UMBS Data Stewardship Policy

1. INTRODUCTION

The University of Michigan Biological Station (UMBS) is committed to building a collaborative and data rich community, as well as a transparent, inclusive, and adaptable data stewardship policy. We know that the nature of data and how it is curated will change over time, thus this is a 'living' document that can change alongside data and scientific needs. One of a field station's many assets is its long-term data, and this accumulated knowledge leads to a rich data environment that the science community can build upon. UMBS's goal is to preserve long-term and emerging datasets that will be vital to basic and applied scientists in the future.

This document outlines the UMBS data stewardship policy. Data hosted by UMBS include data collected through research and teaching activities based at UMBS and other data that advance the mission of UMBS.

2. UMBS DATA STEWARDSHIP GOALS

Goals of the UMBS Data Stewardship Policy include the following:

- The long term preservation and discoverability of data by adhering to the FAIR principles[5] for scientific data management and stewardship (described in Appendix 1)

- Ensuring the long-term (>50 years) value and viability of data sets collected at the Biological Station through proper metadata documentation and data archiving

- Defining how data can be shared and used through use of licensing

- Facilitating access to Biological Station-related data sets and project-related information (e.g., project descriptions, research sites, bibliography) in order to create opportunities for the development of unique research questions and collaborations that will further advance environmental research and education

- Defining roles and responsibilities for researchers, faculty, staff, and students participating in research at and sharing data with UMBS

3. DATA MANAGEMENT OVERVIEW

3.1. UMBS Data Management Services and Expectations

UMBS offers and encourages the service of data management assistance to all researchers who conduct research at UMBS. Data services at UMBS may include the following:

- Data Management Planning - UMBS data management staff will provide guidance and/or assistance with creating data management plans for proposals and projects based at UMBS

- Data Curation - UMBS data management staff will provide training and guidance to researchers and students on the best practices for organizing, cleaning, and documenting data collected at or in coordination with UMBS

- Metadata & [Ecological Metadata Language (EML)](#)[6] - UMBS will provide documentation and templates (web based and downloadable) for completing EML based metadata to be stored in the UMBS data repository and made discoverable through publication of data and metadata

- Data Hosting - UMBS will provide a local repository for public and private data sets and related research products

- Data Publication - UMBS will leverage a public data repository to serve up long-term and high impact data sets collected at UMBS or by affiliated researchers. Working data and short-term data will be published via a locally run repository or in conjunction with the University of Michigan Library.

In exchange for data management services, UMBS maintains the following expectations of data originators:

- Submission of a new or revised research project proposal on a three (3) year rotational basis (PIs) or annually (students)

- Submission of a data management plan for new research projects

- Adherence to standards of clean and well documented data (See Appendix 2)

- Deposition of data sets in UMBS data repository or an approved data repository

- Sharing of other research products including, but not limited to, publications, reports, and presentations

- Adherence to the FAIR principles[5] (see Appendix 1)

- If researchers or students receive support (e.g.- scholarships or housing), data must be accurately archived to receive future support.

3.2. UMBS Data Sources

Data managed by UMBS can be generally categorized by data source as follows:

- Station Data: Data systematically collected or produced as part of baseline monitoring conducted at UMBS, by site based staff researchers.

- Funded Data: Data collected or produced by funded projects at UMBS.

- Publications: Journal publications, dissertations and theses of research occurring at UMBS or supported by UMBS resources

- Class data: Data collected as part of class activities or class projects

- Student papers: Student reports of undergraduate research projects

- External Data: Data from external repositories or data providers, including existing operational data streams and historical sources, industry, international institutions, or others, as relevant.

- Rescued Data: Data retrieved from unpublished sources, e.g., field notebooks, records on outdated storage media, or photographic records, which are often at risk of loss.

- Local Knowledge (LK): is the knowledge that people in a given community have developed over time and continue to develop.

3.3. Data Sharing

UMBS encourages public data sharing, but hosts both publicly available and private data. UMBS acknowledges that the funding origin predominantly dictates the expectations for data sharing. UMBS also recognizes justification for private

or restricted data sets; however, if data were collected with support of UMBS funds (e.g., scholarships & housing support), those data must be made publicly available. Data hosted by UMBS can be shared in multiple ways. UMBS categorizes data into two sharing types: Data Type I (publicly available data) and Data Type II (restricted data). Distribution and use of data hosted by UMBS is subject to conditions designated by the license assigned to the data set.

