DMP TEMPLATE FOR PAIRWISE SYSTEMATIC REVIEWS

(The template was adapted from the Portage DMP template for Systematic reviews. Examples were extracted from Beck, A., LeBlanc, J.C., Morissette, K. et al. Screening for depression in children and adolescents: a protocol for a systematic review update. Syst Rev 10, 24 (2021). https://doi.org/10.1186/s13643-020-01568-3)

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, and Dr. Kelly D. Cobey.

1. DATA DESCRIPTION AND COLLECTION

1a. Describe the Systematic Review for which the data is being collected.

Guidance:

Add a brief description identifying the population, interventions (or exposure), comparison, and outcome, often referred to as the PICO framework.

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

Guidance:

For a systematic review (or other types of knowledge synthesis), data includes the literature you find, the data that you extract from the studies, and the detailed methods and information that would allow someone else to reproduce your results. To identify the data that will be collected or generated by your research project, start by thinking of the different stages of a systematic review and how you are going to approach each: planning/protocol, data collection (literature searching), study selection (screening), data extraction, risk of bias assessment, synthesis, manuscript preparation. The PRISMA 2020 checklist provides broad categories of the stages required to complete a systematic review.

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

Guidance:

If you plan to use systematic review software or reference management software for screening (or artificial intelligence platforms (e.g., ChatGPT) and data management, indicate which program (and version) you will use, and what format files will be saved/exported in.
Keep in mind that proprietary file formats will require team members and potential future users of the data to have the relevant software to open or edit the data. While you may prefer to work with data in a proprietary format, files should be converted to open-source formats wherever possible at the end of the project. Read more about file formats: UBC Library or UK Data Service.

1d. 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?

Guidance:

It is important to keep track of different versions of files (e.g., protocol versions), files held in different formats or locations, and information cross-referenced between files. This process is called 'version control'.
Logical file structures, informative naming conventions, and clear indications of file versions, all contribute to better use of your data during and after your research project. These practices will help ensure that you and your research team are using the appropriate version of your data and minimize confusion regarding copies on different computers and/or on different media.
Read more about file naming and version control: UBC Library or UK Data Service. Consider a naming convention that includes: the project name and date using the ISO standard for dates as required elements and stage of review/task, version number, creator's initials, etc. as optional elements as necessary.

2. DOCUMENTATION AND METADATA

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

Guidance:
Typically, good documentation includes information about the study, data-level descriptions, and any other contextual information required to make the data usable by others. Other elements you should document, as applicable, include: research methodology used, variable definitions, vocabularies, classification systems, units of measurement, assumptions made, format and file type of the data, a description of the data capture and collection methods, explanation of data coding and analysis performed (including syntax files), and details of who has worked on the project and performed each task, etc.

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

Guidance:

List of all variables included in the study, along with their units of measurement, codes, if applicable, and descriptive definitions of each variable. List variables on: (1) report (e.g. author, year); (2) study (e.g., sample characteristics, country, setting, registration, funding, study design); (3) population (e.g., sex, age, socioeconomical status); (4) intervention or exposure (e.g., time, duration, dose); (5) outcomes of interest (e.g. measurements methods, effect estimates); and (6) Risk of bias. The PRISMA Explanation and Elaboration document offers additional information.

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

Guidance:

Consider how you will capture this information and where it will be recorded, ideally in advance of data collection and analysis, to ensure accuracy, consistency, and completeness of the documentation. Consider where the process and procedures for each stage of your review will be kept and shared. Will the team have a shared workspace? Who will be responsible for documenting each stage of the review? Team members responsible for each stage of the review should add the documentation at the conclusion of their work on a particular stage, or as needed. Refer back to the data collection guidance for examples of the types of documentation that need to be created.
Often, resources you've already created can contribute to this (e.g. your protocol). Consult regularly with members of the research team to capture potential changes in data collection/processing that need to be reflected in the documentation. Individual roles and workflows should include gathering data documentation.

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

3. STORAGE AND BACKUP

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

Guidance:

Storage-space estimates should take into account requirements for file versioning, backups, and growth over time. A long-term storage plan is necessary if you intend to retain your data after the research project or update your review at a later date. A systematic review project will not typically require more than a few GB of storage space; these needs can be met by most common storage solutions, including shared servers.

3b. How and where will your data be stored and backed up during the systematic review?

Guidance:

Will you want to update and republish your review? If so, a permanent storage space is necessary. If your meta-analysis includes individual patient-level data, you will require secure storage for that data. If you are not working with sensitive data, a solution like Dropbox or Google Drive may be acceptable. Consider who should have control over the shared account. Software to facilitate the systematic review process or for citation management such as Covidence or Endnote may be used for active data storage of records and PDFs.

The risk of losing data due to human error, natural disasters, or other mishaps can be mitigated by following the 3-2-1 backup rule: Have at least three copies of your data; store the copies on two different media; keep one backup copy offsite. Further information on storage and backup practices is available from the University of Sheffield Library and the UK Data Archive.

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

Guidance:

If your systematic review includes individual patient-level data (IPD), you will require secure storage for that data. As most systematic reviews typically do not involve IPD, you likely don't need secure storage. A storage space such as OneDrive should be acceptable, as long as it is only shared among team members. Consider who will retain access to the shared storage space and for how long. Consider who should be the owner of the account. If necessary, have a process for transferring ownership of files in the event of personnel changes.
An ideal solution is one that facilitates cooperation and ensures data security, yet is able to be adopted by users with minimal training. Relying on email for data transfer is not a robust or secure solution.

