.
- Educate programmers to understand science.
- Standard interdisciplinary metadata format.
- Cross-domain ontologies of measurable phenomena and instrument types.

USE CASES IDENTIFIED BY PARTICIPANTS

Five "Use Cases" were identified by the participants and developed as examples during a breakout session on the second day of the workshop. The title, scientific problem, and tangible actions by EarthCube that would facilitate the goal are summarized below.

1. Subpolar North Atlantic Plankton Dynamics

Scientific Problem: What are the plankton dynamics from the surface to 1000 meters in the subpolar North Atlantic during periods when it is difficult to sample, particularly during the winter and the transition to spring? Specifically, what is the magnitude of primary productivity, when do deep over-wintering zooplankton come to the surface, how do zooplankton control termination of the spring bloom, and how will climate change impact transfer of carbon to higher trophic levels?

Tangible actions by EarthCube that would facilitate the goal:

- Develop a framework to integrate diverse interdisciplinary data in multiple formats with differing sampling rates and varying spatial coverage.
- Provide access to improved systems for data entry/ingestion and QA/QC from sensor systems and near real-time analysis, including calibration records and proxy details.
- Facilitate the merging and visualization of Eulerian and Lagrangian data in an efficient, synoptic, and synthetic way.
- Provide visualization tools for data exploration using broad array of data sets and portal for discovery by others.
- Facilitate an archive for large data sets of annotated images and broadband acoustics.

2. Revisiting anthropogenic ocean carbon data/model intercomparisons

Scientific Problem: Is the integrated air-sea CO_2 flux inferred from surface ocean pCO_2 measurements consistent with the global anthropogenic CO_2 inventory and distribution; why are the current coupled models so different in their predictions for both in the Southern Ocean (defined as south of 35S)?

Tangible actions by EarthCube that would facilitate the goal:

- Create tools for re-gridding spatial data from different sources (e.g., model, *in-situ*, satellite).
- Develop new tools for developing data/model comparisons with uncertainty budgets.
- Provide easily updated mapping tools for creating World Ocean Atlas-like inventories of oceanic carbon, nutrients, satellite sea surface temperature, etc.

3. Impacts of sea ice reduction

Scientific Problem: How would the sea ice loss in the Arctic change physical mixing, biological productivity, ocean chemistry, and flow of carbon through the food webs and to depth?

Tangible actions from EarthCube that would facilitate the goal:

- Facilitate data management and informatics consulting throughout projects to facilitate integration of disparate data types.

- Provide access to technology that supports collaboration across disciplines and subgroups of collaborators throughout the project.

4. Coastal Ecosystem Responses to Climate Change

Scientific Problem: How will coastal ecosystems respond to anthropogenic stressors, especially increasing atmospheric carbon dioxide levels, and its related impacts (e.g. ocean acidification, warming, hypoxia, sea-level rise, and storm frequency/intensity)? How will environmental variability driven by the biology influence the interactions between these multiple environmental stressors, and how will the biological/chemical feedbacks change in the future?

Tangible actions from EarthCube that would facilitate the goal:

- Develop better tools for discovering information relevant to determining how coastal ecosystems respond to climate change.
- Create a community portal to data repositories that incorporate both new and historical data that can be used for improving spatially explicit models.
- Develop tools for integrating diverse types of data for coastal ecosystems.
- Develop collaborative website for data sharing and model development during the project.
- Create a robust integrated information resource that will be useful for coastal planning by resource managers.

5. How have coastal plankton communities changed over decades?

Scientific Problem: How have the structure and function of coastal plankton communities changed over recent decades? What are the effects that propagate through the food web to influence fish, mammals, and seabirds? How have the occurrence and distribution of HAB events changed and what are the contributing factors?

Tangible actions from EarthCube that would facilitate the goal:

- Provide methodologies of integrating data with varying resolution (temporal and spatial).
- Develop a data portal to find, discover, and integrate datasets and that also serves as a collaborative workspace.
- Develop automated methods of mining and integrating data from the literature (dark data).
- Develop systematic approaches for dealing with biological taxonomic information in mixed data sets, taking into account dynamic taxonomic standards.

COMMUNITY NEXT STEPS

Below is a list of tangible items and actions by EarthCube that would facilitate the community achieving its transformative science goals:

Short- term Next Steps (1-3 years):

- Create a catalog of different data repositories and tools.
- Create a catalog of existing and dark tools (for discovering, analyzing, and visualizing data).
- Develop a tool that captures the output from multiple Earth System models for a geographical position for comparison to field-collected data.
- Create better search tools for data discovery (such as faceted searches).
- Provide a help desk to provide investigators with information on repositories where data should be submitted.
- Provide incentives to preserve all data and make accessible.
- Facilitate data and computational literacy curriculum development, including tutorials for undergraduate and graduates.
- Synthesize outcomes from the domain-specific EarthCube workshops, and communicate outcomes to EC stakeholders.
- Enable the funding of international collaborations and access to international databases.

Long-term Next Steps (>3 years):

- Create a centralized access to earth system and ocean model output data with Google style searchability.
- Make databases and repositories interoperable.
- Plan to identify and liberate dark data (resources, funding, and expertise).
- Provide funds to enable Use Cases put forward in the workshop to be implemented.

