

## PLE Data Management Plan

The data that we scientists collect represents small pieces of a larger collective effort to better understand the natural world. As such, it is important that scientists organize and archive past, present, and future data for other scientists to build upon with full acknowledgment of the original efforts of investigators. Not only does this benefit the community of scientists, it can also be self-serving because clear recollection of an investigator's past studies can fade with time. This is the idea underlying a *Data Management Plan* (DMP). An effective DMP includes protocols for recording raw data, maintaining data, analyzing data, and disseminating data (in the form of publications and the archiving of raw data and metadata).

Most research groups already conduct many of the steps in a typical DMP, although the protocols tend to be informal and not uniform among all members of the research group. Major funding agencies including NSF are recognizing the value of formalizing this process and are now requiring a DMP for all proposals. In addition, many journals now require that raw data and metadata accompany publications. Having a detailed DMP will facilitate transfer of scientific data and ensure compliance with data management policies set by funding agencies and journals. Moreover, NSF Field Station and Marine Lab (FSML) grants require that field stations provide transparent access to data produced by their scientists. A concerted effort to manage data produced by PLE researchers will put the field station in a competitive position for grants aimed at renovating and expanding facilities for research, teaching and outreach.

As a field station, PLE has two types of data to manage. The first type of data are those related to the operation of the field station including climate data, field site use data, publications, user data (enrollments, research user days, outreach user days, and conference user days). These data will be managed and archived by the PLE Director and Administrative Secretary.

The second type of data is produced while conducting research at PLE. PLE investigators that performed the research are in the best position to organize and archive these data. To assist investigators in developing their own DMP for their own group and for upcoming grant proposals, we outline some of the most important aspects to consider (see [EXAMPLE DMP](#)). We also encourage researchers to check out DataOne's website (<https://www.dataone.org/>). DataOne provides a useful search engine of data management tools including interactive software that drafts a DMP based on user input.

As part of the PLE DMP, each investigator agrees to archive their data and metadata in a publicly available repository (e.g., [Dryad](#), [GenBank](#), [TreeBASE](#), and [NCEAS Data Repository](#)) or as a supplement to manuscripts. Once archived, PLE requests that researchers send PLE administrators ([pymilab@pitt.edu](mailto:pymilab@pitt.edu)) links to the archived data. These links will be placed at the end of each publication's entry on PLE's "Publications" webpage (<http://www.biology.pitt.edu/facilities/pymatuning-1>).

## **Example Data Management Plan for Researchers**

Every research group has protocols for conducting research and the following items are offered as suggestions to consider when conducting research as part of PLE's DMP. The goal of the DMP is to emphasize reproducibility, transparency, and accountability in every aspect of the research. In accordance with University of Pittsburgh policy, all members involved with this research will undergo, or have already undergone online training in research ethics. Here is an example DMP:

### **Recording and maintaining data**

1. Research activity and data collections will be kept in a single, waterproof laboratory notebook. Entries will consist of experimental design, calculations, data records, experimental notes, and summaries of discussions concerning the research. All entries will be legible and clearly marked with date of entry. Undergraduate assistant are allowed to record in this notebook under strict monitoring by the PIs and co-PIs for clarity, legibility, and accuracy. No revision of entries is permitted, although future entries may reference mistakes in any past entries with clear explanation for the reason of the mistake. Backup photocopies of the notebook entries should be made on a regular basis in accordance with lab protocol.
2. Frequently, raw data measurements will be taken in the field on waterproof paper. Such datasheets will be kept in a binder designated for experimental measurements. Datasheets will be referenced in the lab notebook whenever taken. Back-up photocopies of these datasheets will be made on a weekly basis in accordance with lab protocol.
3. Within a reasonable amount of time, all raw data will be added to a single Microsoft Excel worksheet with clear identification regarding the type of data and when the data was taken. Digital backups will be made of this data on a regular basis via an external hard drive. Prior to being archived, digital and hard copies of all data will be held in at least two different locations for security and redundancy.
4. In agreement with governmental regulations, the University of Pittsburgh requires that research notebooks, all datasheets, and all back-up copies will be retained for a minimum of seven years after final reporting or publication.

### **Policies regarding rights and obligations of involved parties**

1. Full credit will be given to all those who conducted research by inclusion in authorship on published papers. Order of authorship will be determined by the accepted standard of placing those who contributed the most substantially to the work before those who played less prominent roles.
2. All notebooks, datasheets, preserved specimens, and digital data will be kept in the laboratory where it originated. Copies of intangible data may be taken with the PI or co-PI should either one leave his appointment at the University of Pittsburgh. Tangible material may be transferred to another institution, provided that sufficient samples exist, and if the samples can be easily split at minimal cost. In all cases, the transfer shall be subject to the terms of a materials-transfer agreement negotiated by the Office of Research and the recipient institution. These rights to access data also apply to trainees and students who are an integral part of the research project.

### **Notification of publication**

1. When a publication results from a research project, the PI should contact the PLE Administrative Secretary ([pymlab@pitt.edu](mailto:pymlab@pitt.edu)) and include a pdf copy of the publication.
2. All published papers resulting from this research should also be listed on the websites of the PI.

### **Archiving data**

Any set of archived data includes both primary data and metadata, following the standards recommended by Michener et al. (1997). Thus, the data for each experiment will include the

following:

1. Data set descriptors (i.e. abstract, key words)
2. Research origin descriptors (i.e. PIs, funding, objectives, methods, experimental design)
3. Data set status and accessibility (i.e. date of last modifications, data use restrictions)
4. Data structure (i.e. file type, description of response variables, data type)
5. Supplemental descriptors (i.e. data forms, quality control procedures, publications)
6. Location(s) of physical specimens (preserved organisms, DNA, etc.)

Within one year of publication, PIs are expected to archive all data and metadata in a publicly available repository (e.g., [Dryad](#), [GenBank](#), [TreeBASE](#), and [NCEAS Data Repository](#)) or as a supplement to the manuscript. Once archived, PLE requests that researchers send the PLE [administrative secretary](#) links to the archived data. These links will be placed at the end of each publication's entry on PLE's "[Publications](#)" webpage.

As with all data archiving, consideration must be given to the dissemination of any confidential or private data.