Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

RTG 2530
BIOTA-MEDIATED
EFFECTS ON CARBON
CYCLING IN ESTUARIES

Research Training Group 2530
Data management Plan

Universität Hamburg

May, 2023

Contents

Introduction

This document outlines the procedures, suggestions, rules, and management for the data generated through the Research Training Group (RTG) 2530. The Data Management Plan (DMP) is essential to providing the most sustainable, and therefore, most useful data for future research, especially for understanding long-term and changing biota-mediated carbon dynamics in estuaries. The DMP outlines the overall vision for data management as the project progresses, including the Who, What, Where, When, and How. It includes a detailed protocol and set of guidelines for individual researchers to follow when preparing their datasets. It also includes technical information regarding long-term data storage, personnel assignments and curation, as well as legal information regarding data ownership. The RTG 2530 is designed to last nine years and includes many separate research projects with the same overall theme and often utilizing the same study sites. If followed, this DMP will help ensure the information gathered through the RTG will be harnessed for optimal use over the longest possible period of time. Such data sustainability is essential for our continuing understanding of the natural world.

The Who of this document outlines the roles and responsibilities of data management within the RTG. The What describes what sorts of data should be included in the plan and what should be left out. Where details the initial and long-term storage drives and locations for the data, and the When describes any recommended backup schedules, as well as transfer and public access time limit milestones. Finally, the How of the DMP provides a review of the above components by describing in detail the best practices for creating, maintaining, and storing datasets for the RTG. Here, there is a protocol and set of simple rules that should be followed at the beginning of every research project to ensure smooth and lossless data management.

Summary

Who	Each doctoral Researcher (PhDs) and other RTG researcher recording data is responsible for accurate and clear data documentation, as well as data backup and security during their project duration. One person, a named data curator, is then responsible for curating all data sets and uploading them to a repository. Facilitative and supervisory roles are also required of PIs overseeing doctoral research.
What	Any raw data used in final analyses, referenced in publications, or otherwise relevant to describing the research, but not intermediate data or notes. Additionally, ALL data sets must be accompanied by metadata.
Where	Data should be stored on a network drive. All UHH members have personal network drives through their laptops or through the UHH VPN. Storing data on UHH network drives secures data safety as automatic (daily) backups are carried out by UHH. Otherwise, data should be automatically backed up on a separate drive. An intermediate drive will also be established to facilitate the sharing of preliminary data within the RTG and final data sets will be stored in a publicly accessible online repository.

When Data should be backed up daily at a minimum. Migration to an online repository will be done before the PhDs finish their program. All data will be publicly available for at least ten years.

How By following the protocols of this document and by using subject-specific standards when necessary.

Who (roles and responsibilities)

Person	*Roles and Responsibilities*
PhD students	• Create data sets according to this plan and to unique specifications of project and supervisor • Responsible for data management and longevity during individual project • Ensures consistent data backup via University of Hamburg Z:\ drive or other means when necessary • Ensures data transition to curator at end of project
Principal Investigators	• Supervises the data management of the PhD student(s)
Data steward/curator	• Teaches projects members about the DMP • Ensures compliance with DMP by individual researchers via periodic consultations • Oversees long-term data sustainability • Responsible for data transitions from individuals to archives and repositories
RTG spokesperson	• Assigns new curator/stewardship roles as required • Responsible for ensuring final data standards and archiving at project finale

What

Not all data generated by a researcher will fall within the purview of the DMP. Only raw measurements and final results are necessary for curation and long-term archiving and public access. Transitional data sets used in trials or for model development are not required to be curated and/or archived, unless they are essential to interpreting your results or they are referenced in publications. Data should be as close to the raw measured values as possible, and for each replicate in the original study, not summarized means or other aggregations for a particular treatment. If the raw values are otherwise deceptive (e.g. due to faulty equipment), corrected values can be used but corrections must be noted in the metadata.

Additionally, ALL data curated for the RTG MUST be accompanied by metadata that describes each dataset in detail. Metadata simply describes the dataset in more detail, allowing anyone else who uses the data to more accurately interpret it and use it. Usually, metadata includes information about the dataset creator(s) and how to contact them, methodological descriptions about how, when, and where the data were obtained, and details about the variables and their units. An example of metadata for the RTG is provided in the appendix, and a template Excel file can also be found in the project's data folder.

