Data Management Plan Template:
Neuroimaging in the Neurosciences

Abstract

This Neuroimaging data management plan (DMP) template is designed to be completed in two phases: Phase 1 questions probe at a high-level, seeking information about the general direction of the study. Normally, researchers will be able to respond to phase 1 questions at the outset of a project.

Phase 2 questions seek greater detail. It is understood that these answers will often depend on the outcome of several steps in the research project, such as: a literature review, imaging protocol design and experimental design, or running multiple pilot subjects and interpreting the outcome. As these details become known, the DMP can and should be revisited. This approach underscores that DMPs are living documents that evolve throughout a research project.

Administrative Details

Template Author(s): Ted Strauss, McGill University

Published: April 9, 2021

DOI: 10.5281/zenodo.4673558

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

Version:

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

Phase 1

Phase 1 questions probe at a high-level, seeking information about the general direction of the study. Normally, researchers will be able to respond to phase 1 questions at the outset of a project.

Data Collection

Describe the types of data, and potential data sources, to be acquired during the course of your study.

Please explain, in particular:

- What type of neuroimaging modalities will be used to acquire data in this study? Ex: MRI, EEG.
- What other types of data will be acquired in this study? Ex: behavioural, biological sample.
- Approximately how many participants does the study plan to acquire images from?

Documentation and Metadata

How will you document your methods in order to support reproducibility?

For fellow researchers, a write-up of your methods is indispensable for supporting the reproducibility of a study. In preparation for publishing, consider creating an online document or folder (e.g. openneuro, github, zenodo, osf) where your project methods can be gathered/prepared. If appropriate, provide a link to that space here.

Storage and Backup

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

Planning how research data will be stored and backed up throughout and beyond a research project is critical in ensuring data security and integrity. Appropriate storage and backup not only helps protect research data from catastrophic losses (due to hardware and software failures, viruses, hackers, natural disasters, human error, etc.), but also facilitates appropriate access by current and future researchers. You may need to encrypt your data to ensure it is not accessible by those outside the project. For more information, see the University of Waterloo's Guideline for researchers on securing research participants' data.

Please provide URL(s) to any data storage sites. If your data are subject to strict rules governing human subjects and anonymity, then you may need an on-premise solution installed on your institution's server.

What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.)?

Preservation

How will you store and retain your data after the active phase of data collection? For how long will you need to keep your data?

Choices about data preservation will depend on the potential for reuse and long-term significance of the data, as well as whether you have obligations to funders or collaborators to either retain or destroy data, and what resources will be required to ensure it remains usable in the future. The need to preserve data in the short-term (i.e. for peer-verification purposes) or long-term (for data of lasting value) will influence the choice of data repository or archive. Tools such as DataCite's repository finder tool and re3data.org are useful for finding an appropriate repository for your data.

Sharing and Reuse

How will you share data from this study with the scientific community? How open can you make it? Describe whether you plan to share your data publicly, make it available in a repository with restricted access, or offer it by request only.

Most Canadian research funding agencies now have policies recommending or requiring research data to be shared upon publication of the research results or within a reasonable period of time. While data sharing contributes to the visibility and impact of research, it has to be balanced with the legitimate desire of researchers to maximise their research outputs before releasing their data. Equally important is the need to protect the privacy of respondents and to properly handle sensitive data.

What you can share, and with whom, may depend on what type of consent is obtained from study participants. In a case where some (or all) or the data analyzed was previously acquired (by your research team or by others), what you can share for this current study may also be dependent on the terms under which the original data were provided, and any restrictions that were placed on that data originally. Provide a copy of your consent forms and licensing terms for any secondary data, if available.

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.

Data management focuses on the 'what' and 'how' of operationally supporting data across the research lifecycle. Data stewardship focuses on 'who' is responsible for ensuring that data management happens. A large project, for example, will involve multiple data stewards. The Principal Investigator should identify at the beginning of a project all of the people who will have responsibilities for data management tasks during and after the project.

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

This estimate should incorporate data management costs expected during the project as well as those required for the longer-term support for the data after the project is finished. Items to consider in the latter category of expenses include the costs of curating and providing long-term access to the data. Some funding agencies state explicitly that they will provide support 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. OpenAIRE has developed a tool to help researchers estimate costs associated with data management. Access this tool here.

Ethics and Legal Compliance

Please provide the name and a web link for the research ethics board (REB) that is responsible for reviewing and overseeing the legal and ethical compliance of this study. Give the file identifier of the REB application.

