DMP TEMPLATE FOR PRE-CLINICAL LABORATORY STUDIES

(Examples were created for a project to develop a treatment loaded biocellulose duraplasty as a potential treatment for stroke.)

This document is a working draft and will continue to undergo piloting and refinement. It represents an ongoing effort to improve data management plans in the biomedical sciences.

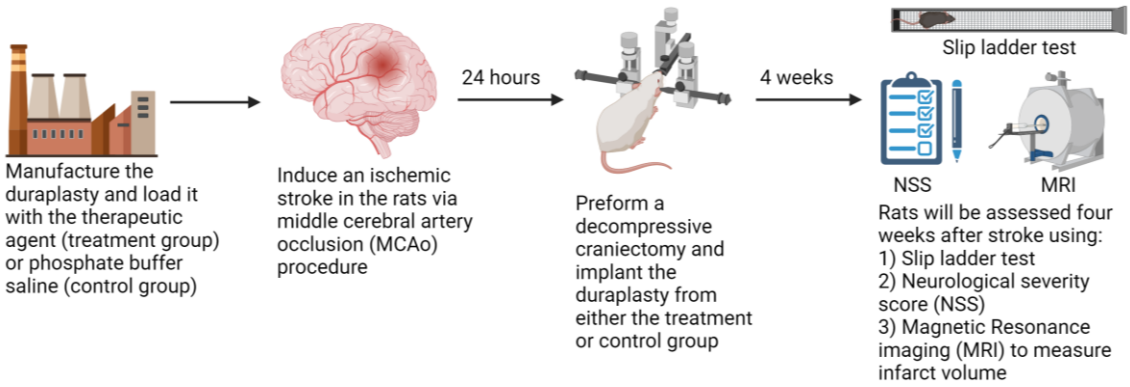
The individuals involved in the development of the template include Dr. Anna Catharina V. Armond, Brian Dorus, Meredith Conboy, Dr. Forough Jahandideh, Dr. Manoj Lalu, and Dr. Kelly D. Cobey.

1. DATA DESCRIPTION AND COLLECTION

1a. Describe the study for which the data are being collected.

Example:

Due to the limited treatment options for stroke patients, more creative therapeutic approaches need to be explored. Some stroke patients can experience brain swelling which requires a portion of their skull to be removed and the dura opened up to allow the brain to swell. To protect the brain from the environment, a duraplasty is placed over the opening. This procedure is known as a decompressive craniectomy, which reduces mortality rates but still often leaves patients with lifelong disabilities. We have developed a biocellulose based duraplasty that can release a therapeutic agent directly to the site of injury for potential functional recovery. Therefore, the aim of this study is to assess the effects of a biocellulose based duraplasty loaded with a therapeutic agent compared to a biocellulose duraplasty with no treatment. These experiments will be performed in male and female rats using a well-established model of stroke. Outcomes of interest include motor skills, infarct volume, and behavioural function. The following figure provides a simplified outline of this experiment.



1b. What types of data will you collect, create, link to, acquire and/or record?

Example:

We will use the middle cerebral artery occlusion (MCAo) model of stroke in rats. Baseline and outcome data collected are generally accepted by the preclinical stroke community for their importance and relevance. Data collection forms will be made available through a project tab that will be established on the Open Science Framework. The STAIR (Stroke Treatment Academic Industry Roundtable)1 and PREPARE (Planning Research and Experimental Procedures on Animals: Recommendations for Excellence)2 guidelines were used to inform experimental design, sample size calculation, and data collection.

Data collected before the experiment (baseline)	***What data is being collected***	***Example***	***The type of data being collected***
	Type of stroke induced	*Ischemic stroke*	*Text (Nominal)*

