Data Management Plan

For the major new synthesis of data that we propose for geographic evidence of source-sink dynamics, it is particularly important that the data are well managed and made accessible to other investigators. This has always been a high priority for us. The PI (Rex) and one Co-PI (Stuart) have extensive experience compiling large databases on deep-sea community structure, and making the data generally available. We compiled the first global database on biological sampling in the abyss from the Challenger expedition to the present. The data set includes sampling station locality, cruise, date, sampling gear deployed, component of the benthos sampled, and where material is archived. It is available on the website of the Census of the Diversity of Abyssal Marine Life (a division of the Census of Marine Life) maintained at the Forschungsinstitut Senckenberg (www.cedamar.org/Abyssal-benthic-biological-sampling). Data can be obtained for any site by clicking on a global map of station localities, or any part of the dataset can be downloaded. Rex and Stuart also compiled the first global database of deep-sea benthic standing stock (abundance and biomass) for bacteria, meiofauna, macrofauna and megafauna (Rex et al. 2006) that includes all station data, sampling procedures, and published references. This database has served as an important part of more recent syntheses of energy in the deep sea (e.g. Wei et al. 2010). We have ensured that all of the extensive gastropod and bivalve material collected by the Woods Hole Oceanographic Institution's Benthic Sampling Program has been incorporated into the collections of the Museum of Comparative Zoology at Harvard, where it has been properly accessioned and catalogued. It is stored securely in alcohol, and a computerized database makes it available (http://collections.mcz.harvard.edu/Molluscs/MolluskSearch.htm). All of the access information was published in the primary literature. Solange Brault (Co-PI) has extensive experience in database management, and in applying the large matrix manipulations necessary to perform nested analyses.

For the present study, our plan is to place all of the existing data described in the proposal on to an Excel spreadsheet. This will include all station locality information, data, cruise, gear deployed, and location of archived material. Careful checking and cross checking will be needed for quality assurance. Locality data for each species will be compiled in two forms: the full depth range, and actual individual sites occupied. Then we will apply nested analysis to presence-absence matrices by using the eight nested metrics and six null models, and Baselga's nested vs turnover analysis; and then regress nested ranks against regional animal density patterns (Rex and Etter 2010), and modes of larval dispersal (where available) as described in the proposal. Locality information for species will be contributed to The Ocean Biogeographic Information System (OBIS), and the full database to National Oceanographic Data Center (NODC), and appropriate Data Management Office within NSF time constraints (6 months from compilation) and published in a timely way. As far as we know, there are no restrictions of privacy, confidentiality, security or intellectual property associated with these data. All contributing parties will be fully involved in collaboration in the research and given authorship. Publication and reporting to NSF will include any statements of proprietary rights. All of the material that we plan to analyze is currently archived properly at the MCZ, Southampton National Oceanographic Centre, or Texas A&M University.

For the reproductive study, the principal variables collected will be oocyte size, body size, completeness of spermatogenic cycle, and evidence of mating. This will be compiled into a

spreadsheet that includes all station locality and sampling information as described above. We believe that the best way to make this information generally available would be to include it in electronic attachments to publications. We will also submit it to the appropriate DMO as part of the NSF reporting process so that it can be accessioned into a National Data Center.

This data management plan is based on The Division of Ocean Sciences Data and Sample Policy document. We have endeavored to address all of the relevant issues, but will certainly comply fully with any additional guidelines.