Data designated as Sharing Type I and carrying a license designation can be downloaded without restriction or authentication, but users are expected to read licensing and acknowledge Data Originators as specified by the license. Data designated as Sharing Type II will be hosted within the UMBS data repository and access will be restricted as specified by the Data Originator, funder, or UMBS. Metadata for Sharing Type II will be shared publicly and potential Data Users will be provided a mechanism to contact the Data Originator. UMBS will prioritize assisting with datasets and dataset curation for datasets that can be used openly by the community.

After UMBS receives data designated as Sharing Type I, UMBS will submit the data to the [Environmental Data Initiative (EDI)](), an NSF supported data archiving and dissemination service, where a Digital Object Identifier (DOI) will be assigned, version and provenance tracking enabled, and generation of code for automated access to the data in the pursuit of reproducible science.

4. ROLES AND RESPONSIBILITIES

4.1. UMBS Data Management Staff

UMBS data management staff and researchers will work in collaboration to implement data management best practices and meet funding agency/institutional requirements. UMBS data management staff will provide researchers templates for a data management plan & metadata. Additionally, UMBS data management staff will provide data management assistance to researchers to provide efficient and accurate collection of data and subsequent archival and sharing of research data products.

4.2. Data Originators

The data originator will submit a properly documented copy of any data set collected using UMBS resources or within the UMBS footprint to the data repository **no later than two (2) years following the completion of that data set.** UMBS recognizes that a data set may require multiple years to complete. At the time of submission, a data set will be classified by the data originator as Data

Type I or II. If data are classified as Data Type II, the data set resource files, not the metadata, will be made private within the data repository for a time period to be determined by the data originator not to exceed five (5) years. At the end of this time period the data set will become publicly available unless an exemption is made by UMBS data management staff. Data protected under federal or state regulations or due to proprietary, ethical, or privacy considerations will typically be classified as Data Type II until those statutes or considerations no longer exist. UMBS will release Data Type II data publically only with the data originator's consent. Data originators with data sets that require more than three (3) years to complete are strongly encouraged to archive an intermediate version of their data set within the UMBS data repository on an annual basis.

Failure of the Data Originator to submit an appropriately documented data set within the specified timeframe can result in the denial of future use of Biological Station resources.

All Data Sets submitted to UMBS must be accompanied by metadata of sufficient quality to allow data users not involved in the formation of the data set to use it for scientific purposes. The Information Manager will guide the development of data set metadata, and metadata will adhere to the EML. All metadata is considered public (Data Type I) unless it contains information protected under federal or state regulations or due to proprietary, ethical, or privacy considerations, in which case it will be considered Data Type II.

4.3. Data Users

The sharing and use of data sets made available by UMBS are subject to the licenses assigned to the dataset. Data users are expected to read and adhere to the assigned license. In the event that license is not yet assigned to a shared data set by the data originator the UMBS data use agreement (Appendix 2) is to be applied.

5. PROCESSES AND PROCEDURES

5.1. Researchers

All those conducting research at UMBS are encouraged to submit accurate, complete and timely project and research product metadata through UMBS's information management system. To be eligible for future support, researchers (PI's, Co-PI's, PostDocs, Grad students) who receive support from UMBS are required to submit clean data and complete metadata to the UMBS data repository or another repository approved by the UMBS data manager.

Submission of initial project metadata and intermediate research products/data sets are expected before acceptance of the subsequent year's research and housing application. Final research products and metadata are expected to be completed within two (2) years, or a time period otherwise specified by the researcher's funding agency, of the completion of data collection.