Researchers must follow the policies and guidance of the research ethics board governing their institutions. There may be important differences across institutions. The Public Health Agency of Canada (PHAC) is responsible for setting standards and coordinating REBs across Canada. They provide 10 best practices for ensuring privacy of human participants:

- Determining the research objectives and justifying the data needed to fulfill these objectives
- Limiting the collection of personal data
- Determining if consent from individuals is required
- Managing and documenting consent
- Informing prospective research participants about the research
- Recruiting prospective research participants
- Safeguarding personal data
- Controlling access and disclosure of personal data
- Setting reasonable limits on retention of personal data
- Ensuring accountability and transparency in the management of personal data

In the context of neuroimaging research, “the potential identifiability of otherwise anonymous image files is of great concern to those in the field who are anxious to encourage electronic data sharing” (Kulynych, 2002). Please consult your REB for recommendations on how to prepare ethics protocols.

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

State how you will prepare, store, share, and archive the data in a way that ensures participant information is protected, throughout the research lifecycle, from disclosure, harmful use, or inappropriate linkages with other personal data. This may mean avoiding cloud storage services, placing data on computers with no access to the internet, or encrypting data that will be shared during the research project. For more information, see the Harvard Catalyst guidance about cloud storage (or Appendix A of this document).

Phase 2

Phase 2 questions seek greater detail. It is understood that these answers will often depend on the outcome of several steps in the research project, such as: a literature review, imaging protocol design and experimental design, or running multiple pilot subjects and interpreting the outcome. As these details become known, the DMP can and should be revisited. This approach underscores that DMPs are living documents that evolve throughout a research project.

Data Collection

Give details about the sources of data, equipment used, and data formats produced for your project.

Please explain, in particular:

- What is the make and model of the neuroimaging system? Ex: Siemens Prisma 3T.
- Can you describe the image acquisition paradigm and parameters being used in the study? Ex. MRI T1w MPRAGE.
- What is the total duration of the scanning sequence? Ex. 40 minutes.
- What file formats will your neuroimaging data be acquired in?
    - Proprietary file formats requiring specialized software or hardware to use are not recommended for preservation, but may be necessary for certain data collection or analysis methods. Use open file formats where possible, or at least industry-standard formats such as dicom, NIFTI, European data format (.edf), or the BrainVision data format (.eeg/.vhdr/.vmrk). Read more about file formats: UBC Library or UK Data Archive.
- Will the data be converted into other formats? Ex. NIFTI, BIDS, Minc.
- Does the study incorporate any data acquired externally?
    - No. New data acquisition only.
    - New data plus retrospective data from the same PI.
    - New data plus retrospective data from multiple sources.
    - Only retrospective data used in this study.
- If external data are used in this study, please provide details about the source of external data, and identifying coordinates (DOI, URL, citation).

What conventions, methods, and standards will be used to structure, name and version-control your files to help you and others better understand how your data are organized? In other words, what types of metadata are being stored alongside the acquisition data? Ex: BIDS, NIDM.

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. 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 Archive.

What anonymization measures are taken during data collection and storage?

“Within the framework of privacy protection, the degree of anonymization of the data is an important consideration and thus is an aspect incorporated in privacy regulations. Different rules apply to data, which are dependent on whether the data is considered personal data, fully anonymized or de-identified. Fully anonymized data has all personalized data removed, is given a separate identification code, and the key between the fully anonymized dataset and any path back to the original data is deleted such that it would be extremely difficult to trace the data back to an individual” (White et al., 2020). The technical steps for anonymizing neuroimaging data should be designed to achieve the level of privacy required by ethics protocols governing the study. See here for a selection of resources pertaining to anonymization.

Documentation and Metadata

What documentation will be needed for the data to be read and interpreted correctly in the future? Document key details of methods pertaining to data and metadata here.

- Does the study have an identifier (study ID) entered into the imaging console and other software? If so, enter the study ID here.
- Does the study use identifiers for participants, e.g. sub-002? If so, give an example of the subject ID format here.
- Are there any other codes or identifiers used in the study? If so, please identify and describe them here.

Storage and Backup

What form of encryption is used, if any, with data transfer and data storage?

If you are using a data management application to manage data, please name which system. Describe the features of the application that are important for this project in particular (ex. provenance tracking, versioning, QC, longitudinal design).

A data management application is a piece of software that stores data and helps to manage some aspects of the data and/or metadata collection, quality control, conversion, processing, reporting, annotation, and other functions. Some applications are designed specifically for the neuroimaging domain, e.g. LORIS, Braincode, while other applications can be used by any research discipline, e.g. XNAT, Redcap. In neurosciences, the term 'database' is sometimes used by convention to refer to data management applications. For the purposes of this question, an application is any software tool used to manage data acquisition or storage.

Preservation

What data will be preserved for the long-term?

In some circumstances, it may be desirable to preserve all versions of the data (e.g. raw, processed, analyzed, final), but in others, it may be preferable to keep only selected or final data (e.g. phantom scans and other diagnostic scans may not need to be preserved).