COMMUNITY RECCOMENDATIONS

Throughout the workshop discussions on science drivers and current challenges, there were many references to education and data availability that cut across several agenda topics. These two themes however may not be explicitly stated or emphasized elsewhere.

Education and the need for data management and computational literacy curricula were highlighted during nearly all the workshop sessions. The workshop participant consensus is that this type of education is critically needed, not only by undergraduates and graduates, but also mid-career and established researchers, who may have had little exposure to, or guidance on management. Beginning education as early as possible with reinforcement throughout the learning process will enable future generations of researchers to effectively manage and share their scientific data. Education curricula should include the use of technologies that enable better management, visualization, analysis and sharing of data.

The critical need for more data was also pervasive throughout the workshop. Most participants were in agreement over the paucity of marine ecosystem data available to conduct research. Discussions focused on two possible causes, (1) significant pools of extant data are dark or inaccessible, and (2) there is simply a lack of comprehensive observations necessary for conducting analyses to answer questions, such as those posed under Science Issues and Challenges. Dr. Ray Schmitt highlighted this lack of observations during the workshop's opening plenary talk[3], stressing the need for sustained, uninterrupted time series of at least 30 years to adequately address climate variability and change. The workshop participants felt continued funding for existing observational programs and funding for new programs to collect long-term ocean ecosystem data should remain a priority in addition to cyberinfrastructure development. Similarly, funding vehicles for data rescue (of dark data), and analysis and synthesis of extant data will continue to be needed.

[3]Relevance of Observations to Addressing Challenges in Oceanography/Climate Research. Delivered October 2013 at the EarthCube Domain Workshop for Ocean Ecosystem Dynamics. https://docs.google.com/presentation/d/1TL258F-ty8iGDaZVQnNGSWBtwfsP1v3vY0FeJKd1JHg/edit?usp=sharing

Appendix A. Final Agenda

EarthCube End-User Workshop for Ocean Ecosystem Dynamics
Agenda

October 7-8, 2013

Woods Hole Oceanographic Institution
Quissett Campus, Clark Building

Pre-Meeting: Sunday, October 6

1800-2000 *Refreshments in Clark 5-07*

Day 1: Monday, October 7

0830-0900 *Continental Breakfast*
0900-0940 *Welcome and Introductions*:
- Welcome to the Workshop - Danie Kinkade
- Welcome to WHOI- Susan Avery (WHOI Director)
- Logistics and Housekeeping - Danie Kinkade
- ~~Foundations of EarthCube - Eva Zanzerkia (NSF)~~ CANCELED
- Workshop Goals – Why We're Here - Jamie Pierson (UMCES)

0940-1040 *Plenary* – Invited Speakers:
- Ray Schmitt (WHOI), Relevance of observations to addressing challenges in Oceanography/climate research
- Joanie Kleypas (NCAR), Challenges to conducting interdisciplinary research

1040-1100 *Break*

1100-1105 *Charge to breakout groups:*
- Logistics, map/directions
- Each group has a leader, recorder, and facilitator designated, plus tele-comm person for groups with remote attendees
- Google doc provided for each group output (EarthCube templates)

1115-1200 *Breakout 1*: Science drivers and grand science challenges now and in 15 years

1210-1300 *Plenary*- Group Reports (3 min each) and Consolidate Results from Breakout 1

1300-1400 *Lunch*

1400 - 1430 *Plenary*- Invited speaker:

- Pascal Hitzler (Wright State University), Cyberinfrastructure Opportunities for Geoscience Research

1440 - 1525 <u>*Breakout 2*</u>: Current barriers to research

1530 - 1550 *Break and Group Photo*

1600 - 1650 *Plenary*- Group Reports (3 min each) and Consolidate Results from Breakout 2

1650 - 1715 *Closing comments and Discussion*

[Planning Committee brief meeting to make adjustments to Tuesday's agenda.]

1800 *Social Time*

1830 *Catered Dinner* - Clark 5-07

1930 ~~Evening Speaker- Joel Cutcher-Gershenfeld (University of Illinois), Ocean Ecosystem Research Community Cyberinfrastructure Needs, Survey Results~~ CANCELED

<u>Day 2: Tuesday, October 8</u>

0800 - 0830 *Continental Breakfast*

0830 - 0900 *Recap Day 1*, Day 2 Overview

0900 - 0920 *Plenary*- Invited speaker:

- Todd O'Brien, Current data management landscape within water column research

0920 - 1040 *Plenary* – Panel (Demos ~ 8 min): Examples of Current Data-Related Infrastructure and Challenges

- Todd O'Brien
- Krisa Arzayus
- Steve Diggs (CCHDO),
- Emily Law (JPL)

- Vembu Subramanian and Ellen Tyler (IOOS)
- Bob Groman (BCO-DMO)

NEW - Cyndy Chandler presenting a subset of Joel Cutcher-Gershenfeld et al. slides on survey results

1040 - 1055	*Break*
1105 - 1145	<u>*Breakout 3*</u>: Identify existing and critically needed data resources, products, and tools
1155 - 1245	*Plenary* - Group Reports (3 min each) and discussion
1245 - 1345	*Lunch*

1345 - 1450	*Plenary* - Cyndy Chandler/Heidi Sosik (WHOI), Use Cases Defined (5 min)
	• Components to capture in use case:
	  o What things (data, tools, analyses, etc.) are involved
	  o Who are the players
	  o What is the product/outcome
	  o What is the workflow
	• Heidi Sosik (WHOI), Use Cases from a Science Perspective
	• Plenary - Discussion of Use Case topics for breakout groups.

