WOODS HOLE OCEANOGRAPHIC INSTITUTION

Developing a Particulate Sampling and In Situ Preservation System for High Spatial and Temporal Resolution Studies of Microbial and Biogeochemical Processes

28 September 2010

PI: John “Chip” Breier, Ph.D.
Applied Ocean Physics & Engineering
Woods Hole Oceanographic Institution
jbreier@whoi.edu

The remotely operated vehicle JASON samples fluids from the orifice of a “black smoker” hydrothermal vent (left) on the seafloor of Lau Basin in the southwestern Pacific Ocean (June, 2009, R/V Thomas Thompson) using the Suspended Particulate Rosette Sampling System (right).

Data Management Plan Philosophy

Principle Investigator **John "Chip" Breier** is strongly committed to the broad dissemination of research results and engineering developments. Within the scope of this proposal to the Gordon and Betty Moore Foundation, titled "Developing a Particulate Sampling and In Situ Preservation System for High Spatial and Temporal Resolution Studies of Microbial and Biogeochemical Processes", the research product will be a novel engineering development: a new tool for microbial sampling and in situ preservation. Thus the key issue for this data management plan is how to provide the broadest access to this engineering development both in knowledge and ultimately in terms of hardware (through funding mechanisms external to this award).

This data management plan follows the thought that there are two primary avenues for proliferation of and continued development of this technology. One is incremental development by other researchers. The other is small run production by a commercial manufacturer of oceanographic equipment. The former can be achieved by an uncontrolled release of engineering design information (i.e., drawings) to the general public (e.g., by placing in a digital repository). The latter typically involves a licensing agreement with a commercial company with the key purpose of limiting risk to the commercial supplier for taking on the manufacturing development of a new technology with an untested market. These two avenues of release have a number of possible permutations and implementation strategies – that can be incompatible if not carefully implemented. This data management plan considers the route of commercial production to be the one that will ultimately result in the most reliable equipment with the lowest cost to obtain, and as a result the maximum dissemination of this hardware to other researchers in the scientific community – not just a limited few that can obtain the resources to build one-off versions.

The potential point of conflict between these two dissemination processes is the issue of Intellectual Property – and specifically the potential issue of a future individual adopter seeking to commercialize a related design based either in whole or in part on the intellectual property developed in this proposal. Such an occurrence could complicate the issue of intellectual property pertaining to this design, to the point that a commercial company might deem commercial production to be too risky. This data management plan seeks to avoid this.

In addition, prior intellectual property rights are likely to be incorporated in the system developed by this proposal – in a controlled and legal manner – but the specific mixture of novel and established technologies will not be fully known until significantly into this proposal – after successful testing of a completed system. For example, it is very likely that the system developed by this proposal will share many features of the prototype system described by Breier et al. 2009. Thus the following data management plan seeks to maintain multiple options for achieving dissemination – options that can be selected based on the specific implementation of the engineering development.

J. A. Breier

Data Management Plan

This data management plan is based on the following 3 possible levels. As it is our goal to maximize the release of this engineering development, our desire and goal will be to implement all 3 levels. However it is unknown at this point what mixture of previous and new patented technologies will be included in the final package – it may be possible to implement all 3 levels regardless – but depending on the ultimate mixture of patents it is possible that levels 2 and 3 may be mutually exclusive. If so our goal will be to follow the plan most likely to achieve maximum proliferation of hardware to the scientific community. Thus we would implement just levels 1 and 2. If however the path to commercial production (level 2) is deemed unviable or starts but ultimately fails – then levels 1 and 3 would be implemented. In following this data management plan, the Woods Hole Oceanographic Institution's policy on intellectual property will be adhered to.

Level 1. The instrument developed by this proposal once tested and field proven will be reported on in a peer-reviewed journal (e.g., Deep Sea Research, or Ocean Engineering) and presented during a conference talk (e.g., OCEANS IEEE 2012). Related and additional information (i.e., related research, images, background) will also be added to the project website Breier is developing for this greater body of research. This website will have two fully public components with parallel content: one designed to describe this research and results to the public and the other (also open to the public) to facilitate dialogue amongst collaborators and the rest of the scientific community. This level will be initiated (paper submitted) within 6 months of a successful field trial using the newly developed equipment, regardless of the decision about which of levels 2 and 3 will be implemented

Level 2. A license, possibly a sole license, to commercially develop the design can be applied for - and handled by WHOI. This will be discussed with potential commercial partners once the design is final and tested successfully.

