Plan Overview

A Data Management Plan created using DMPTool

Title: “Collaborative Research: Magmatic and Tectonic Drivers of Geological, Hydrothermal and Biogeochemical Processes on the Mid-Cayman Rise”

Creator: Kenna Rubin - ORCID: 0000-0002-8554-1337

Affiliation: University of Rhode Island (ww2.uri.edu)

Principal Investigator: Kenna Rubin, Adam Soule, Katherine Kelley, Jessica Warren, Jill Mcdermott, Jeffrey Marlow

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: 1620

Template: NSF-EAR: Earth Sciences

Project abstract:

Multidisciplinary study of volcanic, tectonic, hydrothermal, and ecological processes on the Mid-Cayman Spreading Center, using deep submergence tools (i.e., the human occupied submersible (HOV) ALVIN and autonomous underwater vehicle (AUV) Sentry). The project will generate subsea video and still imagery data, physical samples collected by HOV Alvin, along-track sensor data and navigation from HOV Alvin and AUV Sentry, CTD and other sensor data Sentry and a CTD rosette. Data will undergo QA/QC by the science party and, in some cases, processed by the vehicle and vessel operators. Some data products will also be generated by shore-based modelling and LB analyses. Metadata will be generated for all.

Start date: 06-01-2024

End date: 05-31-2026

Last modified: 10-16-2023

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal.
“Collaborative Research: Magmatic and Tectonic Drivers of Geological, Hydrothermal and Biogeochemical Processes on the Mid-Cayman Rise”

Data and sample types

List the types of data and samples to be collected and/or generated. The listing of each data/sample type should briefly identify what metadata will be provided and when data/sample preparation will be considered complete. (Definitions of “data” and “samples” are explained above within “EAR requirements.”) For proposals providing community-serving infrastructure or research services, the DMP should describe the data/sample types to be managed and what guidance or support will be provided to help users meet their data/sample sharing obligations. EAR recognizes that data/samples may undergo multiple transformations in the research process (including destructive analyses), and disciplinary expectations for assignment of metadata and retention of intermediate data and sample products may vary.

Product 1: Oceanographic Data - Alvin navigation and vehicle attitude and along-track sensor data. Creation Plan: HOV ALVIN data will be constructed by combining multiple sensor streams into a single file interpolated by time. Raw data will also be submitted. Anticipated Volume: 200 MB, Format(s): ASCII

Product 2: Oceanographic Data - Subsea video and still imagery. Creation Plan: HOV ALVIN and AUV SENTRY Video and still imagery data will be recorded digitally. Anticipated Volume: 25 TB. Format(s): Video will be stored in H.264 format and still imagery as JPG

Product 3: Oceanographic Data - Seabed bathymetry, backscatter. Creation Plan: AUV SENTRY sonar data will be recorded digitally. Anticipated Volume: 15 TB. Format(s): Bathymetry will be stored in grid file format. Backscatter will be stored as a grid file.

Product 4: Oceanographic Specimen (volcanic and metamorphic rock, grab specimens from HOV ALVIN). Anticipated Quantity: 250 hand specimens, processed photographed, described, and split at sea, weight 2000 lbs. Metadata for all samples compiled and registered with IGSN via SESAR.

Product 5: Oceanographic Specimen and derived compositional data (Seafloor vent fluid water samples from HOV ALVIN). Anticipated volume: 50 L. Metadata: samples will be given a unique ID based on dive and sample number. Georeferenced temperature and derived geochemical data associated with each water sample (including in situ methane sensor values) will be tabulated and made available through publications and deposition in EarthChem and SESAR. File types: .xlsx

Product 6: Oceanographic Sensor Data (in situ redox, CTD and methane data from HOV ALVIN and AUV Sentry). Data will be acquired digitally by sensors during deployments, downloaded and processed on ship. Data will be stored in xlsx format. Anticipated volume: 100 GB. Metadata and data registered through BCO-DMO

Product 7: Geological Mapping data (Merged geospatial integration and interpretation of geological features, and sampling sites). Data collected during dives and subsequent dive review will be compiled in ARGIS or Q-GIS, exported as GeoTIFF format, and made available as a derived data product through MGDS. Appropriate geospatial metadata will be tagged in the GeoTIFF and available as a txt file.