The process of data submission and review for researchers is integrated with the annual UMBS research and housing application system and is generalized in Figure 1 and described below.
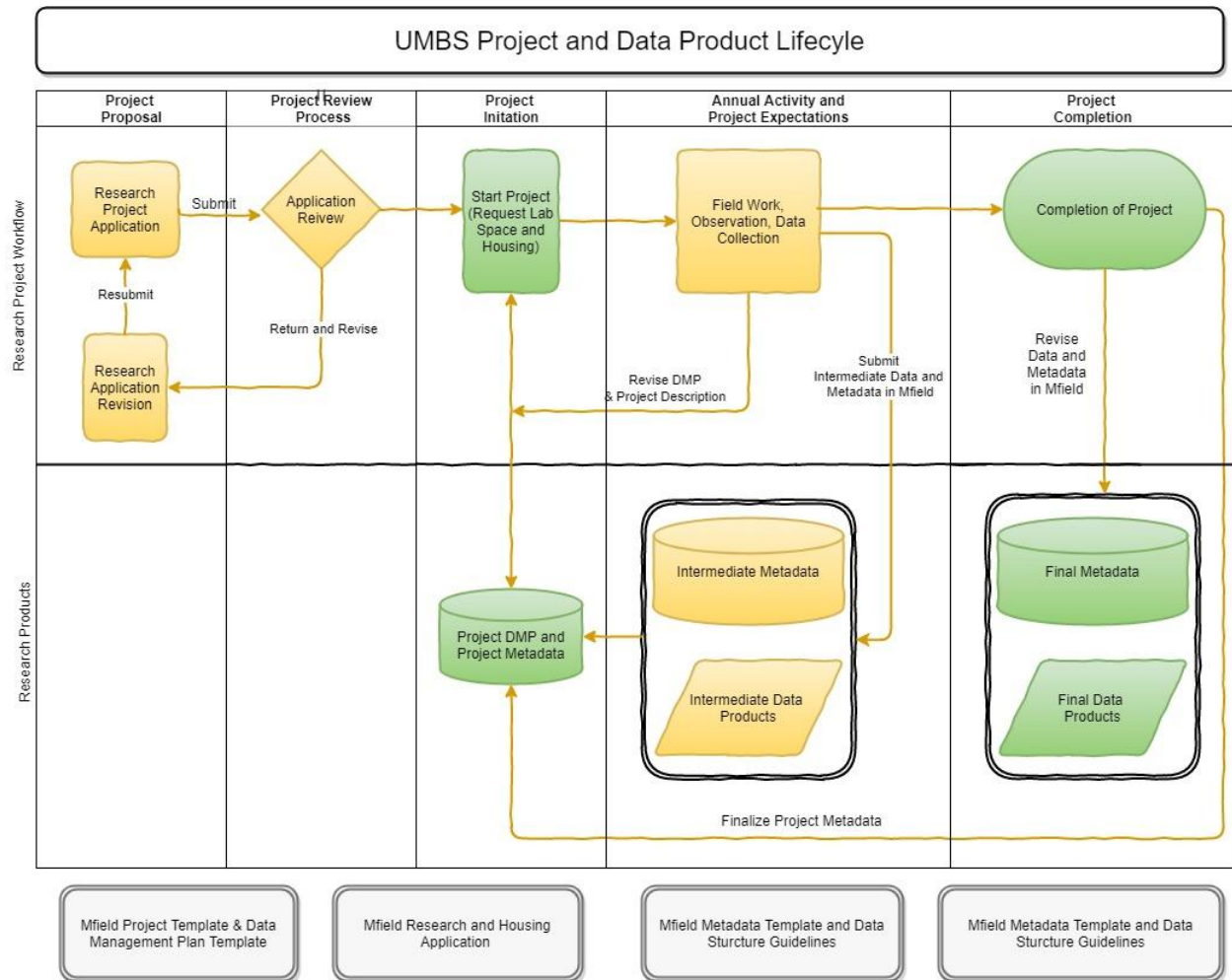
5.1.1. Project Proposal

UMBS provides researchers with a project proposal form and a data management plan (DMP) template. Research proposal with DMP is to be submitted by March 1st of the year the proposed project is to start. If a

researcher is proposing to conduct field work on UMBS property, the researcher must report proposed locations of any installation of equipment and manipulations beyond pure field observation. If a researcher or research team is proposing a large manipulation or construction of semi-permanent infrastructure additional time and project review will be needed by UMBS staff and research project review committee.

This information will be used to identify potential conflicts with ongoing research programs, space constraints, management issues and assess UMBS's ability to support the project. Research project review and approval will happen following the March 1st proposal deadline and during the summer session for projects submitted after the deadline, with limited exceptions. UMBS staff and research project review committee will provide feedback and/or approval by April 1st and September 1st for projects reviewed in summer. After an approved project has been running for three (3) years, UMBS requests a revised project proposal which is subject to review by UMBS staff.

