Oleson Lab Reseach Project Resource

Alemarie Ceria

2024-01-30

Table of contents

Overview

This resource seeks to provide you with the following:

> **Interactive Code Annotations**
>
> Throughout this resource, hover over the numbered annotations to the right of code chunks to reveal detailed explanations and comments about the code.

Software and Platforms

Download/Sign Up Links

Software/Platform	Purpose	Download/Sign Up Link
Slack	Enhances team communication and project coordination	Windows Link Mac Link
Google Drive for Desktop	Streamlines file storage, sharing, and collaboration	Link
R	Provides tools for statistical analysis and data visualization	Link
RStudio	Facilitates R coding, debugging, and project organization	Link
Git	Manages version control for tracking and merging changes in code and documents	Link
GitHub	Hosts and manages Git repositories, facilitating collaboration and code sharing	Link
Zotero	Organizes and cites research sources consistently	Link

Setup Instructions

Project Organization

Naming Conventions

https://www.youtube.com/watch?v=ES1LTlnpLMk&list=PLy_EwS4oOnoQzWYaWDO_x5t7O1RqHwuhw&index=5&pp=gAQBiAQB

TL;DR:

File and Directory Naming Best Practices:

1. **Machine Readability:**
   - Use globbing with regular expressions for pattern matching in scripting languages (e.g., Bash).
   - Example: `report-2020-*.txt` to find files matching a pattern.
2. **Naming Rules:**
   - Avoid spaces and accented characters.
   - Use only hyphens (-) and underscores (_) for punctuation.
   - Use lowercase letters exclusively.
3. **Human Readability:**
   - Employ highly informative slugs for filenames (e.g., `01_collect-cdec-snow-data.R`).
   - Ensure slugs are descriptive, concise, and in plain language.
   - Use hyphens to replace spaces.
4. **Metadata Separation:**
   - Use underscores to separate metadata for parsing in scripting languages.
   - Example: `01_report-metadata.txt` for clear metadata separation.
5. **Logical Sorting:**
   - Apply left padding (e.g., 01, 02) for sequential order.
   - Useful for predefined sequences or steps in a process.
6. **Chronological Sorting:**
   - Follow the ISO 8601 date format (YYYY-MM-DD), e.g., `2020-12-01_draft.docx`.
   - Suitable for tracking updates/progress or changes over time.

These practices ensure both machines and humans can easily work with your files and directories, promoting organization and clarity in your projects.

Type	Naming Convention	Example

Directories

Local - Shared Google Drive

Example Project Directory Structure

```
/g/path/to/example-project-directory                    (1)
  deliverables                                          (2)
    papers                                              (3)
  github-repo                                           (4)
    example-project-github-repo
  meetings-and-events                                   (5)
    meeting-notes                                       (6)
    agendas                                             (7)
    workshop-or-event-planning                          (8)
  presentations                                         (9)
  project-materials                                     (10)
    methods-drafts                                      (11)
    lit-review                                          (12)
  README.md                                             (13)
```

(1) Root directory housing all materials and documentation for the research project.
(2) Contains final outputs like reports, papers, or products resulting from the project.
(3) Stores final versions and supplementary materials for academic papers or reports.
(4) Specific GitHub repository directory for collaborative development and code sharing. *Only `example-project-github-repo/` will be pushed to GitHub.*
(5) Organizes documentation and planning materials for meetings and events related to the project.
(6) Archives notes and decisions from project meetings.
(7) Prepares and stores agendas for upcoming meetings to structure discussions.

⑧ Holds planning documents, schedules, and resources for workshops or project-related events.
⑨ Contains slides, speaker notes, and related materials for presentations about the project.
⑩ Stores various project-related documents not categorized elsewhere.
⑪ Keeps drafts and notes on methodological approaches and procedures.
⑫ Compiles literature reviews, reference materials, and bibliographies.
⑬ Provides an overview of the project directory, explaining the structure and contents of the folders/files.

Create New Project Directory

The following code creates a directory for your project within the `current-projects/` directory in the `Oleson Lab/` Shared Google Drive by making a copy of an existing template.