1455 - 1555	<u>*Breakout 4*</u>: Brainstorm science use cases and solutions/tangible actions from EarthCube that would facilitate the research
1555 - 1605	*Break*
1605 - 1630	*Plenary*- Group Reports on use case development
1630 - 1700	*Plenary*- Discussion and Wrap-up Discussion of Community next steps: How to maintain the community's focus on data and cyber infrastructure needs? How can the community help facilitate filling these needs? What are the immediate tangible steps (short/long term) that EC could take to fill identified needs?
1700	*Adjourn*

[Planning Committee meets to review workshop results.]

Day 3 Planning Committee and available workshop participants

0800 - 0830	*Breakfast*
0830 - 1200	*Planning Committee Review of workshop output* • Assign and Write Sections • Reconvene to complete report draft
1200	*Adjourn*

Appendix B. Planning Committee, and Invited Speakers and Panelists

Committee Member Name	Title	Institution
Danie Kinkade	Information Systems Associate	BCO-DMO, Woods Hole Oceanographic Institution
Cynthia L Chandler	Information Systems Associate	BCO-DMO, Woods Hole Oceanographic Institution
Robert C Groman	Information Systems Specialist	Woods Hole Oceanographic Institution
David M Glover	Sr. Research Specialist	Woods Hole Oceanographic Institution
Jon Hare	Oceanographer, and Chief, Oceanography Branch	NOAA/NMFS/NEFSC
David I Kline	Project Scientist	Scripps Institution of Oceanography, University of California, San Diego
Jasmine S Nahorniak	Sr. Faculty Research Assistant	Oregon State University
Todd D O'Brien	Oceanographer, Project Lead	NOAA / NMFS / COPEPOD
Mary Jane Perry	Professor	University of Maine
James J Pierson	Research Assistant Professor	University of Maryland Center for Environmental Science
Peter H Wiebe	Scientist Emeritus	WHOI
Speaker/Panelist Name	**Title**	**Institution**
Krisa M Arzayus	Oceanographer and Chief, Marine Data Stewardship Division	NOAA/National Oceanographic Data Center
Stephen C Diggs	Data Manager, CCHDO	Scripps Institution of Oceanography
Robert C Groman	Information Systems Specialist	Woods Hole Oceanographic Institution
Pascal Hitzler	Associate Professor, Dept. of Computer Science and Engineering	Wright State University
Joan Kleypas	Scientist, Marine Ecologist/Geologist	National Center for Atmospheric Research
Emily Law	Science Data Systems Manager	Jet Propulsion Laboratory
Todd D O'Brien	Oceanographer, Project Lead	NOAA / NMFS / COPEPOD
Ray Schmitt	Sr. Scientist, Physical Oceanography	WHOI
Heidi M Sosik	Sr. Scientist, Biology	Woods Hole Oceanographic Institution

Appendix C. Participant List

Name	Organization	Email
Robert Arko	Lamont-Doherty Earth Observatory	arko@ldeo.columbia.edu
Krisa M Arzayus	NOAA/National Oceanographic Data Center	krisa.arzayus@noaa.gov
Stace Beaulieu	Woods Hole Oceanographic Institution	stace@whoi.edu
Kim S Bernard	College of Earth, Ocean, and Atmospheric Sciences, Oregon State University	kbernard@coas.oregonstate.edu
Cynthia L Chandler	BCO-DMO, Woods Hole Oceanographic Institution	cchandler@whoi.edu
Hilary G Close	University of Hawaii at Manoa	hclose@hawaii.edu
James E Conners	Scripps Institution of Oceanography (CCE LTER / PAL LTER)	jconners@ucsd.edu
Stephen C Diggs	Scripps Institution of Oceanography	sdiggs@ucsd.edu
Robert B Dunbar	Stanford University	dunbar@stanford.edu
David M Glover	Woods Hole Oceanographic Institution	dglover@whoi.edu
Robert C Groman	Woods Hole Oceanographic Institution	rgroman@whoi.edu
Pascal Hitzler	Wright State University	pascal.hitzler@wright.edu
Lei Hu	Dauphin Island Sea Lab	lhu@disl.org
Dana E Hunt	Duke University	dana.hunt@duke.edu
Robert M Key	Princeton University	key@princeton.edu
Danie Kinkade	BCO-DMO, Woods Hole Oceanographic Institution	dkinkade@whoi.edu
Joan Kleypas	National Center for Atmospheric Research	kleypas@ucar.edu
David I Kline	Scripps Institution of Oceanography, University of California, San Diego	dkline@ucsd.edu
Michael R Landry	Scripps Institution of Oceanography	mlandry@ucsd.edu
Emily Law	Jet Propulsion Laboratory	emily.law@jpl.nasa.gov
Michael W Lomas	Bigelow Laboratory for Ocean Sciences	mlomas@bigelow.org