	Method to induce stroke	*Middle cerebral artery occlusion (MCAo)*	*Text (Nominal)*
	Rat strain	*Wistar rats*	*Text (Nominal)*
	Sex	*Equal amount of male and female rats*	*Text (Nominal)*
	Sample size	*We will use the following website to calculate sample size: Web-based Sample Size/Power Calculations (ubc.ca)*	*Numeric (Discrete)*
	Date of birth (YYYY-MM-DD)	*2022-01-01*	*Numeric (Discrete)*
	Age at time of experiment	*600 to 730 days*	*Numeric (Discrete)*
	Body weight (g)	*650g for male rats* *400g for female rats*	*Numeric (Continuous)*
	Core temperature (ºC)	*37ºC*	*Numeric (Continuous)*
	Neurological severity score (Baseline)	*NSS will be based on this article: https://doi.org/10.1016/0006-8993(94)01433-I*	*Tabular data (Ordinal)*
	Treatment group	*Control: duraplasty with PBS* *Treatment group: duraplasty with therapeutic agent*	*Text (Nominal)*
	Endpoint	*48 hours post MCAo*	*Numeric (Discrete)*
	Anesthetic agent	*Ketamine/xylazine*	*The agent itself: Text (Nominal)* *Time of administration: numeric (Discrete)* *Concentration of agent: Numeric (Continuous)* *Number of Doses: Numeric (Continuous)* *Route of administration: Text (Nominal)*
	Therapeutic treatment	*Biocellulose based duraplasty loaded with a therapeutic agent*	*The agent itself: Text (Nominal)* *Time of administration: numeric (Discrete)* *Concentration of agent: Numeric (Continuous)* *Number of Doses: Numeric (Continuous)* *Route of administration: Text (Nominal)*
	Analgesia agent	*Buprenorphine*	*The agent itself: Text (Nominal)* *Time of administration: numeric (Discrete)* *Concentration of agent: Numeric (Continuous)* *Number of Doses: Numeric (Continuous)* *Route of administration: Text (Nominal)*
	Slip ladder test	*We will use the following article to develop our test: doi: 10.3791/1204*	*A video will be taken and upon review a score is given:Video data (Continuous) and Tabular data (Ordinal)*
	Body weight (g)	*635g for male rats* *387g for female rats*	*Numeric (Continuous)*

Data collected during the experiment	*Core temperature (°C)*	*35°C*	*Numeric (Continuous)*
	Neurological severity score (NSS)	*NSS will be based on this article: https://doi.org/10.1016/0006-8993(94)01433-I*	*Tabular (Ordinal)*
	Time of death if specified endpoint is not reached	*During the MCAo procedure*	*Numeric (Discrete)*
	Slip ladder test	*We will use the following article to develop our test: doi: 10.3791/1204*	*A video will be taken and upon review a score is given: Tabular data (Ordinal)*
	Infract volume	*We will use MRI to obtain an image of the infract volume and then the Olea Sphere 3.0 software to calculate the infract volume*	*Image to determine infarct volume: Numeric (Continuous)*

1c. How will new data be collected or produced and/or how will existing data be re-used?

Example:

Given the exploratory nature of preclinical studies, we anticipate the results of our initial experiments will influence the next set of experiments. We will begin by randomizing 20 rats following induction of stroke into two groups. Allocation will be stratified by biological sex. Group 1: stroke + biocellulose based duraplasty loaded with PBS, Group 2: stroke + biocellulose based duraplasty loaded with therapeutic agent.

For each measure listed above the following methods will be used to produce and collect data:

Data that will be recorded	***How the data will be generated***
How the Stroke was induced	*We will be using the following article to develop our stroke model: doi: 10.3791/2423-v*
Body Weight	*Tared Sartorius Cubis™ II Precision (Fisher Scientific).*
Core Temperature	*We will be using the RET-2 non-isolated rectal probe (Kent Scientific) paired with the 911B Thermocouple thermometer (TEGAM).*
NSS	*We will be modifying our NSS, using the following article as a foundation: https://doi.org/10.1016/0006-8993(94)01433-I*
Ketamine/xylazine dose	*The dosage we will be using for rats is 0.2 mL/100 g of body weight.*
Treatment Dose	*We will perform a drug release study to determine how much therapeutic agent our duraplasty is releasing.*
Buprenorphine	*We will administer 0.05 mg/kg of body weight 1 hour before the surgery.*
Slip Ladder Test	*We used the following article to develop our test: doi: 10.3791/1204*
Infarct Volume	*We will use a high-field small-animal MRI machine (7.0 Tesla Discovery MR901 System from Agilent Technologies/General Electric). The resulting image will then be processed through the Olea Sphere 3.0 software to determine the infract volume.*

Data will be initially collected in standard forms within lab books. The lab books are individually prenumbered in our laboratory. In addition, each book contains pre-numbered pages. The data will then be transferred to an electronic database using REDCap (Research Electronic Data Capture). Data that is processed/analyzed will be saved in the research institute's MS OneDrive in a project specific folder.

Data provenance will be established by including metadata on experiment day and lab book where data was initially recorded. In addition, we will note if/how data was transformed prior to entry in database.

This study will not reuse any existing datasets (we are unaware of any similar datasets that have been produced and made publicly available). Therefore, all the following data will be generated by our research technician, Vasco L.

Burgess who will complete lab book forms to record the animal's biological sex, strain, DOB, weight, etc. They will also fill out lab book forms on outcomes such as the NSS, slip ladder test, and infarct volume.

