

## Data Management Plan

1. **Overview.** This data management plan (DMP) is part of the proposal “*Collaborative research: Keystone chemicals: Identifying general and universal molecules of fear*”.

### 2. Expected Data

The following data will be collected as described in the proposal:

1. Morphological responses to oysters exposed to different chemical cues
2. Behavioral responses of mud crabs exposed to different chemical cues
3. Chemical data and samples

**Data Formats.** Data formats include raw image files (1), standard spreadsheets containing experimental data (2) and mass spectroscopy data (3) and text files of chemical structures (3). Image files will be stored in tiff or jpg formats, with a master list of file names and experimental treatments as txt files, Compiled experimental data will be stored as tab delimited txt files. Physical notebooks containing experimental data will be retained and will be available for all legitimate scientific queries.

### 3. Data Storage.

Data that is expected to be retained for sharing and archiving include all publication data, all data on oyster morphological and mud crab behavioral responses, all data on chemical analysis and results.

**Storage and backup during the project.** The PIs will be responsible for operational data storage and backup during project execution. Behavioral and morphological data will be collected and/or stored on PCs and laptops with frequent scheduled backup to secured institutional servers.

**Data capacity and volume.** We anticipate 1-2 GB of image files and 100-200 MB of txt files and mass spectroscopy data files.

**Security.** All data stored during and post project execution will be within the GT and DISL campus network infrastructure, which employs firewalls and secure authentication and authorization methods for login and access.

**Long term archiving and preservation.** Long-term storage will be done on DVDs with duplicate backups provided to each PI. Given the need to synthesize the data from the experiments, it seems likely that the data will be most useful when presented in a compiled form. Thus, we will work with the Biological and Chemical Oceanography Data Management Office (BCO DMO) to present the data when organized as responses of experimental replicates to the different experimental conditions for both oyster morphology and field experiments. We anticipate that much of the data will be associated with particular publications that present the statistically analyzed data in tabular and/or graphical form. The BCO DMO data will be linked to this work and the BCO DMO data repository will be referenced in the publication as appropriate. Compiled data also will be uploaded to publically available servers at GT (<https://smartech.gatech.edu/>), which will include a reference to the paper that

contains the analysis of this data. *All data will be uploaded to the BCO DMO site with descriptions of the experiment and measurement variables within two years after project completion.*

Pure chemical compounds isolated from crab exudates will be stored frozen in the Kubanek lab at Georgia Tech, and made available to external scientists upon request following publication of molecular structures and chemical composition. Extracts consisting of multiple compounds will be stored frozen in the Kubanek lab and also will be shared upon request. Outcomes of bioassay-guided fractionation and metabolomics experiments will be published as results (in figures and/or tables) in peer-reviewed journals. Structures of novel compounds will be published in journals that are monitored by the Chemical Abstracts Service (maintained by the American Chemical Society). This service abstracts all molecular structures published in peer-reviewed chemistry-related journals (not only American journals) and provides these structures to the SciFinder database. This database, to which most universities and research institutes around the world subscribe, allows searching for molecular structures and sub-structural features and links to journal abstracts. Metabolomic mass spectral data will be shared pre-publication via a Sharepoint secure server hosted at Georgia Tech. This server will have a system of permissions so users are granted access based on a hierarchical system. Published datasets will be made available via the School of Chemistry's WebServer. *We will also submit project details and data to the BCO-DMO system within two years of project completion.*

**Roles and Responsibilities.** The PIs will have decision making authority over all data management. The PIs will draft the overall data management policy during the first three months of the project to confirm their shared understanding of roles and responsibilities. This document will be reviewed yearly.

#### **4. Data Retention**

**Operational Data.** Following the conclusion of the proposed project, data will be retained for a minimum of five years by the PIs, as described in section 3.

**Archival Data.** The archival lifecycle and retention policy for archived data will be managed by the PIs as described in section 3. The retention period for this will be determined at the conclusion of the project.

**5. Data Sharing and Reporting.** All data will be made publicly available 24 months after the completion of the project as documented above. Prior to this, the data will be shared on a secure FTP website in response to legitimate scientific inquiry.