Level 3. Engineering drawings for unique components (not previously patented) that are developed through this grant will be provided to requestor's that sign an agreement not to commercially develop an instrument based on this technology.

Answers to GBMF questions:

1. **Data description**. Questions to consider as appropriate:

- What data will be collected during this project?

  Results of in situ preservation efficacy will be collected during the testing stage and reported on in a peer-reviewed journal article concerning the instrument and posted to the project website.

- How many different data formats are anticipated? Please list formats.

  In situ preservation efficacy, data tables.

  Engineering drawings. Adobe PDFs.

- When will the data be collected, when will they be entered into electronic databases, and what databases will harbor the data?

  Data results will be reported and archived primarily by way of peer-reviewed publication (e.g., Deep Sea Research)

- Does this project involve organization or analysis of pre-existing data, and what are the data sharing arrangements for these data?

  N/A

- What are the anticipated data products (e.g., databases, analyses, tools)?

  Engineering drawings in the form of Adobe PDFs.

- What kinds of metadata will be associated with the data?

  N/A

- Who is the owner of the data?

  Use of data reported in the peer-reviewed journal will be open to public use but the article distribution will be governed by the copyright of the journal.

2. **Data management**. Questions to consider as appropriate:

- Where (physically) will the data be stored?
  Data results will be reported and archived primarily by way of peer-reviewed publication (e.g., Deep Sea Research) and as such will reside in the publishers' servers and hard file copies. Related data will also be stored on WHOI servers and publicly disseminated via the world wide web (WWW)

- What type of data access or data distribution mechanism and software will be used?

  Scholarly reference search engines, WWW search engines, and WHOI's sitemap will provide links to the data.

- Will the location or software for initial data entry differ from the data archive?

  N/A

- How will metadata be stored, and what provisions will be made to enable metadata searching capability?

  N/A

- Who will be responsible for entering and maintaining data archives, and over what period of time will archives be maintained?

  The journal publisher will maintain the article. WHOI will host the secondary repository up to at least the duration of Breier's employment.

- What data quality controls and assurances will be provided?

  Peer-review of the journal article is the final control on quality and assurances.

- Who will contribute to the database?

  N/A

- Will proprietary data be used? If so, describe the permissions obtained to use the data.

In the engineering development, commercial subcomponents may be employed – the nature of their use will be based on agreements with the vendors.

3. **Data Sharing**. Questions to consider as appropriate:

- Who are the potential data users?

  Primarily academic researchers, potentially applied environmental researchers and managers.

- What is the appropriate timing for release of data to the public or relevant users, and why?

  Engineering development and testing is expected to take all of 2 years. Reporting in a peer-reviewed journal is expected during year 2. Navigating a path to commercialization will require discussions a commercial partner – as well as user demand – and can be expected to take longer than this 2 year grant.

- When will archived data be openly available to other users?
  The journal article will be submitted within 6 months of successful field testing with appropriate postings to the project website consistent following the peer-review process.

- If data from non-GBMF-supported or previous projects are integral to the successful completion of the Grant Purposes, will the non-GBMF-supported and/or pre-existing data also be made freely available?

  Data: N/A

  Release of previous engineering developments will follow laws applicable to the owners of the IP and the data management plan.

- How will other users (i.e., beyond the grantee and GBMF) access data and metadata?

  See previous answers.

- Are the publicly available data in raw form? If not, what treatments have been applied to the data prior to their being released to the public?

  N/A

- How long beyond the grant term will the data be maintained and by whom?

  See previous answers.

- Does the proposed grant include provisions for future hardware upgrades in the event that data is to be stored and maintained well beyond the project period of the Grant?

  No but this is not expected to be an issue.

- If data analysis tools are to be created as a consequence of the Grant, will a tutorial be available for training of future users of the data, and if so, how can it be accessed?

  N/A

- Will a data sharing agreement be required between outside vendors? If so, a brief description of the agreement needs to be provided in the grant proposal.

  N/A

- Is a Creative Commons type-license appropriate for sharing the data? Why or why not?

N/A

- How will appropriate attribution to the data provider be provided?

  N/A

- Do you anticipate publishing a “Data Release Paper” for referencing and sharing the data?

