EXECUTIVE SUMMARY: EARTHCUBE WORKSHOP RESULTS

Earth Cube Workshop Title: Developing A Community Vision Of Cyberinfrastructure Needs For Coral Reef Systems Science

Convenors: Dr. Ruth D. Gates (Researcher, HIMB, UHM) and Dr. Mark Schildhauer (Director of Computing, NCEAS, UCSB)

Organizing Committee: Dr. Megan J. Donahue (Associate Researcher, HIMB, UH Manoa), Dr. Peter J. Edmunds (Professor, CSUN), Dr. Erik C. Franklin (Assistant Researcher, HIMB, UH Manoa) and Dr. Hollie M. Putnam (Assistant Researcher, HIMB, UH Manoa)

Workshops: 1. September 17-18, 2013, University of Hawaiʻi: Hawaiʻi Institute of Marine Biology (HIMB) and 2. October 23-24, University of California: Santa Barbara – National Center for Ecological Analysis and Synthesis (NCEAS)

INTRODUCTION: The purpose of the two workshops for the ***Developing a Community Vision of Cyberinfrastructure Needs for Coral Reef Systems Science*** project was to gather input from end users and data generators on the role that cyber-enabled data tools can play in addressing the key science drivers and grand challenges in the field, and in enhancing the value, scope and impact of coral reef systems science. A total of 53 participants, representing a broad geographic range of the U.S. academic coral reef research community, attended one or both workshops (in person or virtually). The coral reef scientific community focuses on a critically important and threatened ecosystem and is extremely diverse from a disciplinary perspective, crossing the boundaries of biological, physical and chemical oceanography, climate science, remote sensing, modeling and engineering. Research in the field spans genomics and ecosystem science, and data generated by these activities crosses broad biological, temporal and spatial scales.

SCIENCE ISSUES AND CHALLENGES

1. Key Science Drivers/ Questions in Coral Reef Science: Participants identified several high-priority science questions that will be the focus of interdisciplinary efforts during the next 5-15 years

- What processes are relevant to understanding the biological responses of coral reefs to biotic and abiotic drivers across temporal and spatial scales?
- What are the mechanisms of coral reef adaptation and acclimatization to climate change?
- How does symbiosis influence the biology and ecology of coral reef organisms?
- How does the abundance and diversity of coral reef organisms influence community resilience at local, regional, and global scales?
- How will invasive species, disease, and parasites disrupt coral reef ecosystem structure and function?

2. Grand Challenges in Coral Reef Science: Several themes emerged as consistent challenges faced within/across the involved disciplines.

- Data utilization and accessibility for automated processing, standardization and measurement was identified as the highest priority, and includes data cohesion across spatial and temporal scales, as well as disciplines, and application and access to this data, from omics to ecosystem modeling.

- Rapid developments in bioinformatics and –omics sciences provide new tools to address taxonomic, genetic, ecological, and evolutionary questions but there is a great need to develop methodologies to efficiently utilize these tools within the coral reef science community
- There should be improved training opportunities in communicating science to peers outside the field as well as to better inform policy and educate students and the public.

TECHNICAL INFORMATION/ISSUES/CHALLE NGES

1. Desired tools, databases, etc. needed for pursuing key science questions with brief elaboration:

There were four distinct classes of community needs: (1) Databases and Portals, (2) Data Processing, Modeling and Visualization, (3) Education and Training, and (4) Internationalization.

(1) Databases and Portals:

- Desirable features include standardization (formats, collections, and representation), richness of metadata, and built on existing efforts with tools to create and query data across repositories that includes standard reference keywords, DOIs, and appropriate credit for data provider. Data integration should support imaging, sequence, environmental sensor data, and local observational data and delivery of streaming real time data from sensors networks. Quality of data and metadata is critical since web applications and interfaces may involve with time.
- Needs to include links/connections to existing resources through a curated data portal that could cluster databases/data by types. The system should allow grass roots contributions through user data entry as well as information filtering and discipline or research theme by sub-setting. Digital tools should also provide a dynamic, collaborative workspace for a variety of sub-disciplines (bioinformatics, ecological studies, genomics, mathematical biology, programming, etc.). Identifying a funding strategy for sustainable data curation is critical.
- There are many “dark data” or challenging data types (such as imagery or sequence data) that could be better managed and harvested from unpublished studies, secondary reports, desk drawers, personal collections, original data from earlier publications not archived that require a various types of standards and integration methods. Improved methods to deal with these data types may arise from industry-academic technology sharing translated to the coral reef research community.