Product 9: Processed Geological Specimens Data (rock thin sections). Thin sections will be made of selected samples after the expedition and will be described petrographically and photographed. Metadata will be tied to parent sample IGSN as children. Anticipated Volume 250 thin sections; 100 GB of jpeg and xlsx documentation.

Product 10: Geochemical Data (derived compositional data, major element, trace element and volatile abundances, from glasses, rocks and minerals, collected on shore by electron microprobe, XRF, SEM, FTIR, LA-ICP-MS and solution ICP-MS, plus associated sample photomicrograph analysis maps when appropriate). Creation Plan: non-destructive or destructive compositional analysis of selected samples using grain mounts, powders, or dissolved splits of specimens. Anticipated volume of data: 100 MB. Data will be collected, reduced, processed, and compiled in .xlsx for preservation at EarthChem, with specimen data tied to sample IGSN metadata.
Product 11: Genetic sequence data. Creation Plan: DNA will be extracted from the sediments and rocks collected during the expedition and submitted for high-throughput sequencing. Sequence data, tied to specimen IGSN metadata, will be processed through field-standard quality control algorithms. Anticipated Volume: 500 GB, Format(s): Fastq text files


For each data or sample type, identify which personnel and institution(s) will be designated for its management, including contingency plans for the departure of key personnel from the project. For collaborative projects, PI(s) of the award(s) associated with the designated personnel and institution(s) are ultimately responsible for overseeing and reporting on their data and sample management activities.

Product 1: Soule (URI), contingency: Rubin (URI) - Intended Repository: MGDS; Duration of Availability: in perpetuity; License: CC_BY-NC-SA_2.0; Release Timeline: ASAP

Product 2: Soule (URI), contingency: Rubin (URI) - Intended Repository: WHOI Data Library and Archive; Duration of Availability: in perpetuity; License: WHOI Copyright, free to use for non-commercial activities; Release Timeline: 1 year after expedition.

Product 3: Soule (URI), contingency: Rubin (URI) - Intended Repository: MGDS; Duration of Availability: in perpetuity; License: CC_BY-NC-SA_2.0; Release Timeline: ASAP

Product 4: Kelley (URI); Warren (U. Del), contingency: Rubin (URI) - Preservation Plan: HOV Grab samples will be described at sea; all specimens will be registered with IGSNs. Physical Specimen Repository: Samples will be archived at URI-MGSL (Rock and Core lab). Specimen Release Timeline: 2 years from acquisition. Duration of Availability: Perpetual, Metadata Repository: SESAR and IMLGS. Metadata Release immediate upon archival at the repository.

Product 5: McDermott (Lehigh), contingency: Michel (WHOI): Preservation Plan: Fluids remaining after analysis will be archived in McDermott’s lab at Lehigh. All specimens will be registered with IGSNs. Metadata Repository: SESAR Specimen and Data Release Timeline: 2 years from acquisition. Duration of Metadata Availability: in perpetuity

Product 6: Marlow (Boston) and Shank (WHOI), contingency: Michel (WHOI): Preservation Plan: Specimens are consumed during analysis. all specimens will be registered with IGSNs. Metadata Repository: SESAR; Specimen and Data Release Timeline: 2 years from acquisition. Duration of Metadata Availability: in perpetuity


Product 8: Soule (URI) and Rubin (URI) Geological Mapping data - Preservation Plan: Derived maps will be published and made available as GeoTIFFs through MGDS with attached relevant metadata (deployment number and observations used to compile these interpretations)

Product 9: Kelley (URI); Warren (U. Del), contingency: Rubin (URI) - Preservation Plan: Thin sections will be archived in the Kelley and Warren labs. Data will be archived in EarthChem. All specimen sections will be tied to sample IGSNs. Metadata Repository: SESAR Specimen. Data Release Timeline: 2 years from acquisition. Duration of Metadata Availability: in perpetuity

Product 10: Kelley (URI); Warren (U. Del), contingency: Rubin (URI) - Data Release Timeline: 2 years from acquisition. All specimen results will be tied to sample IGSNs. Duration of Availability: Perpetual, Data Repository: EarthChem (PetDB).