Where

A regularly backed up network drive should be used as the working directory. All UHH members have personal network drives through their laptops or through the UHH VPN. Storing data on these UHH network drives secures data safety as automatic (daily) backups are carried out by UHH. If this drive is for any reason not available, an alternative system with automatic backup should be used for all ongoing projects. Using network drives with automatic backups means that the live version of all data is the version on the network drive, and that older versions of the data are automatically stored ('backed up') by the system. This procedure ensures a consistent back-up, with multiple historical versions. Using this system, older versions of the data are always available in the case of accidental data loss or errors. When working files are instead stored on local computers with no automatic backups, those files and all the information they contain are at greatest risk for damage and loss.

Additionally, because the RTG is a collaborative project meant to foster interchangeability of data and results, an intermediate network data folder has been established for the project to store in-progress and preliminary data. This will allow others to explore results as they are generated and test hypotheses to better direct future research. The location of this intermediary folder is located on UHH's network system (Z:\min\FBBIO-IPM-ANVEG\RTG_2530).

Before submitting any manuscripts for publications, data should be made available in a publicly accessible repository, and the address and/or digital object identifier (DOI) of the data should be provided within the publication. This ensures studies and analyses can be repeated, as is necessary for the scientific process, and that data can be included in subsequent meta-analyses or literature reviews. In case researchers would like to maintain their data private while other manuscripts are awaiting publication, it is also possible to place moratoriums on repository data for a specified amount of time, meaning the data will not be made publicly available until the moratorium has expired. The primary repository used for the RTG 2530 will be the Research Data Repository (RDR) managed by the Center for Sustainable Research Data Management of Universität Hamburg (https://www.fdr.uni-hamburg.de/). For molecular and genomic datasets, alternative data repositories may be used.

When

As suggested in the "Where" section, working files should be backed up daily at a minimum. If you are working from your personal UHH K drive, this is done automatically. Alternatively, the University of Hamburg also offers free backup software to help you coordinate and automate scheduled data backups to an external drive. These can be found through the RRZ service portal: https://rrz-serviceportal.uni-hamburg.de/wm/app-SelfServicePortal. Also mentioned in the "Where" section is an intermediate data folder (Z:\min\FBBIO-IPM-ANVEG\RTG_2530) accessible for all members of RTG 2530. This drive should be used after some results have been generated but before they are finalized. This will be important to allow other researchers access to your data for their own preliminary

hypothesis testing. Data should be stored here on demand, or when others request it. It is important to note here that all RTG members agree to a data use and authorship agreement when collaborating with other members. A template for a data use agreement can be found in the appendix, and data creators are responsible for ensuring it has been acknowledged by any party with access to the data.

Per funding of the RTG via the DFG, all data generated through the RTG 2530 should be made publicly available "as soon as possible" and should be archived for at least 10 years. As discussed above, this ideally means that all data should be uploaded to a repository upon the publication of any relevant manuscripts. This allows the data to be accessed by anyone with further interest upon reading the publications. In some cases, public access moratoriums may also be utilized to maintain the data as private for up to two years. At the very latest, data should be made publicly available via a repository before the PhD students finish their degree.

How

Modified from the guide provided by the Ecological Society of America.
https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/0012-9623-90.2.205

Data Creation Protocol

1. Provide raw, uncorrected results. Any corrections should be limited but made as deemed necessary by individual data users. All corrections or modifications to the raw data should be described in the metadata.
2. Use a predetermined set of consistent names for your variables and observations.
    - As an example, most, if not all, study sites should be one of the following listed in the tables below, such as:
        i. KWK_HMG_UD_03
            - Kaiser-Wilhelm-Koog, High Marsh-Grazed, Undisturbed, plot 3
        ii. HH_PZ
            - Hohenhorst-Haseldorf, pioneer zone, all plots
                a. This might be the case for the data logger probes, or for the soil profiles, which are meant to represent all the plots at this site.
        iii. MT_LM_15_2
            - The MERIT experiment at the Hamburger Hallig, low marsh, treatment 1.5°C, replicate 2
        iv. ML_0
            - Mühlenberger Loch, surface (0 m depth)
        v. TWI
            - Twielenfleth, all depths