5.1.2. Project Initiation

With project approval granted, researchers will coordinate with UMBS staff their specific needs while conducting research at UMBS by submitting a Research and Housing application. Research and Housing requests should specify environmental health and safety considerations (e.g., chemical use, hazardous waste, biological waste), resources needed (lab space, stockroom materials, tools, etc.) and dates desired to be at UMBS. If this information is not provided by the March 1st deadline, UMBS can not guarantee space or resources even for approved projects.

5.1.3. Project Updates

Prior to submitting a research and housing application for a new field season, researchers will need to submit or revise the previous year's intermediate data products/metadata. This provides researchers with an additional data back-up and provides UMBS data management staff access to the data and metadata well in advance of publication. This lead time allows researchers and data management staff to create the best possible research products.

5.1.4. Project Completion

Yearly and at the completion of a research project researchers are required to provide complete project metadata. Additionally, they are expected to provide final versions of data collected and final, complete metadata. If other research products are produced, they are expected to be submitted and or documented with UMBS data management staff.

5.1.5. Publications and Reports

An electronic copy (e.g., pdf or doc file format) of all reports, peer-reviewed papers or final manuscripts derived from data sets collected using Biological Station resources or within the UMBS research footprint/property must be submitted to the Data Manager at the time of completion. UMBS will maintain archival records and will not distribute, further copy or provide access to the publications except to authorized users of the UMBS in facilities or via computers owned and networked by the University of Michigan.

5.2. Faculty

It is a goal to help facilitate the archival of data sets produced by faculty and students participating in UMBS courses as this has been a rich source of observational data historically. Course exercise data are data sets derived during class exercises such as methods demonstrations.

UMBS recognizes that the brief time allotted for the development of course project data and course exercise data introduces a high degree of variability in product quality. To account for this, faculty instructors are asked to adhere to the same standards as UMBS researchers (DMP, data collection hygiene, metadata, etc) and submit to the archive. Course data collected by students will be categorized in the data repository at student collected data. In exchange UMBS data management staff will provide training and support, if desired, to students during instructional time.

5.3. Students

All undergraduate student researchers must submit copies of their final research papers to the UMBS Student Paper Archive in Deep Blue. Students can submit final drafts of research papers and associated metadata to the UMBS Student Paper Archive via a Google form. All data collected by students, per SPG 303.04[3], is their intellectual property and should be treated as such, even if

collected as part of a class exercise or class project. All major contributors should be listed as authors of course collected data sets that are archived.

5.4. Others

5.4.1. Collected Data

Data collected for research projects taking place on UMBS grounds or coordinated from the field station are requested to be deposited in the UMBS data repository or associated data infrastructure. Collected data can be stored in external repositories if the data originator can demonstrate that the repository is sufficiently supported to meet UMBS's stated data stewardship goals. If data is stored in an external repository, then metadata and a persistent DOI must be provided.

5.4.2. Physical Samples and Specimens

Physical samples or collections housed at UMBS, whether collected on UMBS grounds or offsite, are expected to be documented in the UMBS data repository or associated data infrastructure.

Physical samples or collections that are housed offsite AND were collected on or surrounding UMBS property are expected to be documented in the UMBS data repository or associated data infrastructure.

6. Disclaimer

While UMBS strives to provide data of the highest quality, all data secured from UMBS is provided "as is." UMBS and the Data Originator are not responsible for errors in or conclusions drawn from the use of Data Sets furnished by UMBS.

7. References

1. University of Michigan Data Administration Guidelines for Institutional Data Resources. Accessed online at https://drive.google.com/file/d/0B7RLNc7LWQbIYlhPb0x5RmVNOWc/view on November 5th, 2020.

2. University of Michigan Data Standard Practice Guide. Institutional Data Resource Management Policy. Number 601.12. Accessed online at https://spg.umich.edu/policy/601.12 on November 5th, 2020.

3. University of Michigan Data Standard Practice Guide. University of Michigan Technology Transfer Policy. Accessd online at https://spg.umich.edu/policy/303.04 on January 24the, 2021.

4. LTER Network Data Access Policy, Data Access Requirements, and General Data Use Agreement. Accessed online at https://lternet.edu/data-access-policy/ on November 5th, 2020.

5. Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

6. Jones, M.B., M. O'Brien, B. Mecum, C. Boettiger, M. Schildhauer, M. Maier, T. Whiteaker, S. Earl, and S. Chong. 2019. Ecological metadata language version 2.2.0. KNB Data Repository. DOI=10.5063/f11834t2

8. Revisions

This document is a major revision for the 2010 UMBS Data Management Policy.