1d. What file formats will your data be collected in? Will these formats allow for data reuse, sharing, and long-term access to the data?

Example:

The data will initially be recorded in archived laboratory notebooks. As described above, data will be entered into a REDCap database. This will allow complete datasets to be available in comma-separated values (.csv) format. In addition, PDF reports of the data can be generated. Images from MRI will be linked to the database as joint photographic experts group (jpeg) with a minimum 300dpi resolution. Videos from the slip ladder test will be available in audio video interleave (avi) format.

1e. What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?

Example:

We have standardized the file naming convention to capture all necessary information without compromising any implemented blinding protocols. The file name, as shown in Figure 1, will be composed of four sections, outlined as follows.

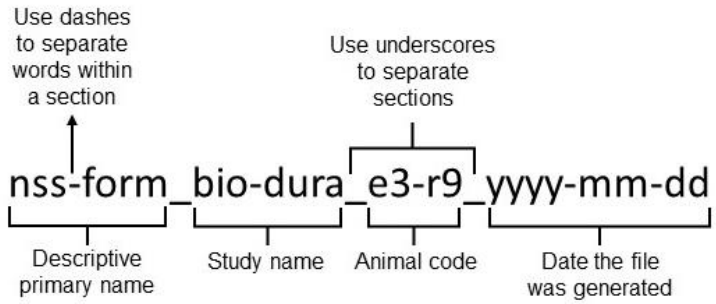


Figure 1. Naming convention for raw data.

The descriptive primary name will be short (≤35 characters), but meaningfully describe the contents of the file. Contain no spaces. Rather than using spaces, the use of dashes should distinguish different words within the primary file name (e.g. nss-form).

The study shorthand name will be all lowercase letters and should not contain any spaces.

The animal code will link all the data associated with every individual rat, where "e#" represents the experiment number and "r#" is the rat number for that specific experiment.

The date code will reflect the day the data file was generated. We will use ISO 86011 format: YYYY-MM-DD.

For example, at the primary site for the third experiment, the following naming convention should be used for the 9th rat in the study - nss-form_bio-dura_e3-m9_2023-06-14.

When processing data to generate graphs or perform statistical analysis, all aspects of the file naming convention remain the same except the animal code will be replaced with a version code. This modified naming convention is represented in Figure 2.

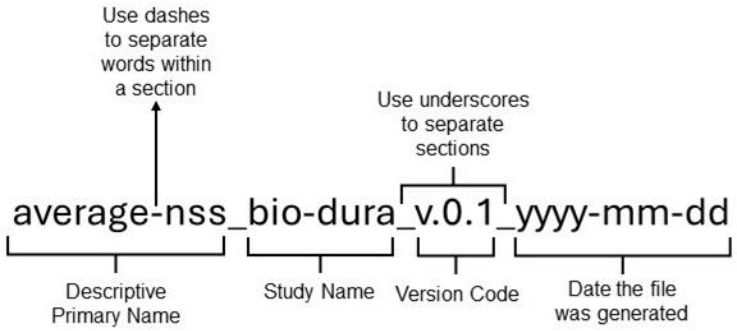


Figure 2. Naming convention for processed data

The version code will be used to distinguish different versions of the document.
- The version code will use the v.X.Y format where the X represents the major version number and the Y represents the minor version number, separated by a period.
- Each time changes are made to the document (but it is not finalized), the minor version number is updated (e.g changing from 0.1 to 0.2)
- Once the processed data is finalized, and no foreseeable changes will be made, the major version number will be updated to v.1.0
- If the finalized data needs to be modified in the future, a new copy should be made where the major version number remains the same but the minor version number increases by 1 (going from v.1.0 to v.1.1). Once this data is then finalized, the major version number should increase by 1 (going from v.1.1 to v.2.0).

2. DOCUMENTATION AND METADATA

2a. What documentation will be needed for the data to be read and interpreted correctly in the future?

Example:

To help future researchers interpret our published data we will include the following:
- Experimental SOPs and any deviations
- Data collection forms, definitions of preclinical characteristics, and outcomes
- Naming conventions
- We will also include a list of all personnel involved in the project, along with a list of their tasks throughout the project.

Documentation on the processing and statistical analysis of all data will include:
- The original raw data (when feasible)
- Processes utilized to clean and code the data
- Statistical software and type of statistical analysis (e.g., student's t-test, one-way ANOVA followed by a multiple comparison test, two-way ANOVA, etc.) and syntax files used.

2b. Describe in detail the data-level descriptions, codes, and definitions.