Krista Longnecker	Woods Hole Oceanographic Institution	klongnecker@whoi.edu
Amy E Maas	Woods Hole Oceanographic Institution	amaas@whoi.edu
Andrew R Maffei	Woods Hole Oceanographic Institution	Amaffei@whoi.edu
Matthew Mazloff	SIO-UCSD	mmazloff@ucsd.edu
Sabine Mecking	Applied Physics Laboratory/University of Washington	smecking@apl.washington.edu
Vanessa K Michelou	University of Hawaii- Center for Microbial Oceanography	michelou@hawaii.edu
Jasmine S Nahorniak	Oregon State University	jasmine@coas.oregonstate.edu
Thomas W. Narock	University of Maryland, Baltimore County	Thomas.W.Narock@nasa.gov
Todd D O'Brien	NOAA / NMFS / COPEPOD	Todd.OBrien@noaa.gov
Cindy Palinkas	University of Maryland Center for Environmental Science	cpalinkas@umces.edu
Deana D Pennington	University of Texas at El Paso	ddpennington@utep.edu
Mary Jane Perry	University of Maine	perrymj@maine.edu
James J Pierson	University of Maryland Center for Environmental Science	jpierson@umces.edu
Lisa Raymond	Woods Hole Oceanographic Institution	lraymond@whoi.edu
Tatiana Rynearson	University of Rhode Island, Graduate School of Oceanography	rynearson@mail.uri.edu
Mark Schildhauer	NCEAS/UCSB	schild@nceas.ucsb.edu
Ray Schmitt	WHOI	rschmitt@whoi.edu
Stephanie Sharuga	Louisiana State University, Department of Oceanography & Coastal Sciences	ssharu1@lsu.edu
James D Simons	Center for Coastal Studies, Texas A&M University-CC	james.simons@tamucc.edu
Heidi M Sosik	Woods Hole Oceanographic Institution	hsosik@whoi.edu
Karen Stocks	Scripps Institution of Oceanography	kstocks@ucsd.edu
Vembu Subramanian	SECOORA	vembu@secoora.org
Ellen P Tyler	Alaska Ocean Observing System	tyler@aoos.org
Mimi W Tzeng	Dauphin Island Sea Lab	mtzeng@disl.org
Amy Waterhouse	Scripps Institution of Oceanography	awaterhouse@ucsd.edu
Amanda L Whitmire	Oregon State University	amanda.whitmire@oregonstate.edu
Peter H Wiebe	WHOI	pwiebe@whoi.edu

Jerry Wiggert	University of Southern Mississippi	jerry.wiggert@usm.edu
Clifton B Woodson	University of Georgia	bwoodson@uga.edu
Erik Zettler	Sea Education Association	ezettler@sea.edu

Appendix D. Databases, data centers, tools and other resources

Databases, Data Centers, and Tools, generated during the pre-workshop survey and breakout session three:

- ACADIS (Advanced Cooperative Arctic Data and Information Service) http://nsidc.org/acadis
- ADD (Antarctic Digital Database) http://www.add.scar.org/
- AGDC (Antarctic Glaciological Data Center) http://nsidc.org/agdc
- AVISO (Archiving, Validation and Interpretation of Satellite Oceanographic Data) http://www.aviso.oceanobs.com
- BATS (Bermuda Atlantic Time-series Study) http://bats.bios.edu/
- BCO-DMO (Biological Chemical Oceanography Data Management Office) http://www.bco-dmo.org/
- CAMERA (Community for Advanced Microbial Ecology Research and Analysis) http://camera.calit2.net
- CCE-LTER (California Current Ecosystem - Long Term Ecosystem Research) http://cce.lternet.edu/
- CCHDO (CLIVAR & Carbon Hydrographic Data Office) http://cchdo.ucsd.edu
- CDIAC (Carbon Dioxide Information Analysis Center/) http://cdiac.ornl.gov
- CLIVAR (Climate Variability and Predictability) htttp://clivar.org
- COPEPOD (Coastal & Oceanic Plankton Ecology, Production, & Observation Database) http://www.st.nmfs.noaa.gov/plankton/
- COSEE (Centers for Ocean Sciences Education Excellence) http://www.cosee.net
- DataONE (Data Observation Network for Earth) http://www.dataone.org/
- Data.gov (Federal data gateway / integrated catalog) http://www.data.gov
- Datazoo (at the University of California at San Diego) http://oceaninformatics.ucsd.edu/datazoo/
- Dryad (Dryad Digital Repository) http://datadryad.org/
- EPA repositories (Environmental Protection Agency) http://www.epa.gov/datafinde
- ESR (Earth and Space Research Antarctic Tide Gauge Database) http://www.esr.org/antarctic_tg_index.html
- Exxon Valdez database http://www.afsc.noaa.gov/ABL/Habitat/ablhab_exxonvaldez_hydrocarbon_databas htm
- FishBase http://www.fishbase.org/
- GADR (Global Argo Data Repository) http://www.nodc.noaa.gov/argo/
- GBIF (Global Biodiversity Information Facility) http://gbif.org
- GenBank http://www.ncbi.nlm.nih.gov/genbank/
- Google Earth/Ocean
- HOT-DOGS (Hawaii Ocean Time-series Data Organization & Graphical System) http://hahana.soest.hawaii.edu/hot/hot-dogs/
- HURL (Hawaii Undersea Research Lab) http://www.soest.hawaii.edu/HURL/
- InterRidge Vents Database www.interridge.org/IRvents
- IEDA (Integrated Earth Data Applications) http://www.iedadata.org/
- IOOS (including Regionals IOOS-RA)