Product 11: Marlow (Boston) and Shank (WHOI). Preservation Plan: Samples are consumed during analysis. Sequence data will be preserved by submission to the appropriate repositories at the National Center for Biotechnology Information (NCBI), where they will be assigned accession numbers and made publicly accessible within
For each data type listed, identify an appropriate long-lived FAIR-aligned repository for data deposit, the timeframe for public data access, and the expected period of data preservation. For each sample type listed, identify an appropriate long-lived FAIR-aligned repository for indexing sample metadata, the location for sample storage (preferably a repository appropriate for the specific sample type), and the expected period of sample preservation. (Required timeframes for data/sample access are specified above within “EAR Requirements.”) Many repositories commit to preserve access to data/samples indefinitely; any deviations from this expectation should be explained. PIs are encouraged to coordinate with designated repositories in advance of planned data/sample submission.

Repositories are listed in the prior section.

In most cases, it is sufficient for the DMP to identify the repository(ies) to be used and the timeframe for access and preservation for each type of data/sample identified. In these cases, the selected repository(ies) should align with FAIR principles and community-specific standards. Occasionally, appropriate long-lived FAIR-aligned repositories do not exist for certain types of data or samples. In such cases, it may be necessary to adopt alternative approaches to data access and retention, such as via use of a local computer server. In such cases, the DMP should explain how the proposed approach fulfills important attributes for FAIR-aligned repositories, consistent with OSTP guidance on “Desirable Characteristics of Data Repositories for Federally Funded Research.” These attributes include, but are not limited to, the following:

- **Findability.** Data should be findable via standard search tools, such as through the assignment of globally unique persistent identifiers (e.g., Digital Object Identifiers (DOIs) and International Geo Sample Numbers (IGSNs)) and rich metadata that is indexed in a searchable resource.

- **Accessibility.** Data should be publicly accessible to other researchers, at no more than incremental cost, within the specified timeframe. Any data access limitations must be justified. “Data available upon request” is not acceptable.

- **Interoperability.** To ensure interoperability, data should be described via appropriate metadata standards, in alignment with expectations of the associated scientific discipline(s).

- **Reusability.** To facilitate the broadest possible data reuse, data should be assigned clear and accessible usage licenses and metadata descriptors that identify provenance. EAR expects the adoption of unrestricted open licenses except with specific justification.

*not applicable*
Planned Research Outputs

Physical object - "Rock, Water, Bio specimens"
Physical samples collected with HOV ALVIN. Repositories listed with products. Metadata by IGSN.

Dataset - "AUV Sonar Products"
Bathy and Backscatter data.

Image - "HOV/AUV Photos and Video"
Images collected by vehicles

Dataset - "rock, Water, Bio analytical data"
Compositional or genetic analysis of specimens. Multiple repositories listed in products.

Model representation - "Water Column Plume models"
Data available through BCO-DMO and GitHub Public Repository with Zenodo

Planned research output details
<table>
<thead>
<tr>
<th>Title</th>
<th>Type</th>
<th>Anticipated release date</th>
<th>Initial access level</th>
<th>Intended repository(ies)</th>
<th>Anticipated file size</th>
<th>License</th>
<th>Metadata standard(s)</th>
<th>May contain sensitive data?</th>
<th>May contain PII?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rock, Water, Bio specimens</td>
<td>Physical object</td>
<td>2026-06-30</td>
<td>Open</td>
<td>None specified</td>
<td>100 MB</td>
<td>Creative Commons Attribution 4.0</td>
<td>None specified</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>AUV Sonar Products</td>
<td>Dataset</td>
<td>2025-06-30</td>
<td>Open</td>
<td>Marine Geoscience Data System</td>
<td>15 TB</td>
<td>Creative Commons Attribution 4.0</td>
<td>DataCite Metadata Schema</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>HOV/AUV Photos and Video</td>
<td>Image</td>
<td>2026-06-30</td>
<td>Restricted</td>
<td>Marine Geoscience Data System</td>
<td>25 TB</td>
<td>Creative Commons Attribution Non</td>
<td>None specified</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>rock, Water, Bio analytical data</td>
<td>Dataset</td>
<td>2027-06-30</td>
<td>Open</td>
<td>None specified</td>
<td>1 GB</td>
<td>None specified</td>
<td>DataCite Metadata Schema</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Water Column Plume models</td>
<td>Model representation</td>
<td>2027-06-30</td>
<td>Open</td>
<td>github Biological and Chemical Oceanography Data Management Office</td>
<td>200 MB</td>
<td>Creative Commons Attribution 4.0</td>
<td>None specified</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>