(2) Data Processing, Modeling, Visualization:

A system of data processing pipelines for bioinformatics/omics data for computationally intensive analysis tasks is critically needed especially those that take advantage of HPC resources (XSEDE). The KEPLER scientific workflow system is a potential tool to utilize. These approaches would include traditional statistical modeling approaches, machine learning, and geospatial analysis and multimedia analysis for image/video/audio analysis and information extraction.

- We could visualize disparate data (space/time) with “easy to use” software tools (*vs.* immersive environments) that support online visualization simulations with user-directed parameters. Google Earth is a reasonable model for the portal interface. The visualizations can be used to communicate directly with public through interactive and applied community engagement. Maintain data and software version control will tools such as GIT or Mercurial.
- Improved software tools (such as API’s: application programming interfaces) for linking ecological and -omics software packages are needed with open standards that facilitate coupling through modular-based software or middleware to connect processes; better interfaces for communication among software models; Free Open Source Software (FOSS); Glue code and provenance: automated metadata extraction/provenance from digital objects and (e.g., OpenDAP, HDF (file format): geodata, temporal metadata; Integration/ alignment: scale; measurement equivalence, reward coders and nurture new type of coral reef scientist/hacker
- Develop a coral reef simulation system that merges model components (forecast, climate change), is applicable to many locations; 3D; modular, with case studies; the WRF is an example (Weather Research and Forecasting)

(3) Education and Training (human resources/workforce)

- The support of a cyberinfrastructure tactical team to support training, scientific programming, and database administration would help facilitate many of the data analysis, education, and training needs. The coral reef domain aware CI team could rotate between thematic-based resources and help with challenging projects to achieve scientific end products that non-CI researchers would have difficulty creating alone. This role may also be fulfilled by a “Campus Champions” such as a grad student or similar to connect geoscientists and computer scientists.
- The coral reef community would support web-based workshops on portals for data; data integration; data management and mechanisms for local group participation (e.g., web-based). Various topics for education programs were proposed including programming, data management; online repositories/versioning; imaging technologies/analysis; tools for temporal/spatial scaling; communication with managers; promotion of cross-disciplinary training and research.
- University programs in coral reef sciences may institute computer programming requirements in curricula or develop and offer a Certificate Program in Coral Reef Informatics to encourage cross-disciplinary training (between geoscience and computer science)

(4) Internationalization

- Need to improve access and use of international data; issues include need to share data to improve our understanding of reef globally to promote international cooperation on global scale reef studies. We would like to connect people, institutions, government management to democratize research and get many contributors to interpret science and results.

COMMUNITY NEXT STEPS

- Maintain momentum and communicate
- Develop a formal network to facilitate the interface with one another and with EarthCube.
- Respond to solicitations for input from EarthCube such as the call for Research Coordinated Network proposals and Geoscience community activities
- Identify existing resources
- Summarize, visualize and communicate results to broader coral reef community
- Solicit feedback from the community on the proposed coral reef science scenarios and other workshop outputs
- Promote community buy-in by moving from "talking activities" to "implementation activities"
- Link EarthCube activities to ongoing activities, e.g., EPSCoR and other funding program
- Link to software infrastructure funding programs
- Identify possible partnerships for proposals
- Reach across EarthCube user groups to identify common needs and initiatives
- Conduct metrics on progress on recommendations, e.g., resource lists, status of specific workshop suggestions in 1 year, 2 year, etc.