To execute this code chunk, please do the following:

1. Ensure you have Google Drive for Desktop software downloaded, you are signed in, and are able to see the `Oleson Lab/` directory in your `Shared Drives/` directory
2. Change `#| eval: false` to `#| eval: true`
3. Input your desired project name in the code `project_name="example-project"`.

````
```{bash}
#| label: create-new-project-directory
#| code-overflow: wrap
#| eval: false # <1>

echo "Starting script..." # <2>

cd "/g/Shared drives/Oleson Lab/projects/project-management-resources" # <3>

new_project_template="/g/Shared drives/Oleson Lab/projects/project-management-resources/new-p

project_name="example-project" # <4>

new_current_project_directory="/g/Shared drives/Oleson Lab/projects/current-projects/$project

# Check if the directory already exists
if [ -d "$new_current_project_directory" ]; then # <6>
    echo "Error: A project with the name '$project_name' already exists. Please choose a dif
    exit 1 # <6>
fi # <6>
````
````

```
# Copy the new-project-template directory to the new project directory
cp -r "$new_project_template" "$new_current_project_directory" # <7>

echo "Completed copying." # <8>
```

① **To run code chunk, set `false` to `true`.**
② Confirms that the script is initiated.
③ Navigates to directory containing the new project template.
④ Defines a variable `new_project_template` which stores the `new-project-template/` directory file path.
⑤ Sets project_name to desired project name and constructs a path for the new project in the `current-projects` directory. **Replace `[insert-project-name]`.**
⑥ Checks to see whether or not the directory exists within the `current-projects/` directory and if so, prompts you to choose another name.
⑦ Copies everything within the `new-project-template/`directory into the `current-projects/` directory and renames it accordingly.
⑧ Confirms that the script ran successfully.

Remote - GitHub Repository

Example GitHub Repository Structure

```
/g/path/to/example-project                                    ①
  outputs                                                     ②
    tables                                                    ③
    plots                                                     ④
  documentation                                               ⑤
  code                                                        ⑥
    functions                                                 ⑦
    models                                                    ⑧
  data                                                        ⑨
    exploratory                                               ⑩
    raw                                                       ⑪
    tidied                                                    ⑫
    output                                                    ⑬
  LICENSE                                                     ⑭
  .gitignore                                                  ⑮
  example-project.Rproj                                       ⑯
  README.qmd                                                  ⑰
  README.md                                                   ⑱
```

① Serves as the root directory, encapsulating all project components for easy management and navigation.
② Contains generated files like tables and plots, separating results from the input data and code.
③ Stores tabular results and data summaries.
④ Holds visualizations and graphs generated by the analysis.
⑤ Stores project documentation, reports, and notes, centralizing information for reference and clarity.
⑥ Contains all the scripts and code used in the project, promoting modularity and code reuse.
⑦ Stores custom functions to ensure code modularity and readability.
⑧ Holds existing and output models.
⑨ Acts as the main repository for all datasets, organized to reflect different stages of data processing.
⑩ Contains initial explorations and analyses, fostering a sandbox environment for preliminary insights. (Will not be pushed to GitHub)
⑪ Stores unmodified raw data, preserving the original datasets for reproducibility and reference.
⑫ Holds processed and cleaned data, ready for analysis, ensuring consistency and reliability. (Will not be pushed to GitHub)
⑬ Stores the final dataset used for analysis.
⑭ Holds processed and cleaned data, ready for analysis, ensuring consistency and reliability.
⑮ Specifies the terms under which the project can be used or distributed, clarifying legal and usage aspects.
⑯ Lists files and directories to be ignored by version control, keeping the repository clean and relevant.
⑰ Provides a Quarto-rendered, detailed project overview and instructions, enhancing comprehension and usage.
⑱ Output of `README.qmd` used to be displayed on GitHub Repository.

Acknowledgements

We crafted our directory structures by drawing from a blend of best practices as outlined in the Environmental Markets Lab's (emLab) file structure guide and the valuable insights shared in the Riffomonas Project's project organization tutorial.

Data

Code

Reports and Publications

More Resources

DMP and SOP:

- WA-Department-of-Agriculture / washi-dmp
  - Washington Soil Health Initiative: State of the Soils Assessment Data Management Plan
- emLab Standard Operating Procedures
- Data Management for Scientists by David LeBauer

Documentation:

- The Documentation System
  - What nobody tells you about documentation
- The University of Arizona Data Documentation Resource
- genophenoenvo/neon-datasets/README.md Example

Videos:

- Reproducible Research YouTube Playlist