Activity 4 on Research Data Management Principles

Chinyere Ruth Ottah

GEOG 712 Reproducible Research

Professor: Antonio Paez

School of Earth, Environment & Society

McMaster University

Date: October13, 2023

Project Data Analysis

List the data that you expect to use, collect or create in your project. Identify if you are generating or collecting the data and if you are using existing datasets.

For my project, I will be utilizing a combination of existing datasets and conducting fieldwork to collect additional data. The variables I intend to use, generate, or collect are as follows:

Variable Type	Data	Data generated if applicable	Data Originator
Response varaibles	Field combustion data for soil, belowground and aboveground	I will be using an existing dataset, and new filed data I intend to collect	
Explantory variables	Elevation	Slope, Aspect, Topgraphic Wetness Index	I collected the elevation data and generated the derivatives
	Climate Variables	Tempetaure and relative hunidity	I generated the data
	Soil	Soil types	I generated the data
	Landsat Data band 1 - 8	Landsat composite, Normalised Burn Ratio, Tasselled cap indicies, NDVI	I generated the data
	Tree Cover	Field data collection	I collected the data

Are there any Legal and Ethical restrictions that you will need to address?

There are no ethical or legal restrictions for the data used in this project as no personal or sensitive information is involved.

Go through the Quick Hits for Data Management and identify possible strategies to build and protect the value of your data.

`- Where will you keep raw data and how will you back it up?`

During the research process, my raw data will be compressed into a zip file. I will adhere to the 3-2-1 rule for data storage, ensuring that I maintain three copies of my raw data. These copies will be distributed across different geographic locations: my computer, my OneDrive account, and my GitHub repository. Additionally, I will create a backup copy that will be stored offsite on my external hard drive. The data will be backed up regularly, and if necessary, this process will occur automatically on my online account. Furthermore, I will utilize checksums, which are a type of free software designed to detect discrepancies during data transmission.

What file formats do you anticipate your data will be in? Are the formats open or can they be converted to open formats?

All the data I've gathered to derive my explanatory variables is in GEOTIFF format. Meanwhile, the existing field data, encompassing aboveground, belowground, and soil combustion measurements, is stored in a proprietary CSV format. All the data used for this study will be open format.

Create a File naming convention for your project dataRaw Data Storage

All the spatial information that I collected will be arranged by their category and the type of data. Additionally, the names of each file will not exceed 32 characters, and there will be no special character usage. For example, the naming convention for my data will include:

Carbonloss_2023_Canada which will first signify the project name

Followed by Predictor_variables listed below

Landsatbands_MayAug2023_Canada.tif

Der_Landsatbands_MayAug2023_Canada

Landsatcomp_MayAug2023_Canada.tif

NDVI_MayAug2023_Canada.tif

NDII_MayAug2023_Canada.tif

Digital elevation

Der_DEM_MayAug2023_Canada.tif

Der_Slope_MayAug2023_Canada.tif

Aspect_MayAug2023_Canada.tif

Response_variables

BGcombustion_MayAug2023_Canada.csv

AGcombustion_MayAug2023_Canada.csv

Soilcombustion_MayAug2023_Canada.csv

This folder has also been created in my Github repository

What standards are relevant to your project? List any existing standards or best practices in use in your field or in your lab? This could include instrument procedures or file management standards. What standards might you want to create to help you to manage yourdata?

The standards relevant to my project include data format and entry, as well as field type definitions. Some of the formats I will use to manage my data include:

Data Format and Entry: - Dropdown lists (e.g., spreadsheets) and ArcGIS field maps will be used for entering field data. This will allow for data sharing and interoperability among users in the field. - All images obtained must be saved in GeoTIFF format. - Related information must be stored under the category's feature datasets. - A unique identifier must be used for each field sample collected. An example of this unique identifier is AB_1 for Alberta sample point 1, and AB_2 subsequently for the next sample.

Field Data Definitions: - Each field site and its corresponding information must be labeled with an underscore (_) and not a space. - Date, month, year, climate, and soil observations during collection should be included in the attribute information. - Field type information and basic spatial elements for each sample location must be defined. Examples of field type information include text and date, while spatial elements include point, line, and polygon features.

List possible strategies you might use to document your data throughout your project.

To effectively document my data, I will employ a structured approach centered around addressing key questions: what, why, and who.

For the "what" question, this involves:

1. Providing a comprehensive description of the file name utilized.
2. Enumerating the data elements for all fields and acquired images.
3. Designating missing data with a designated "nodata" value, identified by the code 9999.
4. Explicitly stating the data type.

Moving on to the "who" question, this encompasses:

1. Furnishing pertinent information about the individuals responsible for data collection, including both myself and my lab mates.
2. Compiling a roster of the contributors involved in processing and generating derivatives for the carbon combustion modeling.
3. Supplying contact details such as email addresses or phone numbers for inquiries, further information, or potential collaborations.

Addressing the "why" and "where" questions entails:

1. Elucidating the rationale behind gathering combustion data and its significance .
2. Specifying the precise location (longitide, latitude, pronvince and city)where field data was gathered.
3. Documenting the spatial reference of the data for geospatial context.