3. Data should be stored in tabular format, when at all possible.
    - CSV if possible.
    - Header line
        i. This is the first row of the data set and should include the variable names.
    - Use Plain ASCII text encoding.
        i. This eliminates the accented characters (Umlaut, Eszett, etc.)
    - When possible, data should be constructed so that it is updated by adding rows, not columns.
        i. This means, for data sets with multiple variables (like water quality, e.g. pH, Temp, DO etc.) one of the variables should be "measurement" representing the specific variables, and the other should be "value" representing the actual measured value. This way, multiple variables can be stored in the same dataset without adding new columns for each variable.
    - One column should contain only one type of information (numeric, alphabetic, dates, times, Boolean, etc.), and should not mix data types.
    - To avoid unnecessarily large datasets, avoid including constant value columns, like site coordinates, eco-types, taxonomy, which will always be the same. Instead, use only the unique site names (or species name). This data can be easily joined to a separate table containing the other variables associated with that species or that site.
        i. This method relies on unique "key" fields or key variables that are shared between the two tables, like the site name or the species name.
    - Measurement variables should include the units in the name when possible.
        i. If not possible or if units are too complex, they MUST be clearly found in the metadata.
    - Taxonomic information should include the source (Linnaeus, Darwin, etc.) to avoid future confusion when taxa are reclassified.
    - Dates should ideally be stored in Coordinated Universal Time (UTC) and in the International Organisation of Standards (ISO) format (YYYY-MM-DD hh:mm:ss). This usually means one hour behind Central European Time (CET), but can also be two hours behind during daylight savings time.
        i. If this is not practical for initial data recording, it must be addressed during data storage. Writing the timezone (CET) in the metadata is a good alternative, when UTC conversion is not practical.

4. Always include Metadata
    - Metadata are extremely important to the long-term usability of your data. Sufficient time and thought should be given to this step to ensure clear and precise data interpretation. A metadata template is provided in the project network drive and an example is presented in the appendix of this document. Metadata should include the following information at a minimum:

i. Methods, contact information for those responsible for the data, descriptions for any variables (including units) or abbreviations used in the data.

RTG 2530 Sites

Marsh Sites

The marsh sites consist of the original three locations along the lower Elbe, at Hohenhorst-Haseldorfer Marsh, Hollerwettern, and Kaiser-Wilhelm-Koog, as well as additional sites at Heuckenlock and at the MERIT experiment on the Hamburger Hallig. The original marsh sites have a consistent design among them that includes three marsh zones (pioneer zone, low marsh, high marsh) and each separated into five equal replicates. Data entry should reflect where an observation was made within this site design, by including the unique site IDs for every location within the RTG. These IDs have the form of LL_MM_ZZ_RR, where LL is the location, MM is a combination of marsh zone (pioneer zone, low marsh, high marsh) and management (Grazed, ungrazed), ZZ is the sampling area (undisturbed, vegetation & soil sampling, greenhouse gas measurement) or treatment, and RR is the replicate (1, 2, 3, 4, 5). A marsh site ID table is as follows:

Location (LL)	Marsh zone and Management (MM)	Zones (ZZ)	Replicates (RR)	ID (LL_MM_ZZ_RR)
Hohenhorst-Haseldorfer Marsh (HH)	High Marsh – Grazed (HMG)	• Undisturbed (UN) • Veg. & Soil Samp. (VS) • Gas Meas. (GG)	• 01 • 02 • 03 • 04 • 05	HH_HMG_XX_XX
	High Marsh – Ungrazed (HMU)			HH_HMU_XX_XX
	Low Marsh (LM)			HH_LM_XX_XX
	Pioneer Zone (PZ)			HH_PZ_XX_XX
Hollerwettern (HW)	High Marsh – Ungrazed (HM)			HW_HM_XX_XX
	Low Marsh (LM)			HW_LM_XX_XX
	Pioneer Zone (PZ)			HW_PZ_XX_XX
Kaiser-Wilhelm-Koog (KWK)	High Marsh – Grazed (HMG)			KWK_HMG_XX_XX
	High Marsh – Ungrazed (HM)			KWK_HM_XX_XX
	Low Marsh (LM)			KWK_LM_XX_XX
	Pioneer Zone (PZ)			KWK_PZ_XX_XX
Heuckenlock (HK)	*to be determined*			HK_XX
Hamburger Hallig (MT)	High Marsh (HM)	• Ambient Temp. (00) • + 1.5 °C (15) • + 3.0 °C (30)	• 01 • 02 • 03	MT_HM_XX_XX
	Low Marsh (LM)			MT_LM_XX_XX
	Pioneer Zone (PZ)			MT_PZ_XX_XX

Aquatic Sites

Aquatic sites consist of a set transect along the Elbe River. The transect sites are accessed via research vessels and commercial fishing vessels, or in some cases from piers. Samples are taken from each station at various depths, thus each depth represents a different location and should be identified accordingly.