Data types

Licensing

Roles and Responsibility

APPENDIX 1:

FAIR data

F: Findability - The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.[4]

A: Accessibility - Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.[4]

I: Interoperability - Data usually needs to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.[4]

R: Reusable - The ultimate goal of FAIRE is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.[4]

Plus

E: Ethical - "Data are ethically managed by respecting legal and ethical obligations, including consent, privacy, and confidentiality; secondary use of data; and data linkage"

APPENDIX 2:

1. **Organize data into logical units based on structure and theme. See https://environmentaldatainitiative.org/five-phases-of-data-publishing/phase-1/**
2. **Clean and format data and apply quality control (QC) checks. See https://environmentaldatainitiative.org/five-phases-of-data-publishing/phase-2/**
3. **Describe data in the Ecological Metadata Language (EML) standard. See https://environmentaldatainitiative.org/five-phases-of-data-publishing/phase-3/**

APPENDIX 3:

Data Use Agreement (DUA) for Data Type II

The use of Type II data sets made available by UMBS is subject to the following restrictions and qualifications:

1. The Data User will acknowledge the Data Originator and UMBS in any publications, reports, or presentations that use data falling under the auspices of UMBS. Where such products result from the use of data secured through UMBS, the data user is strongly urged to consider collaboration and/or co-authorship with the data originator as appropriate in conformance with common academic

practices for attribution of authorship and recognition of contribution. Data sets should be cited in the general form: Data Originator, Year Published. Data Set title. Publisher: Data Set Identifier. Date Accessed. For example:

Miller, D., 2010. Water profiles of the Southfishtail Bay Depression in Douglas Lake, Michigan, 1933-1934. Pellston, MI: Mfield Data Repository and Information Management System: URL. Accessed on 2018-05-14.

2. The Data User will provide a PDF version of all published papers and reports derived from data sets obtained via UMBS to the Data Originator and to the UMBS Data Manager and Resident Biologist within 6 months of publication.
3. The Data User agrees not to disseminate or re-distribute data supplied by the UMBS beyond the immediate collaboration or in accordance with the specified license.
4. the licensing applied to the specific Data Set.
5. The Data User is fully responsible for all errors in analysis and judgment that are derived by the data user from data sets made available by UMBS.

GLOSSARY:

Data Category or Source: data origin that dictates role and responsibility of Data Originator and Data User

Data Originator - individual or institution that produced the Data Set.

Data Set – Digital data and its metadata derived from any research activity such as field observations, collections, laboratory analysis, experiments, or the post-processing of existing data and identified by a unique identifier issued by a recognized cataloging authority such as a site, university, agency, or other organization.

Data Structure - Stand alone tabular data or data stored in relational table database formats. GIS data in shapefile or geodatabase, documents in PDF, and images in jpeg or tif format.

Data User - individual to whom access has been granted to this Data Set, including their immediate collaboration sphere, defined here as the

institutions, partners, students and staff with whom the Data User collaborates, and with whom access must be granted, in order to fulfill the Data User's intended use of the Data Set.

Ecological Metadata Language (EML) - a metadata standard used at UMBS, EML defines a comprehensive vocabulary and a readable XML markup syntax for documenting research data.

Metadata: Metadata provides the information about a dataset, specifically the what, where, how, when, by whom it was collected, its current location, and any access information. Metadata facilitates the understanding, use, and management of data and is a means for networking and collaboration. Standardized metadata records consist of a defined set of information fields that must be completed to allow automatic sharing of records via interoperability between metadata management facilities and data portals.

Research Products: data sets, geographic information, physical samples, reports, and publications

Data Type I – data to be released to the general public according to the terms of the general data use agreement (see below) within two (2) years from collection and no later than the publication of the main findings from the dataset.

Data Type II – data to be released to restricted audiences according to terms specified by the owners of the data. Type II data are considered to be exceptional and should be rare in occurrence. The justification for exceptions must be well documented and approved by the lead PI and Site Data Manager. Some examples of Type II data restrictions may include: locations of rare or endangered species, data that are covered under prior licensing or copyright (e.g., SPOT satellite data), or covered by the Human Subjects Act. Researchers that make use of Type II Data may be subject to additional restrictions to protect any applicable commercial or confidentiality interests