- JPL-PODAAC (Jet Propulsion Laboratory - Physical Oceanography Distributed Active Archive Center) http://podaac.jpl.nasa.gov/
- LTER Network (Long Term Ecosystem Research Network) http://www.lternet.edu
- MBARI (Monterey Bay Aquarium Research Institute) http://www.mbari.org/data/
- MCVO (Martha's Vineyard Coastal Observatory) http://www.whoi.edu/mvco
- MGDS (Marine Geoscience Data System) http://www.marine-geo.org
- MMI (Marine Metadata Interoperability) http://marinemetadata.org
- NASA DAACs (SeaWiFS/MODIS, Meris, VIIRS) http://nssdc.gsfc.nasa.gov/earth/daacs.html
- NCBI GenBank (National Center for Biotechnology Information) http://www.ncbi.nlm.nih.gov/genbank/
- NODC (National Oceanographic Data Center) http://nodc.noaa.gov
- NCDC (National Climate Data Center) http://ncdc.noaa.gov
- NGDC (National Geophysical Data Center) http://ngdc.noaa.gov
- NOS (National Ocean Service) http://oceanservice.noaa.gov
- NBDC (National Buoy Data Center) http://ndbc.noaa.gov
- NSDL (National Science Digital Library) http://NSDL.org
- Ocean Teacher (an OpenCoursewareSite) www.oceanteacher.org
- ODV (OceanDataView) http://odv.awi.de
- OBIS (Ocean Biogeographic Information System) http://www.iobis.org
- PANGAEA (earth and environment database) http://www.pangaea.de/
- Phylo MetaRep (Metagenomics Reports) http://www.jcvi.org/phylo-metarep/
- ProPortal (*Prochlorococcus* Portal) http://proportal.mit.edu/
- R2R (Rolling Deck to Repository) www.rvdata.us/
- ReefBase (a global information system for coral reefs) www.reefbase.org
- SAHFOS Continuous Plankton Recorder (Sir Alister Hardy Foundation for Ocean Science) http://sahfos.org
- SeaBASS (bio-optical oceanography database) http://seabass.gsfc.nasa.gov/
- SeaDataNet (pan-european infrastructure for ocean & marine data management) http://www.seadatanet.org/
- SeaGrant (NOAA's National Sea Grant College Program) www.seagrant.noaa.gov
- SeaVOx (SeaDataNet's MarineXML) https://www.bodc.ac.uk/data/codes_and_formats/seavox
- USGS Streamflow http://water.usgs.gov/nsip/
- WOCE (World Ocean Circulation Experiment) http://woce.nodc.noaa.gov/wdiu
- WOOD (World-wide Ocean Optics Database) http://wood.jhuapl.edu/wood/

Programming/Code Resources:

- GitHub (Online Version Control Service)
- SVN (Apache Subversion)
- SourceForge (Web-based open-source code repository)

Appendix E. Critically needed data products and tools

We need a centralized searching mechanism to discover [all] digitally accessible data, as well as tools and models. This mechanism needs to have the ability to subset data that are co-located in space and/or time, (e.g., a GoogleEarth type search). The implementation should support advanced queries (e.g., "find all North Atlantic plankton data that were collected below 1000 meters").

We need better visualization tools for interactive data analysis, including non-gridded data. These visualization tools should be able to accurately integrate data from multiple sources, taking into account differences in units, collection methodology, and processing methodology.

We need a tool to help track and control data versions and to prevent the finding of duplicate datasets during a search (perhaps by using persistent identifiers).

We need a community-level interface and facility to share tools and programming code to enable reuse of other peoples' efforts. Documentation for using the tools and code should also be included.

We need a sharable system for handling massive amounts of data including images and video.

Data providers need guidance on where to submit data including any restrictions, guidance for repository use, and details of long-term storage time limits.

We need centralized forum providing information about models, scripts, software, and documentation for "how-to" methods in data handling and related topics that are located in web accessible repositories.

We need production of a wider range of more sophisticated data products and derived calculations accompanied by quality control notes.

Development of graduate-level curricula for training the next-generation of scientists to be able to find and work with data. Educate programmers to understand science. Development of basic and advanced online training videos.

Data repositories are needed for organic chemistry and metabolomics.

Automatic incorporation of new data into databases/repositories.

Develop the technology to assimilate metadata produced by smart sensors.

Creation of standard interdisciplinary metadata format.

Create a data Rosetta Stone that can translate from different format types to a standard format (and vice versa) that will also support all available ontologies to enable mapping and matching of variables/field names.

Foster the creation of cross-domain ontologies of measurable phenomena and instrument types (recognizing the work ongoing in organizations such as the BODC).

Develop tools to help foster data quality assurance and quality control.

Appendix F. Fully Developed Use Cases

Use Case 1: Subpolar North Atlantic Plankton Dynamics

Group Members:

Cyndy Chandler,
Peter Wiebe,
Bob Arko,
Mary Jane Perry,
Jamie Pierson,
Tatiana Rynearson,
Jasmine Nahorniak,
Amanda Whitmire,
Pascal Hitzler

What is the scientific problem?

