Data Management Plan Template: **Systematic Reviews**

Abstract

This template provides general guidance for those who are undertaking systematic reviews. It is suggested that different team members contribute to the DMP based on the stage of the review process in which they will be involved in creating data.

For additional guidance and examples, please see the online research guide located at https://library.ucalgary.ca/dmpforsr.

The PRISMA-P for systematic review protocols is a tool that may help with planning and describing your data management process. It requires you to specify which databases you'll search, present a draft search strategy, report your inclusion/exclusion criteria, as well as procedures for all stages of your review. This information can be included in various sections of your DMP.

Administrative Details

Template Author(s): Heather Ganshorn, Zahra Premji, Libraries and Cultural Resources, Paul E. Ronksley, Cumming School of Medicine, University of Calgary

Published: April 9, 2021

DOI: 10.5281/zenodo.4663434

Contact: Portage Network - portage@engagedri.ca, portagenetwork.ca

Version:

Version	Date	Changes
1.0	2021-04-09	Formatted for inaugural publication.

Data Collection

What types of data will you collect, create, link to, acquire and/or record in each of the different stages of the systematic review?

For a systematic review (or other knowledge synthesis types of studies), data includes the literature you find, the data that you extract from it, 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.

Example: The types of data you collect may include: literature database records; PDFs of articles; quantitative and qualitative data extracted from individual studies; a document describing your study protocol or other methods.

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

If you plan to use systematic review software or reference management software for screening and data management, indicate which program 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 non-proprietary formats wherever possible at the end of the project. Read more about file formats: UBC Library or UK Data Service.

Examples:

Word, RTF, PDF, etc.: project documents, notes, drafts, review protocol, line-by-line search strategies, PRISMA or other reporting checklists; included studies

RIS, BibTex, XML, txt: files exported from literature databases or tools like Covidence

Excel (xlsx, csv): search tracking spreadsheets, screening/study selection/appraisal worksheets, data extraction worksheets; meta-analysis data

NVivo: qualitative synthesis data

TIF, PNG, etc.: images and figures

For more in-depth guidance, see “Data Collection” in the University of Calgary Library Guide.

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?

It is important to keep track of different copies or versions of files, files held in different formats or locations, and information cross-referenced between files. Logical file structures, informative naming conventions, and clear indications of file versions, all help ensure that you and your research team are using the appropriate version of your data, and will 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 Archive.

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.

Examples:

> ***Suggested format - PDF full-texts of included studies:***
> AuthorLastName_Year_FirstThreeWordsofTitle
>
> Example: Sutton_2019_MeetingTheReview
>
> ***Suggested format - screening file for reviewers:***
> ProjectName_TaskName_ReviewerInitials
>
> Example: PetTherapy_ScreeningSet1_ZP

For more examples, see "Data Collection" in the University of Calgary Library Guide.

Documentation and Metadata

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

Good documentation includes information about the study, data-level descriptions, and any other contextual information required to make the data usable by other researchers. Other elements you should document, as applicable, include: research methodology used, variable definitions, 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. Some of this should already be in your protocol. For specific examples for each stage of the review, see "Documentation and Metadata" in the University of Calgary Library Guide.

How will you make sure that documentation is created or captured consistently throughout your project?

Where will the process and procedures for each stage of your review 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.

Most systematic reviews will likely not use a metadata standard, but if you are looking for a standard to help you code your data, see the Digital Curation Centre's list of disciplinary metadata standards.

Storage and Backup

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?

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.

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

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.

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

If your meta-analysis includes individual patient-level data, you will require secure storage for that data. As most systematic reviews typically do not involve sensitive data, you likely don't need secure storage. A storage space such as Dropbox or Google Drive 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.

Preservation

Where will you deposit your data for long-term preservation and sharing at the end of your research project?

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.

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

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 converting files from proprietary to non-proprietary formats may lose information or affect functionality. If this is a concern, you can archive both a proprietary and non-proprietary version of the file. Always document any format changes between files when sharing preservation copies.

Sharing and Reuse

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

Raw data are 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.

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.

Examples of what should be shared:

- protocols
- complete search strategies for all databases
- data extraction forms
- statistical code and data files (e.g. CSV or Excel files) that are exported into a statistical program to recreate relevant meta-analyses

For more examples, consult “Sharing and Reuse” on the University of Calgary Library Guide.

Have you considered what type of end-user license to include with your data?

Licenses determine what uses can be made of your data. Funding agencies 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. You may also want to check the requirements of any journals you plan to submit to - do they require data associated with your manuscript to be made public? Note that only the intellectual property rights holder(s) can issue a license, so it is crucial to clarify who owns those rights.

Consider whether attribution is important to you; if so, select a license whose terms require that data used by others be properly attributed to the original authors. Include a copy of your end-user license with your Data Management Plan.

Example: There are several types of standard licenses available to researchers, such as the Creative Commons or Open Data Commons licenses. 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. More about data licensing: UK Data Curation Centre.

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

Choose a repository that offers persistent identifiers such as a DOI. These are persistent links that provide stable long-term access to your data. Ensure you put a data availability statement in your article, with proper citation to the location of your data. If possible, put this under its own heading. If the journal does not use this heading in its formatting, you could include this information in your Methods section or as a supplementary file you provide.

Example: The datasets analysed during the current study are available in the University of Calgary's PRISM Dataverse repository, [https://doi.org/exampledoi].

Responsibilities and Resources

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.

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.

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

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.

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

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?

Ethics and Legal Compliance

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

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. 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.

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

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 data.

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

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.