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

Guidance:

The issue of data retention should be considered early in the research lifecycle. Data-retention decisions can be driven by external policies (e.g. funding agencies, journal publishers), or by an understanding of the enduring value of a given set of data. Consider what you want to share long-term vs. what you need to keep long-term; these might be two separately stored data sets.
Long-term preservation is an important aspect to consider for systematic reviews as they may be rejected and need to be reworked/resubmitted, or the authors may wish to publish an updated review in a few years' time (this is particularly important given the increased interest in the concept of a 'living systematic review').
For more detailed guidance, and some suggested repositories, see "Long-Term Preservation" on the University of Calgary Library Guide.

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

Guidance:

Some data formats are optimal for long-term preservation of data. For example, non-proprietary file formats, such as text ('.txt') and comma-separated ('.csv'), are considered preservation-friendly. The UK Data Service provides a useful table of file formats for various types of data. Keep in mind that preservation-friendly files converted from one format to another may lose information (e.g. converting from an uncompressed TIFF file to a compressed JPG file), so changes to file formats should be documented.
Identify steps required following project completion in order to ensure the data you are choosing to preserve or share is anonymous, error-free, and converted to recommended formats with a minimal risk of data loss.
Read more about anonymization: UBC Library or UK Data Service.

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

Guidance:

- Raw data are the data directly obtained from the instrument, simulation or survey.

- Processed data result from some manipulation of the raw data in order to eliminate errors or outliers, to prepare the data for analysis, to derive new variables, or to de-identify the human participants.

- Analyzed data are the results of qualitative, statistical, or mathematical analysis of the processed data. They can be presented as graphs, charts or statistical tables.

- Final data are processed data that have, if needed, been converted into a preservation-friendly format.

- Consider which data may need to be shared in order to meet institutional or funding requirements, and which data may be restricted because of confidentiality/privacy/intellectual property considerations.

- Consider the supporting documentation for data analyses (Metadata).

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

Guidance:

Licenses determine what uses can be made of your data. Funders and/or data repositories may have end-user license requirements in place; if not, they may still be able to guide you in the development of a license. Once created, please consider including a copy of your end-user license with your Data Management Plan. Note that only the intellectual property rights holder(s) can issue a license, so it is crucial to clarify who owns those rights.
There are several types of standard licenses available to researchers, such as the Creative Commons licenses and the Open Data Commons licenses. In fact, for most datasets it is easier to use a standard license rather than to devise a custom-made one. Note that even if you choose to make your data part of the public domain, it is preferable to make this explicit by using a license such as Creative Commons' CC0.
Read more about data licensing: UK Digital Curation Centre.

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

Guidance:

Possibilities include: data registries, repositories, indexes, word-of-mouth, publications.
How will the data be accessed (e.g., Web service, File transfer protocol)? If possible, choose a repository that will assign a persistent identifier (such as a DOI) to your dataset. This will ensure stable access to the dataset and make it retrievable by various discovery tools.
One of the best ways to refer other researchers to your deposited datasets is to cite them the same way you cite other types of publications (articles, books, proceedings). The Digital Curation Centre provides a detailed guide on data citation. Note that some data repositories also create links from datasets to their associated papers, thus increasing the visibility of the publications.
Contact your institution for assistance as in making your dataset visible and easily accessible.
Reused from NIH. (2009). Key Elements to Consider in Preparing a Data Sharing Plan Under NIH Extramural Support. National Institutes of Health.
Other resources can be found at https://ohri.ca/journalology/data-and-materials-sharing and https://journalologytraining.ca/all-courses/.

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.

Guidance:

Your data management plan has identified important data activities in your project. Identify who will be responsible -- individuals or organizations -- for carrying out these parts of your data management plan. This could also include the time frame associated with these staff responsibilities and any training needed to prepare staff for these duties. Systematic review projects have stages, which can act as great reminders to ensure that data associated with each stage are made accessible to other project members in a timely fashion.

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?

Guidance:

Indicate a succession strategy for these data in the event that one or more people responsible for the data leaves. Describe the process to be followed in the event that the Principal Investigator leaves the project. In some instances, a co-investigator or the department or division overseeing this research will assume responsibility.
If data is deposited into a shared space as each stage of the review is completed, there is greater likelihood that the team has all of the data necessary to successfully handle personnel changes.
NOTE: Shared storage spaces such as Dropbox and Google drive are attached to an individual's account and storage capacity so consideration needs to be given as to who should be the primary account holder for the shared storage space, and how data will be transferred to another account if that person leaves the team.

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

Guidance:

Consider the cost of systematic review management software and citation management software if you are applying for a grant, as well as the cost for shared storage space, if needed. What training do you or your team need to ensure that everyone is able to adhere to the processes/policies outlined in the data management plan?
Some funding agencies state explicitly the support that they will provide to meet the cost of preparing data for deposit. This might include technical aspects of data management, training requirements, file storage & backup, and contributions of non-project staff. Find more instructions about budget estimation here:
https://www.utwente.nl/en/service-portal/services/lisa/resources/files/library-public/dcc-rdm-costs-estimation.pdf

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?

Guidance:

Most reviews do not include sensitive data, but if you are using individual patient-level data in your meta-analysis there may be data sharing agreements that are required between institutions. Similarly, patient consent to share IPD may be required. These approvals require coordination with the legal teams between multiple institutions and will necessitate secure data management practices. This type of data will not be open for sharing. Sensitive data should never be shared via email or cloud storage services such as Dropbox.

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

Guidance:

Systematic reviews generally do not generate sensitive data, however, it may be useful for different readers (e.g., funders, ethics boards) if you explicitly indicate that you do not expect to generate sensitive.

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

Guidance:

Be aware that PDF articles and even database records used in your review may be subject to copyright. You can store them in your group project space, but they should not be shared as part of your open dataset.
If you are reusing others' published datasets as part of your meta-analysis, ensure that you are complying with any applicable licences on the original dataset, and properly cite that dataset.