Data Management Plan
RI NSF EPSCoR

Data Collected
We utilize the standard definition of research data (based on the 2009 "Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age" by the National Academy of Engineering, National Academy of Sciences, and Institute of Medicine) as "information used in scientific, engineering, and medical research as inputs to generate research conclusions". In addition, we include a) art and design data and b) the management of EPSCoR/**RI C-AIM** participant data (e.g., information on publications, presentations, grants submitted/awarded) that will be collected by the **RI C-AIM** office for NSF annual reporting purposes.

Primary research data (in the natural, physical, and social sciences and engineering) will be collected through a variety of techniques (e.g., sensor measurements, physiological assessments, image acquisition). Art and design data include, but are not limited to, representations of information, ideas, and objects in material or digital formats and original/creative works. All investigators will comply with standard management policies of their respective institutions, **RI C-AIM** and the NSF Office of Integrative Activities. These policies include the requirement that researchers submit data to the appropriate project repositories (see below) in order to maintain their funding.

To insure data sustainability, archiving and accessibility, we will establish the **Rhode Island Center for Data Discovery**. This Center, hosted by CCV, will be easily accessible to the general public and **RI C-AIM** investigators from all of our institutions and will create a framework and web portal for investigators to archive, search and access data and disseminate their data to the general public. The Center will house various forms of data with associated metadata including videos and images as well as model outputs and observational data in standardized formats (e.g., HDF5 or netcdf). This Center will benefit from improved fiber optic connectivity throughout New England made possible via a NSF EPSCoR Track 2 grant (NSF Award #0918061) along with other funding sources. RI will leverage prior Track II support as well as enhance our capability for long-term data archiving.

Data Types, Formats and Standards
Due to the wide scope of research that will be conducted, data types will include (but not necessarily be limited to) ecological and environmental data from field and laboratory experiments, genomics and proteomics data, image data, bioinformatics data, economic, and qualitative data. We will collect both raw and processed data via a wide variety of instrumentation, including real time qPCR, mass spectrometry, and TEM. (See *Facilities* plan for additional information.)

Data will be captured in numerous file formats, including, wherever possible, in tab-delimited text files (<10MB) that are compatible with spreadsheet programs such as Excel. Associated metadata will be collected using current specification standards such as ISO 19115-2, FGDC, Dublin Core, or Ecological Metadata Language (EML) where possible, allowing the files to be immediately ingested by the RI Center for Data Discovery. Additional metadata will include experimental and/or instrument details (i.e. as readme files), date and user identity, so that the raw and processed data can be properly interpreted. We will store community diversity data (raw reads from sequencing) in standard folders with a file naming protocol.

Data Storage and Archiving
All data (including images) will be stored digitally, and all associated metadata will be linked to the primary data files. All data will be stored on multiple systems at multiple locations (as detailed below) and will be retained for at least three years after the end of the grant. Individual investigators (as well as **RI C-AIM** administrative and outreach staff) will store their own data, in compliance with their institution's data management policies. Genomics and bioinformatics data will be archived at URI's High Performance Computing Center and Brown University's Center for Computation and Visualization (CCV; *http://ccv.brown.edu*). All other research data will be archived at the RI Center for Data Discovery, where servers mirrored to URI will ensure data backup and security. All individual lab notebooks will also be stored by the individual researchers, and will be preserved for at least three years following the end of the grant. In

addition to individual researchers choosing to store their own data for a longer term, data will be selected for long-term preservation at URI, based on its potential to enhance future scientific research, cyberinfrastructure, and/or workforce development.

We will also work with existing data repositories in the life, computer, and social sciences. Repositories that are relevant for our project include the:
• BCO-DMO (Biological and Chemical Data Management Office -- NSF; http://www.bco-dmo.org/),
•NCBI (National Center for Biotechnology Information; www.ncbi.nlm.nih.gov),
• NERACOOS (Northeastern Regional Association of Coastal Ocean Observing Systems; http://www.neracoos.org/),
• NODC (National Oceanographic Data Center; www.nodoc.noaa.gov),
• GSO Long-Term Time Series (URI's Graduate School of Oceanography long-term data set repository; http://www.gso.uri.edu/research/longterm-research),
• DEM-BART (RI Department of Environmental Management Bay Assessment and Response Team; http://www.dem.ri.gov/bart/),
• NarrBay.org (Narragansett Bay, RI portal for ocean researchers; www.narrbay.org),
• SBML (Systems Biology Markup Language; sbml.org).

RI NSF EPSCoR is currently part of the ERCore (EPSCoR Reporting Core) Consortium, which incorporates 10 NSF EPSCoR jurisdictions. ERCore is a Drupal module that has been developed to meet the annual reporting requirements of NSF EPSCoR. All participants will have accounts in ERCore, where they will report all publications, grants, presentations, patents, etc., and EPSCoR administrative staff (PA, EOD Coordinator, Communications Coordinator, PI, coPIs) will enter in administrative, education, and outreach data. The module is, and will be, be stored on a physical server at Brown University's CCV and is administered by Kia Huffman (Cyberinfrastructure Architect, Brown U.). ERCore files and database are also backed up nightly, to a IBM Tivoli tape storage system that Brown CCV maintains, along with a secondary backup in the Brown BioMed Center.

Dissemination Methods
Results from research data will be presented at conferences appropriate to the investigators' discipline (e.g., Ocean Sciences, FASEB) and published in peer-reviewed journals (e.g., Nature Biotechnology, PNAS) in a timely fashion. Electronic data will be available for public sharing within two years of being acquired or once it is published, whichever is sooner, assuming that there are no intellectual property concerns. Data archived at the RI Center for Data Discovery will be available for web distribution, with protocols established to manage permission levels governing data access. All data will be packaged with associated metadata prior to distribution to ensure proper interpretation and analysis.

Policies for Data Sharing and Public Access
When inventions and/or patents result from RI NSF EPSCoR-sponsored work, the individual investigator(s) will work with their institution's office of Intellectual Property/Commercialization to ensure compliance with all NSF grant policies and federal laws regarding inventions/patents. Investigators will also share information on their inventions/patents with the **RI C-AIM** administrative team, for logging in ER Core, our internal reporting system. Ethical/privacy issues should be minimal, as personal data collected on **RI C-AIM** participants will be released only in aggregate form (i.e., the participants in a particular **RI C-AIM** event were 55% female, 45% male). For any research projects that do incorporate human subjects' data, individual investigators will be responsible for ensuring that the privacy and rights of those participants is continually upheld. All researchers will also adhere to Institutional Review Board (IRB) policies within their individual institution.

Roles and Responsibilities
The **RI C-AIM** PI, PA, and co-PIs will be responsible for general oversight and maintenance of **RI C-AIM** administrative, education, and outreach data. In addition, each co-PI will coordinate the management of data for their research focus group and will work with individual researchers, including faculty, postdoctoral researchers, and graduate students in their group to ensure and monitor common, local (i.e., Rhode Island) data storage and archiving techniques, including transmitting all relevant data and associated metadata to the RI Center for Data

Discovery. Adherence to this data management plan will be assessed at quarterly, all-hands research meetings for each research theme. If there is turnover within co-PIs, the outgoing co-PI will train the incoming coPI on the existing data management strategies to ensure a seamless transition. If original personnel were no longer available/participating in **RI C-AIM**, the current PI/coPIs would have responsibility for making decisions regarding data archiving.