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

DataCite's repository finder tool and re3data.org are both useful tools for finding an appropriate repository for your data. Searches on re3data can be easily narrowed by discipline, such as this search for 'neurosciences.' There are also generalist repository services like Zenodo, OSF, Figshare, and Academictorrents.com. If your data is ready to be shared under an open license, then posting to an existing platform like openneuro.org, openfmri.org, nitrc.org, or portal.conp.ca could be a good solution.

Not all repositories offer long-term preservation options, so you may want to consult a repository's posted policies before deciding to deposit.

Indicate how you will ensure your data, and any accompanying materials (such as software, analysis scripts, or other tools), are preservation ready.

Consider using preservation-friendly file formats (open, non-proprietary formats), wherever possible. 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 Archive 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 Archive.

Many repositories cannot accept data that has not been anonymized or de-identified, making de-identifying and cleaning the data necessary steps towards long-term preservation. Always include supporting documentation that describes the anonymization and de-identification procedures carried out.

Sharing and Reuse

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

What type of repository or storage service are you considering as the host of your shared data?

You may wish to share your data in the same repository selected for preservation or choose a different one. DataCite's repository finder tool and re3data.org, recommended in the preservation section, are also useful to consult here. Scientific Data has some specific recommendations for neuroimaging repositories here.

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

If data will be shared with any collaborators, then it should have a data license that defines the terms of use. If the data will eventually be published to a data hosting platform, then a creative commons open data license would be applicable. Even though it is “open” the license can place important constraints on what kind of re-use is allowed. For example, you can restrict access to only non-commercial uses of the data, or you can require that credit be given. Click here for more about Creative Commons licenses.

If data will be shared on a more restricted basis, e.g. with a closed consortium of collaborators, then a custom data license and usage agreement will be needed. Please consult your institution’s research librarian or technology transfer office for assistance. Click here to access OpenAIRE’s guide “How do I license my research data?”

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

Possibilities include: data registries, repositories, indexes, word-of-mouth, publications. How will the data be accessed (Web service, ftp, etc.)? If possible, choose a repository that will assign a persistent identifier (such as a DOI) to your dataset. This will ensure a 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. The Digital Curation Centre provides a detailed guide on data citation. Some repositories also create links from datasets to their associated papers, increasing the visibility of the publications. Contact your Library for assistance in making your dataset visible and easily accessible (reused from National Institutes of Health, Key Elements to Consider in Preparing a Data Sharing Plan Under NIH Extramural Support [2009]).

Responsibilities and Resources

Describe your succession plan, indicating the procedures to be followed and the actions to be taken to ensure the continuation of the data management if significant changes in personnel occur.

Some examples of events to consider: replacement of principal researcher, change of in responsibility for any researchers or data managers, the departure of students who have finished projects associated with the research material described in this DMP.

Ethical and Legal Compliance

If human imaging data are acquired, how will the data be anonymized? Will any defacing techniques be used?

Give examples of the tools or software will you use to clean DICOM headers of personally identifiable information. E.g. PyDeface is a tool that can be used to strip facial structures from the brain. For more detailed de-identification guidance and recommended tools for all types of data, see Portage's De-identification Guidance.

If external data are used in this study, please provide the data license & data use agreement.

This can be provided as a link, a description, or as a full copy of appropriate documents.

Do any other legal, ethical, and intellectual property issues require the creation of any special documents that should be shared with the data, e.g., a LICENSE.txt file?

Appendix A. Guidance About Cloud Storage

Source:
https://catalyst.harvard.edu/pdf/regulatory/Guide_to_Technologies_Used_in_Research.pdf

What is a cloud service and storage technology?

Cloud service and storage generally refers to a set of technologies that enable the collection, and processing and storage of data through a set of services or infrastructure where a third-party vendor manages computing resources on behalf of a data customer. Cloud computing services are ever evolving. While there are many types of configurations, the cloud may best be understood as a diverse network of computing resources (e.g., servers, smart phones, PCs, tablets) linked through the internet, which may be used in concert to perform a given set of computing tasks. Through the internet, cloud services can leverage massive data centers and a variety of software capabilities around the world, enabling flexible, scalable, and interoperable access to data and data services from any location. Commonly used cloud service and storage companies include Google Cloud, Amazon Web Services (AWS), Microsoft Azure, Dropbox, Netflix, Flickr, Syncplicity for EMC, and Microsoft 365.