What's going on in the subpolar North Atlantic (surface to 1000 m), where are the zooplankton, when do they come out of diapause? What are the plankton dynamics from the surface to 1000 meters? How do we figure out what's going on in the planktonic food web in the subpolar North Atlantic when we're not there (episodic events, winter months)?

What is the goal (how will you address that problem)?

Who is there (taxonomy; genomics) and what are they doing (metabolomics) at high resolution over the course of a year (or even winter - spring transition).
Quantify trophic dynamics (zooplankton provide food for fish stocks - will help in prediction of stock size in the future).
What is the response to climate change?
Also consider competing hypotheses, e.g. Sverdrup spring vs. Behrenfeld winter bloom dynamics.

Who are the people and what roles do they fill?

Field biologists (phytoplankton and zooplankton)
Plankton taxonomists
Physical oceanographers
Molecular ecologists
Biogeochemists
Modeler
Remote sensing specialist
Field technicians / engineers to build and maintain equipment
Platform developers
Instrument developers
Data manager
International collaborators (foreign clearances?)

What things are needed (data, tools, etc.) to accomplish the goal?
Variables:
Location (GPS)
Temperature
Salinity
Nutrients
Primary production
Species composition
Abundance and size distribution of phytoplankton, zooplankton and zooplankton predators

Sensors/models:
Models: 3-D basin scale model ecosystem model, acoustic models for zooplankton analysis,
Platforms: satellites, ships, moorings, autonomous vehicles
Satellite data - ocean color, sea surface temperature, sea surface salinity, wind, altimetry
Continuous underway data from ships of opportunity (surface temperature, salinity, nutrients, chlorophyll, phytoplankton imaging, plankton recorder, acoustics for zooplankton)
Ships for platform deployments, calibration of autonomous sensors, and for measurements that cannot yet be made autonomously, including water samples
Mooring with crawlers
Array of autonomous floats (Lagrangian and profiling), floating sediment traps, underwater gliders, surface wave gliders, and AUVs that can collect and preserve water samples
Floating sediment traps (carbon flux)
Sensors for ships and in-water platforms: nutrient sensors, optical sensors (chlorophyll fluorescence, backscatter, beam attenuation, spectral radiometers), phytoplankton sensors (flow cytometer, imaging flow cytobot), zooplankton and aggregate sensors (holographic camera, video plankton recorder, multi-frequency or broadband acoustics (to determine taxa/species), nets and water sampling from ships
MOCNESS system

What is the workflow?
1. Recruit students and postdocs.
2. PI planning workshop (**in person** with PIs from multiple locations; real face to face time is important in collaborative research endeavors).
3. Study existing data for region (including model data).
4. Develop detailed data management strategy.
5. Instrument and platform integration.
6. Calibration of instruments (identify or build calibration facilities; agree on sensor calibration strategy for pre-deployment, post deployment and during mission, post recovery).
7. Deployment of instruments (moorings, floats, gliders, etc.).

8. Cruises to do the field studies and calibrate sensors on autonomous platforms (collect measurements that cannot be done autonomously; develop proxy relationships for sensors – e.g., particulate organic carbon from optical backscatter; zooplankton biomass from acoustic backscatter).
9. Post-cruise calibrations, Data analysis, synthesis, and submission.

What is the desired outcome?
A better understanding of subpolar ecosystem dynamics for models that will help manage and predict the ecosystem, particularly in the face of changes in the region due to climate change).

Improved skill in use of autonomous sensing systems.

Note any tangible actions from EarthCube that would facilitate the goal.
- As part of the **assessment of available data before deployment** we may find that some of the data resources are difficult to discover, and integration of those data presents a significant challenge.

- Help with dealing with multiple data formats to cut down on the time scientists spend on this instead of research. Some data are a single parameter (e.g. temperature), others may be multiple parameters (e.g. prey). Creating a framework to make it easy to extract the needed information.

- Archive and annotate images and broadband acoustics.

- Help with working with and visualizing Eulerian and Lagrangian sampling efforts (data streams from moored instrumentation, gliders, floats and ship-based sampling) - in a synoptic and synthetic way. Currently this type of synthesis is done in an *ad hoc* manner, but there is desire for a better, more efficient way.

- Improved systems are needed to accommodate increased levels of data entry/ingestion from sensor systems and near real-time analysis and data QA/QC, including calibration records and proxy details.

- Visualization tools for data exploration using broad array of data sets and portal for discovery by others.

Use Case 2: Revisit of anthropogenic ocean carbon data/model intercomparison

Group Members:

Bob Key,
Sabine Mecking,
Krisa Arzayus,
Andrew Maffei,
Steve Diggs

What is the scientific problem?

Is the integrated air-sea CO_2 flux inferred from surface ocean pCO_2 measurements consistent with the global anthropogenic CO_2 inventory and distribution; why are the current coupled models so different in their predictions for both in the Southern Ocean (defined as south of 35S)?

What is the goal (how will you address that problem)?

- Determine why model and data approaches give different results.
- A new set of tools to aid model and data comparison.
- Determine if the interior ocean and surface ocean data sets give different results.

