Data Management Plan

Data policy compliance – The project investigators will comply with the data management and dissemination policies described in the NSF Award and Administration Guide (AAG, Chapter VI.D.4) and the NSF Division of Ocean Sciences Sample and Data Policy.

Description of data types – The proposed project is expected to produce several observational and experimental datasets and physical specimens, described in the list below. In addition to the materials described below, the project will produce an open-access learning module to be developed in collaboration with the Network of Conservation Educators and Practitioners (see *Project Description*), which will be available through the website of the American Museum of Natural History.

Observational datasets
- **Coral reef fish abundance from visual surveys**: Counts, body size (total length), and species ID for fishes on randomly placed transects on the forereef of each island for 18 islands. Meta-data include date, time, coordinates, depth, habitat type, current, and conditions. Recorded by divers on waterproof paper datasheets. File type: entered data in .csv, raw datasheets as scanned pdfs. Repository: BCO-DMO
- **Coral reef fish biodiversity calculated from visual survey data**: Species richness and Shannon diversity at each surveyed site on 18 islands. File type: .csv. Repository: BCO-DMO
- **Individual-level data on collected coral reef fishes**: For each fish collected, we will record collection location and time, body size measurements (TL, SL, FL, mass), reproductive status, stomach mass and contents, and liver mass. Will be recorded by dissectors on paper datasheets. File type: entered data .csv, raw datasheets as scanned pdfs. Repository: BCO-DMO
- **Images of each coral reef fish**: Photograph taken of the right side of each fish after freezing and thawing and immediately prior to dissection, to document species identification. File type: .jpeg. Repository: BCO-DMO
- **Parasite abundance in coral reef fishes**: For each fish, the number and species identity of all metazoan parasites. Will be recorded by dissectors on paper datasheets. File type: entered data .csv, raw datasheets as scanned pdfs. Repository: BCO-DMO
- **Images of parasites from coral reef fish hosts**: For a subset of parasites detected, photographs will be taken to aid in identification and expansion of our parasite species ID guide. File type: .jpeg. Repository: BCO-DMO

Experimental datasets
- **Genetic voucher specimens of parasites from coral reef fish hosts**: At least 50 individual parasites per putative parasite taxon at each of 18 islands, in 1.5-mL microcentrifuge tubes with 95% ethanol, labeled with a paper tag inside the vial and sample number in permanent ink on the outside of the vial. Twenty-five of these will be sequenced (see below), but the remaining vouchers will be catalogued in the Smithsonian Invertebrate Zoology Collection and made available for lending (see below). File types: specimens in 95% ethanol. Repository: meta-data and access information to be provided to BCO-DMO.
- **Genetic data from parasites**: DNA sequences from 25 individual parasites in each of six common parasite taxa at each of 18 islands. File type: .fasta files. Repository: NCBI; accession numbers to be provided to BCO-DMO.
- **Genetic voucher specimens of coral reef fish hosts**: A plug of muscle tissue from every fish collected, in 1.5-mL microcentrifuge tubes with 95% ethanol, labeled with a paper tag inside the vial and sample number in permanent ink on the outside of the vial. Will be retained in PI Wood's lab and made available to those who request them. File types: specimens in 95% ethanol. Repository: meta-data and access information to be provided to BCO-DMO.
- **Stable isotope voucher specimens of coral reef fish hosts**: A plug of muscle tissue from every fish collected, frozen in 1.5-mL microcentrifuge tubes, labeled with a paper tag inside the vial and sample number in permanent ink on the outside of the vial. Will be retained in PI Wood's lab and made available to those who request them. File types: frozen voucher specimens. Repository: meta-data and access information to be provided to BCO-DMO.