This is all the data we will collect:

1. *Rat strain*
2. *Sex (Female/Male)*
3. *Date of Birth (YYYY-MM-DD)*
4. *Age at time of experiment (weeks)*
5. *Body weight (kg)*
6. *Core temperature (oC)*

7. *Neurological Severity Score (NSS) at Baseline and post treatment (4 weeks after stroke). A NSS is a scoring system where a Senior Research Technician (Vasco L. Burgess) assesses the reflexes, balance, and sensorimotor functions of an animal and provides a score, within a range, for each category. The score for each category is then added; the higher the score the more severe the deficit is.*

8. *Slip Ladder Test at baseline and post treatment (4 weeks after stroke). A video recording of a rat crossing a horizontal ladder is taken and the Senior Research Technician (Vasco L. Burgess) will review it to assess the placement of the animal's paw as it moves along the ladder. A score is then provided for every paw placement and the final score is the sum of all paw placements. The lower the score the more severe the deficit is.*

9. *Infarct volume. To determine infarct volume the Senior Research Technician (Vasco L. Burgess) will use MRI to generate images of the brain. Then the Olea 3.0 software will calculate the infarct volume, which represents the amount of the brain affected by the stroke.*

2c. How will you make sure that documentation is created and captured consistently throughout your project?

Example:

To ensure accuracy, consistency, and completeness, we will institute the following measures:
- SOPs will be reviewed with all involved personnel prior to experiments.
- Training on data entry into the lab book and the database will be provided by an experienced research technician.
- Data entry will be audited by a senior team member.
- In addition, when data is entered into the REDCap database, field validation will be used to ensure accuracy (e.g. flagging entry of numerical data outside of predetermined expected ranges). If any inconsistency is observed, an investigation will be opened until the inconsistency is resolved.
- All deviations from SOPs will be recorded with an explanation. In addition, a senior team member will consult with personnel to ensure workflows for data entry remain feasible and that changes to procedures are documented.

2d. If you are using a metadata standard and/or tools to document and describe your data, please list here.

Example:

We will standardize our vocabulary based off the Darwin Core: https://www.tdwg.org/standards/dwc/

Since we will be uploading our data to Open Science Framework, we will follow their metadata standards, which states to follow a metadata scheme that is common to our project. Therefore, we will follow the metadata scheme developed by DataCite: https://schema.datacite.org/.

3. STORAGE AND BACKUP

3a. What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

Example:

The estimated storage-space is 1 terabyte of data. There are no restrictions on how long to retain the data as we will generate non-sensitive laboratory animal data.

3b. How and where will your data be stored and backed up during your research project?

Example:

The data will be stored using the 3-2-1 backup rule. Three copies of every piece of data will be generated, the original data and two backups. During the data acquisition stage, all data generated from different experiments will be stored on the personal computer of the person collecting the data and stored in their own project folder. This project folder must be linked to the cloud-based Corporate Microsoft 365 platform, OneDrive, which will synchronize to our lab computer.

The two backup copies of this data will be stored on two different types of media, one on a OneDrive folder that is dedicated to the lab and one on an external hard drive owned by the PI (Dr. Devon B. Swift).

The dedicated lab computer is not assigned to one individual but is controlled by the PI (password-protected). The data will be backed up on a OneDrive folder dedicated to this project accessible only by the PI (Dr. Devon B. Swift), students, and staff members (Dr. Katrina Cortex, Vasco L. Burgess) directly involved.

The external hard drive will be stored by the PI and will be updated every three months to add any new versions created since the last update.

3c. How will the research team and other collaborators access, modify, and contribute data throughout the project?

Example:

During this research project, all data will be stored in a OneDrive folder that will be shared between our PI (Dr. Devon B. Swift), students, and staff members (Dr. Katrina Cortex, Vasco L. Burgess) directly involved. The PI (Dr. Devon B. Swift) will need to grant access to others not directly involved in the project, in which they will have to log in (username and password). All personnel directly involved in the project will be able to modify and process the data. For any data modification, a new version file will be generated and reflected in the file name (see file naming conventions). Once the raw data is uploaded into OneDrive (non-editable version), a copy of the data will be made for modifications (editable version). Furthermore, modification is restricted to the data processing (generating graphs, statistical analysis, etc.), and not modifications to the original raw data. For any modification done to data, a new file will be generated with its file name indicating which version we are on.

4. PRESERVATION

4a. Where will you deposit your data for long-term preservation and access at the end of your research project?

Example:

Upon completion of the project and publication in a peer-reviewed journal, original data, metadata, and the standard operating procedures (SOPs) will be made publicly available on the Open Science Framework in their respective formats (csv,.txt, avi, jpeg).