Who are the people and what roles do they fill?

- Surface/Ocean CO_2 expert
- Ocean Interior Carbon specialist
- Inverse modeler
- Someone who understands chemistry and current versions of earth system models
- Someone to write tools to allow comparison of data
- Informaticist to conceive of more general software to do data/model comparisons.

What things are needed (data, tools, etc.) to accomplish the goal?

- pCO_2 surface ocean data (SOCAT, etc.)
- GLODAP2 (discrete and gridded)
- Set of earth system models
- Matlab, S+, or R

What is the workflow?

1. Download data plus estimate the uncertainties.
2. Locate the model output from Earth System Model.
3. Massage data so they look like same sort of units, formatting, grid sizes etc. for comparison.
4. Find some way to deal with model's gridded data (e.g. how to deal with model's gridded data?; e.g. how to deal with weird grids?).
5. Grunt work (PhD student work).
- Conduct error estimation (e.g. Monte Carlo or other) of both measurement datasets (try different things).
- Global quantification of differences.
- New inversions.
6. Identify differences from the inverse methods.

What is the desired outcome?

- 3D anomaly maps.
- Difference between anthropogenic carbon flux for each ocean region as seen in previous ocean inversions.

- Recommendations for making new measurements to improve the situation given constrained funding.
- Recommendations on how the models can be improved.
- Recommendations on a software system that would make such a comparison easier independent of parameters.

Note any tangible actions from EarthCube that would facilitate the goal.

- Fund the work of building the tool to do data/model comparisons (future candidates would be World Ocean Atlas nutrients with satellite sea surface temperature).

Use Case 3: Impacts of sea ice reduction

Group Members:

Facilitator: Bob Groman,
Recorder: Lisa Raymond,
Stace Beaulieu,
Kim Bernard,
James Conner,
Krista Longnecker,
Jim Simons,
Ellen Tyler (remote)

What is the scientific problem?
How would the sea ice loss in the Arctic change physical mixing, biological productivity, ocean chemistry, and flow of carbon through the food webs and to depth?

What is the goal (how will you address that problem)?
To understand how sea ice reduction is impacting circulation and mixing and how these will impact productivity and carbon cycling.

How to address - Establish baseline against which to compare, figure out the components / parts of the system that need to be assessed, and then figure out how to do the assessing, i.e. through budget model approach and/or three dimensional modeling.

There is a need to think "outside the control volume" of the Arctic.

Who are the people and what roles do they fill?
Physical oceanographer (observational and modeler)
Chemical oceanographer-
Sea-ice expert - satellite expert & also HF radar/smaller scale circulation expertise
Climate modeler - hindcasting and forecasting
Food web modeler
Biological oceanographer
Top predatory specialist/Marine ecologist

Benthic ecologist
Informaticist
Climatologist

What things are needed (data, tools, etc.) to accomplish the goal?
Access to historical and current data
Ships
Data under ice today - long term moorings, sediment traps, ice-tethered profilers
Instruments to measure fauna and flora
Surface and under ice trawl
Tools to integrate across these datasets
Modeling system and way of integrating individual models - (to include forecasting)
Incorporation of individual scientists' datasets into the larger project - a space to stage and layer project data in the context of a larger project
Real time data needed to get information, and analyze it in a timely manner to answer policy questions
Data over longer time scales to accommodate translation up the food web to impacts on higher trophic levels

What is the desired outcome?
Describe impact of short and long term impact of reduction of sea ice thickness and enable forecasting (and planning) of potential responses in impacted communities.

Note any tangible actions from EarthCube that would facilitate the goal.
Informatics consulting (workshops) at beginning and throughout projects. (i.e., EarthCube can support workshops during the technology development phase to help disciplinary experts come together to understand interactions through what their data are telling them, and at the technology deployment phase in terms of training use of technologies developed.)

Technology that supports collaboration across disciplines. It's important to have access to this kind of consulting throughout the project.

Use Case 4: Coastal Ecosystem Responses to Climate Change

Group Members:

David Kline (Lead; recorder),
Mark Schildhauer (recorder),
Karen Stocks,
David Glover,
Rob Dunbar,
Emily Law,
Danie Kinkade,
Mimi Tzeng (remote)

What is the scientific problem?

How will coastal ecosystems respond to anthropogenic stressors, especially increasing atmospheric carbon dioxide levels, and its related impacts (e.g. OA, warming, hypoxia, sea-level rise, and storm frequency/intensity). How will environmental variability driven by the biology influence the interactions between these multiple environmental stressors, and how will the biological/chemical feedbacks change in the future?

What is the goal (how will you address that problem)?

1. Integrate relevant multidisciplinary data, both new and pre-existing.
2. Expand the spatial and temporal scale of the data.
3. Harmonize these data to inform an integrated ecosystems model that provides insights and predictions relative to coastal vulnerability (human factors, biodiversity, fisheries sustainability, etc.).

To acquire and integrate biological, physical, and chemical datasets from different coastal ecosystems in order to inform a predictive model. The model should be spatially explicit, and embody increased understanding of the fundamental differences in impacts of these stressors within and across ecosystems (kelp, coral reefs, mangroves, estuaries, seagrass beds, engineered/developed coastlines, etc.), including geospatially-localized influences (e.g. basin-specific impacts). The predictive model should integrate high resolution carbonate chemistry, environmental data, biophysical data and ecosystem function (rates and interaction strengths). This effort will also help **identify critical data gaps** for next generation efforts.