Data and metadata formats and standards – Observational data will be stored in .csv files, which can be easily read by different software packages. Quality flags will be assigned according to the ODS IODE Quality Flag scheme (IOC Manuals and Guides, 54, volume 3; http://www.iode.org/mg54_3). Metadata will be prepared in accordance with BCO-DMO conventions (i.e., using the BCO-DMO metadata forms) and will include detailed descriptions of collection and analysis procedures. Meta-data on voucher specimens will also be stored in .csv files, to allow researchers to query our voucher database and easily identify samples of interest. Meta-data will include samples that were already collected (6 islands) in addition to the samples we propose to collect in this grant application (12 islands).

Data storage and access during the project – PI Wood will be responsible for collecting and entering parasite data and meta-data and all vouchers, co-PI Sandin will be responsible for collecting and entering fish data and meta-data, and Collaborator Haupt will be responsible for collecting and entering molecular data. PI Wood will be responsible for storing and backing up these datasets. We will store all project data on dual Synology Diskstation DS1517+ network-attached storage (NAS) servers, each with five redundant 4-TB drives in a single 5-bay rack, located in two separate buildings on the University of Washington campus. NAS system 1 will serve mapped volumes mutually accessible among the Wood, Sandin, and Haupt labs, with each lab possessing read-only privileges in the other labs' sub-folders, and editing privileges in their own sub-folder. NAS system 1 will be stored in a dedicated server space in UW's Fisheries Sciences Building. NAS system 2 will operate as a back-up for NAS system 1, and will be stored in a dedicated server space in UW's Marine Sciences Building. Both systems will be backed up to the cloud using Microsoft Azure and will have full IT support and monitoring from IT services at UW's School of Aquatic and Fishery Sciences.

Mechanisms and policies for access, sharing, re-use, and re-distribution – Data produced by this project may be of interest to biological oceanographers, parasite ecologists, and coral reef scientists. Within 2 years from the date of collection, all data and metadata will be submitted to BCO-DMO, DNA sequences will be deposited in the National Center for Biotechnology Information (NCBI) database GenBank, and all vouchers will be submitted to the Smithsonian Institution (see below). GenBank accession numbers will be provided to BCO-DMO in a .csv file and metadata will be provided using the BCO-DMO Dataset Metadata submission form. The project investigators will work with BCO-DMO data managers to make project data available online in compliance with the NSF OCE Sample and Data Policy. Data, samples, and other information collected under this project will be made publicly available without restriction once submitted to the public repositories. We will adhere to and promote the standards, policies, and provisions for data and metadata submission, access, re-use, distribution, and ownership as prescribed by the BCO-DMO Terms of Use.

Parasite vouchers will be catalogued in the Smithsonian Invertebrate Zoology Collection, which is home to the US National Parasite Collection. The Collections Committee of the Smithsonian Institution's Invertebrate Zoology Collection charges for cataloging new specimens to cover the costs of materials (e.g., jars, labels), storage space, and curatorial labor. Smithsonian has estimated a cost of $5,500 for cataloging the material generated by this project, which we have planned for (see *Budget*). Once catalogued, any individual can request these samples, with the agreement that destructive use is at the discretion of Smithsonian curators, any unused portion of the specimen will be returned, that the PIs will be included as authors on any papers that result, and that NSF is acknowledged as the funding source.

Plans for archiving – The PI will work with BCO-DMO to ensure data are archived appropriately in national data archives and that proper and complete documentation are archived along with the data. All voucher specimens will be registered with the System for Earth Sample Registration SESAR and given a persistent identifier (International Geo Sample Number) and catalogued in the Smithsonian Invertebrate Zoology Collection.

Roles and responsibilities – Each PI will be responsible for sharing his/her subset of data among the project participants. PI Wood will be responsible for collecting and analyzing the parasite sampling data. Co-PI Sandin will oversee the fish survey and collection work. Collaborator Haupt will oversee the molecular work and will submit the resulting sequences to GenBank. PI Wood will document, maintain, and catalogue all voucher specimens. She will also coordinate the overall data management and sharing process and will submit the project data and metadata to BCO-DMO, who will be responsible for forwarding these data and metadata to the appropriate national archives.