4b. Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

Example:

Along with uploading files in their original format, non-proprietary, preservation friendly, file formats will be used. Data will be saved as .csv, images will be saved as .jpeg, any text files (e.g. SOPs) will be saved as .txt, videos will be saved as .AVI. All experimental protocols, SOPs, a record outlining the original format of all documents and if the files were converted to a different format, as well as any deviations to these documents, will be included.

5. SHARING AND REUSE

5a. What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final, and metadata).

Example:

Once this study is completed, all data forms (raw, processed, analyzed, final, and metadata) will be published in peer-reviewed journals. The data stored in the Open Science Framework will be linked to the published article via DOI. The final analyzed data will also be uploaded to Open Science Framework in their respective preservation friendly format (.csv, .txt, .jpeg, .avi). Given that our imaging data will exceed the 50GB capacity of Open Science Framework, we will house original images in the Federated Research Data Repository. We will clearly link to this data from our project on the Open Science Framework. The required protocols, SOPs, and scoring lists to process the raw data to generate the final analyzed data will be included. We will also provide metadata, with a readme file with the coding, variables, naming conventions, and standards.

5b. What type of end-user will you use for your data?

Example:

We will share all materials via a Creative Commons license (CC BY 4.0.). The data generated is not sensitive (i.e. all lab animal data) and therefore, a CC license is sufficient. A CC BY license enables users to modify and redistribute data in any form, with proper credit given to our research group (i.e. the original generators of the data).

5c. What steps will be taken to help the research community know that your data exists?

Example:

To make our data findable and accessible, the data and metadata will be archived and shared via the Open Science Framework. The DOI number provided by the Open Science Framework will be included in the publication. DOIs promote academic credit, direct citation, and tangible metrics that our group will track. The DOI will also link to the final publication(s), as well as information on study funders and our institute where the study was performed. In addition to publishing our data in a peer-reviewed journal, we will promote our research via conference presentations, and poster presentations. The ORCID ID of every researcher involved will be linked in the publication.

To make the data interoperable, we will share detailed metadata (workflows, vocabularies, processes, and standards) and the data will be shared in preservation friendly formats as detailed above. To make the data will share it under a Creative Commons CC-BY-4.0 license.

6. RESPONSIBILITIES AND RESOURCES

6a. Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

Example:

The team will include:
Dr. Devon B. Swift - Principal Investigator
Dr. Katrina Cortex - Research Associate
Vasco L. Burgess – Senior Technician

From project inception to final publication, our Principal investigator will be responsible for:
(i) Finalizing the protocol and the data analysis plans;
(ii) Reviewing progress of the study and, if necessary, deciding on protocol changes;
(iii) Reviewing and approving study publications;
(iiii) Overseeing the project's original and analyzed data;

From project inception to final publication, our Research Associate and Senior Technician will be responsible for:
(i) Data collection and entry;
(ii) Data processing and storage;
(iii) Data documentation;
(iv) Oversee and analyze the project's data;
(v) Making, testing, and validating changes to the database

As junior team members are on-boarded to this project, their names and specific roles will be added to this DMP.

6b. How will responsibilities for managing data activities be handled if substantive changes occur in the personnel overseeing the project's data, including a change of Principal Investigator?

Example:

If any staff or trainees leave the project prior to completion, a current staff or trainee will replace them (or new personnel will be hired, if required). New personnel will receive adequate training to ensure competency in data collection and entry as detailed above. Training will be enabled through project specific SOPs as well as in-person sessions.

If the project needs to be transferred to a new principal investigator, then all responsibilities outlined in this DMP will be transferred as well.

6c. What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

Example:

Our project includes a data management budget of $15,000 which will be allocated to a Research Associate. They will ensure all the data is being consistently uploaded and that the DMP is being implemented throughout the project. For data sharing, since we are using the Open Science Framework, there is no cost for uploading our data. Given that our imaging data will exceed the 50GB capacity of Open Science Framework, we will house original images in the Federated Research Data Repository. We will clearly link to this data from our project on the Open Science Framework.

7. ETHICS AND LEGAL COMPLIANCE

7a. If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

Example:

Not applicable as only non-sensitive laboratory animal data will be collected/generated in our study.

7b. If applicable, what strategies will you undertake to address secondary uses of sensitive data?

Example:

Not applicable as the 'participants' are laboratory animals.

7c. How will you manage legal, ethical, and intellectual property issues?

Example:

The research protocol has been approved by our instituion's Animal Care Committee (ACVS #5678). All data will be made open access as described above under a Creative Commons CC-BY-4.0 license.