**DMP TEMPLATE FOR RANDOMIZED CLINICAL TRIALS**

*(Examples were extracted from a fictional trial and the template was adapted from the Portage general template - https://dmp-pgd.ca/template_export/878086536.pdf)*

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, Dr. David Moher, Dr. Florian Naudet, Dr. Dean Fergusson, Dong Vo, and Dr. Kelly D. Cobey.

## 1. DATA DESCRIPTION AND COLLECTION

**1a. Describe the Randomized Clinical Trial (RCT) for which the data are being collected.**

*Example:*

*"MOZART trial is a two-arm, parallel randomized controlled trial. The trial will investigate whether listening Mozart (Flute & Harp Concerto - 2nd Movement) reduces the time taken to fall asleep (sleep onset latency)."*

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

*Example:*

*Data will be collected with an estimated enrollment of 100 participants. Data will include baseline and follow-up information. The baseline includes participants' details, such as name, date of birth, and sex; and the Polysomnographic parameters (total sleep time (TST, min), sleep onset latency (min), rapid eye movement (REM) onset latency (min) and wake after sleep onset (min)). Follow-up information includes the Polysomnographic parameters and potential side effects. Data pertaining to the conduct of the study, such as the subject's loss to follow-up, and any deviation from the protocol, will also be collected.*

*The data collection will be achieved through a combination of sources, including electronic case report forms (eCRF) and manual data collection. Both will be made available along with the protocol.*

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

*Example:*

*The participants will be randomized into two groups (group 1: no music, group 2: music) and stratified by sex. The apparatus used for the polysomnography exams consists of different devices, including an electroencephalogram with the international 10–20 system (to rule out the occurrence of epileptic seizures), electrooculogram, electromyogram of chin muscles and upper and lower limbs, nasal pressure cannula, oral thermistor, thoracic and abdominal respiratory inductive plethysmography straps, pulse oximetry, electrocardiogram, and snoring and body position sensors. Video and sound will also be recorded during the exam. Polysomnography recordings will be obtained with "System X".*

*The trial does not plan to reuse any existing datasets, as to our knowledge, there are no similar datasets available.*

*The data will be captured in the eCRF using the web-based application REDCap (Research Electronic Data Capture). We will document any preprocessing, transformation, or cleaning steps applied to the data before analysis to ensure easy understanding and interpretation by others.*

**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 datasets will be generated in CSV format. This format allows for data reuse, sharing and long-term access to the data. The eCRF will be available in annotated PDF 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:*

*All documents, including the annotated PDF eCRF will follow the name convention of subjectID_ecrf_YYYYMMDD. The generated CSV datasets will be submitted to the Canadian Federated Research Data Repository (FRDR) to ensure that different copies or versions of files are subject to version control. FRDR has built-in version control and retains all copies of a file added to FRDR.*

**2. DOCUMENTATION AND METADATA**

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

*Example:*

*The metadata will include the RCT protocol, eCRF blank form, consent form, statistical analysis plan (SAP)(Gamble et al., 2017), definition & derivation of clinical characteristics & outcomes, training material, a readme file with the coding, variables, naming conventions and standards and regulatory documents.*

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

*Example:*

*We will collect the following data:*

1. *Subject ID number*
2. *Date of birth (dd-mm-yyyy)*
3. *Sex (Female/Male)*
4. *Total sleep time (measured in minutes and seconds) refers to the overall duration of sleep measured during the entire recording period. It includes time from sleep onset to sleep offset and is distributed throughout the sleep time.*
5. *Sleep onset latency (measured in minutes) is the period starting from when the lights are switched off for the patient to sleep, until the moment the patient falls asleep, as indicated by changes in EEG patterns and behavioral parameters associated with sleep.*
6. *Rapid eye movement (REM) onset latency (measured in minutes) represents the duration from the moment of falling asleep to the first occurrence of REM sleep during the sleep cycle. Consequently, it is influenced by the sleep latency of the individual. REM sleep cycles typically repeat at intervals of 90 to 120 minutes throughout the night.*
7. *Wake after sleep onset (measured in minutes) refers to periods of being awake occurring after the initial onset of sleep. It quantifies wakefulness during sleep, excluding any wakefulness experienced prior to falling asleep. WASO (Wake After Sleep Onset) time provides a more accurate assessment of sleep fragmentation.*
8. *Side effects*

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

*Example:*

*The data will be collected on trial-specific electronic case report forms (eCRFs) by a single investigator using the REDCap secure electronic data capture system (EDC). The data will be monitored for quality control every week by the statistician, and in case of any inconsistencies, a query will be opened in the system for the resolution of the*

*problem. For example, when collecting age, if a participant is identified as 200 years old rather that 20 years old, a quality control alert and subsequent investigation may result in an input error.*