Location	Depth	ID
Bunthaus (BH)	0-10	BH_0, BH_10 etc.
Mühlenberger Loch (ML)		ML_0, ML_10 etc.
Twielenfleth (TF)		TF_0, TF_10 etc.
Schwarztonnensand (SS)		SS_0, SS_10 etc.
Brunsbüttel (BB)		BB_0, BB_10 etc.
Medem Reede (MR)		MR_0, MR_10 etc.
Seemannshöft (SH)		SH_0, SH_10 etc.

Appendix

Meta Data Example

Meta data should describe each dataset in more detail. This includes information about the methods, contact information for those responsible for the data, and descriptions for any variables or abbreviations used in the data. A metadata template is available for the RTG 2530, and when filled in, looks something like the example below for the water level monitoring data. The template consists of three spreadsheets within an Excel workbook and may include figures if necessary to aid in methodological interpretation.

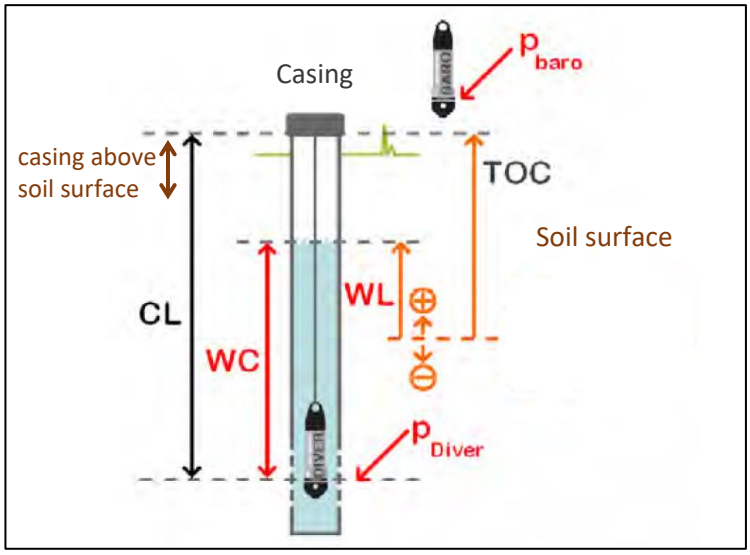
Metadata worksheet page 1: General Information

Title of Dataset	*Water pressure and temperature at RTG 2530 marsh sites*
Responsible person	
Name	Friederike Neiske
Institution	Universität Hamburg
Email	friederike.neiske@uni-hamburg.de
Alternative contact person	
Name	Volker Kleinschmidt
Institution	Universität Hamburg
Email	volker.kleinschmidt@uni-hamburg.de
Date(s) of data collection	
KWK-HM	8/19/2021 3:00:00 PM - ongoing
KWK-HM-grazed	8/19/2021 1:00:00 PM - ongoing
KWK-LM	9/2/2021 12:30:00 PM - ongoing
KWK-PZ	8/30/2021 12:10:00 PM - ongoing
KWK-baro	8/30/2021 12:00:00 PM - ongoing
HW-HM	8/17/2021 1:00:00 PM - ongoing
HW-LM	8/20/2021 11:00:00 AM - ongoing
HW-PZ	9/8/2021 12:30:00 PM - ongoing
HW-baro	19/8/2021 3:00:00 PM - ongoing
HH-HM	X
HH-HM-grazed	3/11/2021 1:45:00 PM - ongoing
HH-LM	X
HH-PZ	X
HH-baro	X
Location of data collection	
KWK-HM	N 53.925980°, E 8.918190°
KWK-HM-grazed	N 53.92865°, E 8.91720°
KWK-LM	N 53.927810°, E 8.914240°
KWK-PZ	N 53.925460°, E 8.916480°
KWK-baro	N 53.925460°, E 8.916480°
HW-HM	N 53.834497°, E 9.372230°
HW-LM	N 53.834200°, E 9.370990°
HW-PZ	N 53.83325°, E 9.36960°
HW-baro	N 53.83325°, E 9.36960°
HH-HM	X
HH-HM-grazed	X
HH-LM	X
HH-PZ	X
HH-baro	X
Funding sources	Research Training Group 2530: Universität Hamburg.
Was data derived from another source?	No