  See previous answers concerning peer-reviewed paper

DATA SHARING PHILOSOPHY

DATA SHARING PHILOSOPHY

The Gordon and Betty Moore Foundation's goals of scientific advancement, environmental conservation and health care improvement will best be served through a culture of open access to data. It is our philosophy that:

- All data used in or developed in whole or in part by foundation-funded projects (and that can be shared in a manner consistent with applicable laws) will be made widely available and freely shared as soon as possible[1]. If data used in foundation-funded projects are owned by an additional party other than the grantee, we do not require it to be released, but the grantee will use its best efforts to encourage the data owners to make it openly and freely available.
- Data are shared with full and proper attribution to the data provider.
- Data developed in whole or in part by foundation grant funding are the property of the grantee unless otherwise specified. The grantee may protect its property through patent, copyright and/or other intellectual property protection instruments, except that it may not impede the effective access and use of the data by the public.
- The foundation is not responsible for any liabilities associated with errors in the data or misrepresentations or misinterpretations of publicly available data.
- The foundation supports grant funding for costs associated with data sharing and open access publication of scientific findings, where appropriate.
- The foundation and prospective grantees will jointly develop a Data Management and Sharing Plan prior to the finalization of a grant agreement.

The Data Sharing Philosophy applies to all activities that are financially supported in whole or in part by the foundation that include, but are not limited to:

> Data collection and analyses, data, meta-analyses and information derived from pre-existing datasets, and database development

Data sharing includes, but is not limited to, data contained within the following:

> Publications, databases, derived data products, mathematical models and model code, metadata (defined as appropriate documentation describing the data, relevant specifics of their collection and the data format) and statistical and other forms of data reduction and analysis

[1] Examples of when data should be released: For data created for scientific and environmental conservation purposes, public release should occur not more than six months from the "date of collection" (defined as the date when data enters an electronic database), unless otherwise specified in the grant agreement between the grantee and the foundation; for DNA sequence data, "public release" (i.e. submission to an appropriate public database), should occur not more than six months after "completion" of the DNA sequence determination (as defined in the grant agreement between the grantee and the foundation).

DATA SHARING AND MANAGEMENT PLAN

As part of the foundation grant development process, potential grantees are required to develop a Data Management and Sharing Plan with their foundation grant team. In these cases, before submitting a final grant proposal to the foundation for approval, both the potential grantee and the foundation grant team must approve a final version of the plan that is consistent with this Data Sharing Philosophy. Any exceptions to the Data Sharing Philosophy must be clearly articulated in the plan and approved by the grant team. Funds needed for data sharing and management may be requested as part of the proposal. Once finalized, the plan will be referenced in the grant agreement for the approved grant.

The plan should address the following three topics and any other topics identified by the foundation and/or grantee:

1. **Data description**. Questions to consider as appropriate:

   - What data will be collected during this project?
   - How many different data formats are anticipated? Please list formats.
   - When will the data be collected, when will they be entered into electronic databases and what databases will harbor the data?
   - Does this project involve organization or analysis of pre-existing data, and what are the data sharing arrangements for these data?
   - What are the anticipated data products (e.g., databases, analyses, tools)?
   - What kinds of metadata will be associated with the data?
   - Who is the owner of the data?

2. **Data management**. Questions to consider as appropriate:

   - Where (physically) will the data be stored?
   - What type of data access or data distribution mechanism and software will be used?
   - Will the location or software for initial data entry differ from the data archive?
   - How will metadata be stored, and what provisions will be made to enable metadata searching capability?
   - Who will be responsible for entering and maintaining data archives, and over what period of time will archives be maintained?
   - What data quality controls and assurances will be provided?
   - Who will contribute to the database?
   - Will proprietary data be used? If so, describe the permissions obtained to use the data.

3. **Data Sharing**. Questions to consider as appropriate:

   - Who are the potential data users?
   - What is the appropriate timing for release of data to the public or relevant users, and why?
   - When will archived data be openly available to other users?
   - If data from non-foundation-supported or previous projects are integral to the successful completion of the Grant Purposes, will the non-foundation-supported and/or pre-existing data also be made freely available?
   - How will other users (i.e., beyond the grantee and the foundation) access data and metadata?
   - Are the publicly available data in raw form? If not, what treatments have been

applied to the data prior to their being released to the public?
- How long beyond the grant term will the data be maintained and by whom?
- Does the proposed grant include provisions for future hardware upgrades in the event that data is to be stored and maintained well beyond the project period of the grant?
- If data analysis tools are to be created as a consequence of the grant, will a tutorial be available for training of future users of the data, and if so, how can it be accessed?
- Will a data sharing agreement be required between outside vendors? If so, a brief description of the agreement needs to be provided in the grant proposal.
- Is a Creative Commons type-license appropriate for sharing the data? Why or why not?
- How will appropriate attribution to the data provider be provided?
- Do you anticipate publishing a “Data Release Paper” for referencing and sharing the data?