Prior to submitting an IRB application or amendment for research studies using cloud service and storage technology, the following risks and technology considerations should be addressed:

Information risks associated with cloud services and storage can arise because the technology functions over a wireless network and through a third-party platform, making the communication susceptible to wiretapping or interception of data. The sensitivity of the data being collected must be considered when determining the risks to subjects' privacy and confidentiality. Be conservative about storing critical information in the cloud; without an appropriate contract, you should only use cloud storage for information that can be replaced with little or no consequence. In determining the best servicer, consider effective management controls (e.g., oversight of third parties, adequate insurance, disaster recovery, etc.). Also, consider the possibility that another company might purchase the cloud servicers and how that would affect data stored in the cloud service provider (e.g., data ownership, disaster recovery, privacy policies, etc.). Assess the relevance of federal privacy regulations, federal laws, contractual obligations, and grant restrictions before moving institution-related files and data to any cloud server. For financial reasons, many cloud providers locate some of their servers outside the US. In this case, since you won't know the physical location of the servers on which a provider stores your information, you should exercise caution if any of the information you store in the cloud is subject to any international or export restrictions.

Important risks associated with cloud service and storage technology:

1. Data Ownership – Research data collected and stored on cloud service and storage technologies is typically owned and/or governed by the investigator's institution or by the sponsor of the research. Factors affecting 'ownership' status includes, among others, who contributed the data, agreements associated with data creation and distribution, contract terms, and intellectual property rights. A cloud service provider may include terms in its vendor contract or end user license agreement (EULA) that automatically transfers some or all ownership rights to the provider. Failure to properly review the contracts and agreements may result in unintentional forfeiture of intellectual property rights or inability to retrieve data. Understand the cloud vendor access and data rights. The two categories of cloud data are data created by the user before uploading it in the cloud and data created on the cloud platform itself.

2. Data Collection – Cloud service and storage technology collects data manually with wired or wireless access to the server. During collection, there may be risks to the confidentiality, integrity, and availability of data.

3. Data Access – Data may be accessed in different locations. If data is accessed on a personal device, additional risks may be considered (e.g., the terms of agreement for the personal device were accepted under personal terms not considering the use for research). Additional precautions should be in place to protect the information (e.g., password protected login, ability to logoff users after a set time, ability to lock access if password is entered incorrectly over a set amount of times, etc.). In some cases, users may opt in or opt out of services but by doing this, may sacrifice access to services and data. If opting into cloud service and storage, choose only the minimum services necessary, and limit the number of staff who can access the media. Cloud storage media may allow for password protected access and remote locking capabilities.

4. Data Storage – Data should be stored on the appropriate cloud storage technology specified to protect the sensitivity of the data, with appropriate access requirements. Data storage should not be done in personal accounts; you should set up new accounts specifically for the research study. Access rights should be defined for all folders and files in the cloud storage media (e.g., only select research staff have the authority to modify backup files). Remove necessary subject identifiers from data files, and encrypt data files. Identifiers should be stored in a physically separate and secure location from the data files, and associated with the data files through a key code that is also stored in a separate and secure location. There is also the risk of more information being stored than is necessary. Server ports should be actively monitored and secured as they pose a disclosure risk through the exposure on internet search engines (e.g., Google, Yahoo, etc.). The longer data are left unused in storage, the more likely unauthorized individuals outside the network can retrieve it. Regulations have requirements on how data can be accessed and where it can be stored. For example, it is not appropriate to store data regulated by the Health Insurance Portability and Accountability Act (HIPAA) or the Family Educational Rights and Privacy Act (FERPA) in DropBox or other cloud services.

5. Data Transmission – Data is transmitted when the cloud service and storage media account is accessed and a data transfer request is made (e.g., request to download, share files, etc.). Risks include access by unauthorized users, mishandling of data, and failure to log out from the media when no longer in use. Whenever possible, don't transfer files via email; instead use an encrypted USB or external drive. When using email, never use your personal email account, as it is not secure. Make sure your work email is set up to be secure. To ensure the security of data being transmission, you can type in the subject line "[send secure]" in front of the subject title. Never include PHI in the subject title; subject titles are not secure. Check the technology to see if the channels are encrypted, secured, and how often the software is updated and patched. Whenever possible, don't transfer files via email; instead use an encrypted USB or external drive. When using email, never use your personal email account as it is not secure. Make sure your work email is set up to be secure. To ensure data being transmitted is secure, you can type in the subject line "[send secure]" in front of the subject title to make the email secure. Never include PHI in the subject title, the subject title is not secure.

6. Data Sharing – Data sharing on cloud service and storage technology should be limited by proper access controls (e.g., password protection, encrypted files, etc.). Prior to sharing data, ensure the location and method for sharing is secured and protected based on the sensitivity of the data. A cloud provider may be configured to include protections, but if the researcher downloads or syncs that data to their end device such as a laptop or smartphone, the device may not be secure.

7. Data Retention and Destruction – The duration of time the data will be stored on the cloud service and storage media should be taken into account when determining the risks. The media should be reviewed to determine if reading is possible while the data is being stored. This is especially critical for long-term storage or archiving. Account for the fact that cloud storage media products can have varied shelf lives, and may become obsolete. Depending on the cloud storage company policy, they may have rights to the data, including how the data will be retained and destroyed.