DMP title

Project Name Metabolomics and IBS/IBD Study

Principal Investigator / **Researcher** Megan Magee

Project Data Contact mageem@mcmaster.ca

Description My research project involves the study of the metabolomic profile of IBS, Chrones, and UC patients pre-, post and during the fecal transplantation. Data will be collect on the key metabolic biomarkers of disease progression and recovery throughout the project to determine the effectiveness of fecal transplants as a treatment method for GI diseases

Institution Portage

Data Collection

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

I will create instrumentation data as well as numerical excel data in relation to analyzed instrumental data. I wil be aquiring textual as well as numerical/instrumental data from past and current researchers studying similar topics.

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?

My data will be collected in proprietary file formats associated with the software available on our instrumentation. The files will be converted/ transfered to more readily available software such as excel as it is an open file formate accessable by most.

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?

Data Files:

Creation Date_Owner Name_(Experiment) Description_Run #

e.g MMDDYY_Meg_BorateBGEAssay_001

Paper/Thesis Written Files:

Owner Name_Project Description_Version #

e.g. Megan_FecalTransplantationBiomarkers_Ver-01

Documentation and Metadata

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

A General document containing the naming conventions used and how/where data is stored will be created for making my data easy to find.

In order to make my data usable by others, clear documentation and recording using R software will be kept in order to track data/documentation creation.

Documentation will need to be detailed and easy to understand while providing any information needed to understand the project/experiment.

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

Documentation will be captured consistanly by making sure to use and follow my data Management plan

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

N/A

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?

The storage requirements for my project can range from 2-6 terabytes and will be stored for an undefined period of time. this data may be needed in the future for data reproducibility studies and thus does not have a set date or time that it will no longer need to be stored.

This question may be updated or changed in the future when the DMP is revisited throughout the research project's progression.

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

The data will be stored on an 8 terabyte harddrive, along with the desktop and a USB stick. Data from the desktop will be backed up on to an external hardrive along with a cloud service (e.g. Mac Drive). this method follows the 3-2-1 backup rule with the three copies being on desktop, UBS, cloud, and external hardrive with several different media sources in use and the cloud serice being an offsite storage location.

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

From my current knowledge I will not be having any contributers on my specific project. I have not yet begun my project, so this information will have to be determined/confirmed at a later date.

This question may be updated or changed in the future when the DMP is revisited throughout the research project's progression.

Preservation

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

After the project is completed the processed data can be shared and stored on the

website GitHub so that it is both retained and accessable to future researches.

Anonymized personally identifying data will not be made accessable as it is restricted information.

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.

To ensure my data is preservation ready, the file formats data will be stored in will be converted/transfered to Excel (.xls) as it is an easily acessable/understandable data file format that can be opened by anyone with basic microsoft office capabilities.

Excel may be constantly updating their software, but the excel files are not being transformed into any other file types so coversion of the file between current and future excel versions should not pose a risk for creating errors in the data. Patient data will either be anonomous or not shared

Sharing and Reuse

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

Our Raw data generated by the instrument is transfered to an excel file or saved as its proprietary data file format on the desktop for further analysis and data processing.

Raw data will be processed to that only peaks or data that is relavent to what is being studied is displayed and so the data is readable and more easily understood by veiwers. processed data will be available for sharing after the project has been comleted along with the analyzed and final data files.

Analysed data such as the statistical analysis of our data clusters will be shared as an R file as well as being outlined in the final report (Final data).

Before completion of the study, data will only be shared internally and with collaborators.

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

General liscence requirements state that after one year papers have to be open acess or available to public. The specificis on liscences are dependent on the journal that the data is being published in.

This question may be updated or changed in the future when the DMP is revisited throughout the research project's progression.

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

Data will be published in journals and shared via word-of-mouth. Data will also be available in the GitHub repositiry where is can be accessed by other researchers.

This question may be updated or changed in the future when the DMP is revisited throughout

the research project's progression.

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.

During the project, data managment will be done by myself. Collaborators may be responsible for handelling client data if they are the data providers. This would be so that patient data is kept anonymous.

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?

Once I have completed my project and have completed my master's program, the student that will replace me, along with my supervisor or any other collaborators will be given the data and informed on how the data has been stored so that they may continue to preserve the data untill it becomes no longer necessary.

upon the unlikely event that the project's principle investigtor must be changed without fore warning, all data will be accessable for inheritance via hard drive/lab notebooks that are kep in the lab group's possession.

This question may be updated or changed in the future when the DMP is revisited throughout the research project's progression.

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

The resources required for implementing my data managment plan include file storage and back up devices. this includes external hard drives and cloud services to ensure data is stored in a way that it is both safe and recoverable.

In order to save files into easily acessable formats such as excel, office applications are required. These applications are offered free to students at th unviversity.

The overall cost for data managment should be ~ $200.00 for the external hard drive and USB stick

This question may be updated or changed in the future when the DMP is revisited throughout the research project's progression and as other costs arise.

Ethics and Legal Compliance

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?

Sensitive data, such as client/patient data will be stored on a desktop computer or hard drive that is password protected to ensure that the data cannot be accessed by anyone not directly granted access or are not an approved member of the project.

Specific example of how sensitive data should be handeled can be found at https://www.cmpa-acpm.ca/en/advice-publications/browse-articles/2013/protecting-

patient-health-information-in-electronic-records (this is not neccessarilt an accurate description for my project but provides some examples of common practice for reference)

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

From my current understanding, sensitive data will not be distrubuted by our group to be used by secondary parties. The data recieved and used by our reserach group will likely come without any personal identifiers so that data cannot be directly linked back to a specific patient.

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

These details of my project have yet to be discussed. Once the project has been ut into motion this information will be filled out. These issues will likely need to reflect policys put into place by McMaster University.

This question may be updated or changed in the future when the DMP is revisited throughout the research project's progression.

This document was generated by DMP Assistant (https://assistant.portagenetwork.ca)