Metadata worksheet page 2: Methodological Information

Description of methods used to collect the data	*Data was collected from automated sensors/dataloggers installed in 2 cm diameter slitted PVC wells with a maximum depth of 2 m below the soil surface. The sensors hang from a wire attached to the well cap, with variable wire lengths for each location. These wire lengths, as well as the length of well protruding from the soil surface are necessary information to accurately calculate water depth from the recorded pressure.*
Methods used for processing the data	Data are collected from the data loggers via a USB chord. Variables recorded include the date and time as well as the pressure and temperature of the sensor. To properly convert the recorded pressure into a depth in reference to the soil surface, the following additional variables are needed: the wire length between the top of the well cap and the instrument sensor (CL), the top of casing (TOC) length exposed above the soil, and the depth of water above the sensor (WC). With these three variables, the water level (WL) in reference to the soil is calculated as WL=TOC-CL+WC, or by including the calculation for WC, it becomes: WL = TOC-CL+9806.65*(ρDiver - ρBaro)/(ρ*g). Here, ρDiver and ρBaro are the pressure measurements for the instrument in the well and another instrument with barometric pressure nearby, both in the same units of cm H_20, ρ is the density of the water in kg/m^3 (1000 for fresh water at 4°C, 1026 for salt water at 15°C) and g is the acceleration due to gravity (9.81 m/s^2). Location specific well pecifications are given below
Location	DL (cm)
KWK-HM	-147
KWK-HM-grazed	-149.5
KWK-LM	-120
KWK-PZ	-97
HW-HM	-102
HW-LM	-148.5
HW-PZ	-140.5
HH-HM	-102
HH-HM-grazed	
HH-LM	-178.5
HH-PZ	-138
Instrument or software specific information necessary for data interpretation	Instruments are vanEssen Micro-Divers (D16xx). No special software required for data processing. A user manual can be found here: https://diver-water-level-logger.com/files/media/Downloads/m2-1111ediver.pdf
Standards and calibration information	None

Environmental/experimental conditions	Soil conditions are variable across the sites and may influence the results. Soils with very low permeability, such as those with high clay content found in many of the RTG sites, may not allow for proper water drainage from the wells. This is known as the hydrostatic time-lag. Some manual examination and manipulation of the data may be necessary to eliminate periods when the hydrostatic time-lag produces erroneous data.
Quality assurance procedures performed on the data	None
Other people involved with sample collection, processing, and analysis	



Metadata worksheet page 3: Variables

Number of variables	*5*
Number of observations	10000+
Missing data codes	NA
Special formats/abbreviations used	NA, see time and date descriptions
Variable	Description
Date	Date of the observation in format DD/MM/YY
Time	Time of the observation in 24-hour format of hh/mm/ss
TimeStamp	The combined date time of the observation in format YYYY/MM/DD hh:mm:ss, with the time in 24 hour format and in Central European Time (CET), with no change to Summer Time (CEST).
Pressure [cmH20]	The measured pressure of the sensor in units of cm H_20, or the pressure exerted by X cm of FRESH water at a temperature of 4°C at sea level. 1 cm H_2O is equal to 98.0638 Pa. This is the total pressure (air + water) exerted on the sensor. To get only the water pressure, and thus the water depth, the air pressure first needs to be subtracted. This can be obtained from a separate sensor placed nearby or from a nearby weather observation. To arrive at the water depth, the equation presented in the "methodological information" should be followed.
Temperature	The temperature of the water in °C. This can be used, along with the salinity to arrive at the water density, which is necessary for accurate calculation of the water depth.

Data Sharing Agreement

Name of organization providing the data (data owner):

_____

Name of organization requesting data (data processor):

_____

Name and position of person requesting data:

_____

Date request received:

_____

Description of data requested:

_____

_____

Decsribe any special considerations for data retention or deletion (e.g. length of data access and use, must data be deleted upon completion, etc.):

_____

_____

_____

Can data be shared with a third party outside of this agreement?

- ☐ Yes
- ☐ No

Signatures, printed names, and date

Data Owner

Data Processor