Who are the people and what roles do they fill?

Data wranglers (people to get data)
Chemists and biologists
Instrument technicians and engineers
Simulation and numerical (ecosystem) modelers
Workflow and knowledge modelers
Informaticists/database managers
Geospatial experts
Research software engineers (tools and interface development)
Facilitators/liaisons across domains (cross-disciplinary practitioners?)
Communication specialist for outreach and education
Students and early career researchers
Socioeconomic researchers/resource managers

What things are needed (data, tools, etc.) to accomplish the goal

Coastal terrestrial and oceanic data (including fresh water inputs); broader scale oceanic data
Socioeconomic data from adjacent land/sea areas
Instrument arrays with the appropriate software- find existing software and improve/enhance for the community
Improved data discovery and data integration tools
Improved tools for discovering and using existing code-bases for model construction

Evaluation that the network is present and robust
Need for shared storage and CPU resources for data intensive computing (cloud?)

What is the workflow?
Short term-goals (1-3 years): Identify, collate and harmonize relevant data while developing and implementing "EarthCube-compliant", standardized, scalable cyberinfrastructure approaches. Also perform a gap analysis to identify what is needed to produce a prototype model (and develop the current state with a meta-analysis) and pick 2 key sites with the best data (i.e. kelp or coral reef LTER sites) to begin development.

Medium to long-term goals (3-10 years): Intake the appropriate datasets to fill the major data gaps and advance the model development to advance, validate, and improve the model. Develop a coastal ecosystem climate change portal to deal with newly measured data and for the community to interact with and improve the models.

What is the desired outcome?
- To develop an adaptive model that incorporates the chemical, biological and socioeconomic data in order to make predictions useful for management and to minimize negative socioeconomic impacts.
- To create a community portal for applying, improving, and supporting the model through time.
- To create a robust integrated information resource as well, that is useful for coastal planning outside the modeling framework. Engage the resource managers to ensure that the models are providing data useful to their mission.

Note any tangible actions from EarthCube that would facilitate the goal.
Better tools for discovering relevant information across independent data frameworks.
Development of integration tools for the diverse data types.
Development of coastal ecosystems climate change portal.
Collaborative website for data sharing and model development.

Use Case 5: Changes in Coastal Plankton Communities over Decades

Group Members:
Heidi Sosik,
Vembu Subramanian,
Dana Hunt,
Todd O'Brien,
Hilary Close,
Erik Zettler

What is the scientific problem?
How have the structure and function of coastal plankton communities changed over recent decades? What are the effects that propagate through the food web to influence

fish, mammals, and seabirds? How have the occurrence and distribution of HABs events changed and what are the contributing factors?

What is the goal (how will you address that problem)? To mine historical physical, chemical, and biological data to consider how things have changed and how current changes are similar or different from those in the past. Develop and use models to assist in the analysis.

Use these scientific data to guide management in order to benefit humans and the environment.

Who are the people and what roles do they fill?
Modelers – predictions
Researchers - data discovery, analysis
Federal and state resource managers/ researchers- support with monitoring data, refine how science could guide management
IOOS Regional Associations - provide inventories of assets, data, etc.
Informaticists - data discovery, data recovery/ mining, build data tools, develop, document, and maintain analysis workflows
Educators / Communicators - reach out to owners of dark data, raise public involvement and support, train students to support project goals

What things are needed (data, tools, etc.) to accomplish the goal?

Data: *(datasets spanning large temporal coverage)*
Habitat characterization (turbidity, nutrients, upwelling), remote sensing
Biological data (species identification, counts)
Overlap data acquisition between old and new technologies in order to standardize them

Tools:
Visualization tools for both raw data and numerical models
Analysis tools to fuse multiple kinds of data and levels of resolution (spatial and temporal resolutions)
Integration tools for metadata
Integration tools for molecular and microscopic data on plankton
Interaction networks (representing biotic interactions in an ecosystem) to examine replacement, redundancy, interactions of taxa

What is the workflow?
Identify and compile dark and existing data sets (old papers, online datasets, etc.).
Analyze and interpret data to identify current gaps in observing data time series.
Work to fill the gaps with additional monitoring.
Analyze and interpret data to answer questions.

What is the desired outcome?
Quantification of changes in plankton through time.

Assessment of plankton changes in the context of environmental drivers.
Identification of key locations and variables for future monitoring.
Feedback to policy makers- predictions of what happens under various conditions, or explanations of existing conditions (e.g., fisheries yield, HAB occurrence), development of best practices.
Increased collaboration between researchers, and end-users (including resource managers, policy makers and the general public).
New hypotheses about how environmental changes structure plankton communities.

Note any tangible actions from EarthCube that would facilitate the goal.
Methodologies to deal with different data resolution (temporal and spatial) and integrating those data.
Data portal to find, discover, and integrate datasets that also serves as a collaborative workspace.
Automated methods of mining and integrating data from the literature and other sources (dark data).
Systematics approaches for dealing with biological taxonomic information in mixed data sets and in regard to fluid taxonomic standards.