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

*Example:*

*Whenever possible, data will be standardized to Clinical Data Interchange Standards Consortium (CDISC) standards. The generated csv dataset is a machine-readable and openly accessible format.*

**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 300GB. The source documents will be retained for at least 15 years from the completion of the study. Identifiable data will be retained only until completion of the study.*

**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 back up rule. Three copies of every piece of data will be generated, the original data and two backups. The original data will be stored on the computer disk (password protected). The two backup copies of this data will be stored on 2 different types of media, one in an external hard drive and one on a secure cloud server (with disaster recovery).*

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

*Example:*

*All data entry and access will be controlled by usernames and passwords, and any changes to data will require the user to enter their username and password. The investigators and statistician will have access restricted to the functionality and data that are appropriate for their role in the study. The statistician, Dr. Wang, and the principal investigator, Dr. Smith, will have access to the raw data*.

**4. PRESERVATION**

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

*Example:*

*When ready to make available, the data will be shared via the Federated Research Data Repository (FRDR). De-identified participant data will be made available openly within three months of publication.*

**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:*

*The datasets will be generated in the comma-separated (CSV) preservation-ready format. After completion of the data collection, data quality control will be checked for inconsistencies, de-identified, and locked. Only de-identified data will be shared.*

*For the de-identification, the direct identifiers, such as name and contact details will be deleted. The quasi-identifiers, such as date of birth will be generalized to 3-year age intervals. The de-identification process will be documented and reported in the metadata.*

**5. SHARING AND REUSE**

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

*Example:*

*We will share raw data, cleaned data, summary tables and analyses which are not released in publications. We will also provide metadata, with a readme file with the coding, variables, naming conventions and standards.*

**5b. How will you license your data?**

*Example:*

*De-identified data and metadata will be shared under Creative Commons' CC-BY-4.0 license.*

**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 Federated Research Data Repository (FRDR). FRDR assigns and promotes persistent identifiers - Digital Object Identifiers (DOIs) – to FRDR-hosted studies. This initiative promotes academic credit, direct citation and tangible metrics for researchers who collect and share data. DOIs contain metadata that identify and acknowledge the staff and sites that contribute to the dataset, as well as information on location of the study, funders, and publications or documents associated with the data. 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. To make the data reusable, we will share the de-identified data publicly under 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:*
*John Smith - Principal Investigator – (CONTACT DETAILS)*
*Emily Johnson - Co-investigator– (CONTACT DETAILS)*
*Jennifer Wang - Statistician – (CONTACT DETAILS)*

*Trial sponsor (ADD ROLES, IF ANY)*

*Principal investigator will be responsible for:*
*(i) Agreement of the final Protocol and the Data Analysis Plans;*
*(ii) Reviewing progress of the study and, if necessary, deciding on Protocol changes;*
*(iii) Review and approval of study publications and substudy proposals;*
*(iiii) Oversee and analyze the project's data;*
*(iiii) Developing the eCRF and database*

*Co-investigator will be responsible for:*
*(i) Data collection and entry;*
*(ii) Data processing and storage;*
*(iii) Data documentation*

*Statistician will be responsible for:*
*(i) Agreement of the final Protocol and the Data Analysis Plans;*
*(ii) Data processing and storage;*
*(iii) Oversee and analyze the project's data.*
*(iiii) Making, testing, and validating changes to the database*

**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:*

*Prior to the trial conduct (I.e., randomization of first participant), a trial co-investigator will be tasked with this responsibility. In case the principal investigator is no longer involved on the project, the co-investigator will be promoted to manage this project, or a qualified researcher will be hired to lead the project.*

**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 0.25FTE of Research Coordinators time to provide day-to-day input and implementation of the data management plan. We will use an open data repository (Federated Research Data Repository). Therefore, there will be no cost.*

**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:*

*Identifiable data will be retained only until completion of the study. While the study is ongoing, the data will be password protected, and will only be accessed by the investigators. After the completion of the study, the data will be de-identified. Only de-identified data will be archived and shared. Participants' names will be deleted and assigned an unique ID number. The date of birth will be generalized to a five-year interval. The de-identification will minimize the overall risk of re-identification. Therefore, this study will not include sensitive data. Moreover, we will continuously assess non anticipated risks of re-identification, and ensure that all research team is have adequate and up-to-date*

*Training and oversight changes in relevant regulations and legislation.*

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

*Example:*

*Only de-identified data will be used for secondary uses, and all study participants will provide informed consent to take part in the study and for broad data sharing openly in FRDR. The informed consent will describe that the personal identifiers will be removed and cannot be easily linked back to the identity of the participant. They will also be informed that the data will be shared publicly, unless they opt out of having their data shared.*

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

*Example:*

*The research protocol will be approved by the respective Research Ethics Board. All the participants will be informed about what data will be collected, how the data will be used for the purpose of the current study, and that the data will be available publicly and potentially reused indefinitely in ways that have not yet been conceived.*