Procedure and Components of Reproducible Research Project

https://github.com/HEGSRR/RPI-Spielman-2020

1. Start a new repository with https://github.com/HEGSRR/HEGSRR-Template

2. Customize the `readme.md` and `LICENSE` with project information

3. Populate either the python notebook or Rmarkdown file with an *analysis plan*

4. Use the `metadata_template.md` template to create a metadata document for each data layer input and saved output

5. B*efore* you start analyzing data
   - render/knit analysis plan to pdf
   - register on OSF

6. Update the `readme.md` header and index tables `data_index.csv`; `procedure_index.csv`

7. Implement analysis in Jupyter python notebook or Rmarkdown computational notebook file(s). With code and results, your analysis plan becomes an analytical report.

8. Document and manage the computational environment: in R use `groundhog`; or in Python use a Conda `environment.yml` file, `requirements.txt` file, or reference the CyberGIS system and kernel type

9. Add results, discussion, and conclusion sections

10. Render final study report to the `docs` folder and register on OSF

11. Final round of checking and revising `readme.md` and metadata documents, `data_index.csv`, `procedure_index.csv` and `results_index.csv`

Compendium Structure

```
Root\
  .gitignore                ignore unpleasant files and private folders
  .here                     path names relative to project root
  CITATION.cff              citation info for GitHub.com users
  LICENSE                   open access license
  readme.md                 project metadata, index and links

  data\
    data_index.csv          table of contents for data
    metadata\               store all data documentation
      metadata_template.md  template ISO metadata
    raw\                    unmodified original data
      private\              restricted by size, license, or ethics
      public\               unrestricted
    derived\                processed & analyzed data
      private\              restricted by size, license, or ethics
      public\               unrestricted
    scratch\                untracked working folder

  docs\                     GitHub.com can render webpages from here
    report\                 store preregistrations, post-analysis reports
      analysis_plan.md      analysis plan template, includes metadata for
                            project and for data sources
    presentation\           slides, interactive data visualizations, etc.
    manuscript\             store LaTeX manuscript

  procedure\
    procedure_index.csv     table of contents for procedure(s)
    code\                   codes, scripts, computational models
      01_Jupyter.ipynb      template Jupyter notebook & analysis plan
      01_R-Markdown.Rmd     template Rmarkdown notebook & analysis plan
    environment\            requirements.txt or environment.yml file
    protocols\              non-code protocols, e.g. survey instruments

  results\
    results_index.csv       table of contents for results files
    figures\                graphic results for manuscripts, presentations
    other\                  other multimedia outputs
    tables\                 data table outputs
```

Project Metadata

Study Metadata

Key words: Comma-separated list of keywords (tags) for searchability. Geographers often use one or two keywords each for: theory, geographic context, and methods.
Subject: select from the BePress Taxonomy
Date created: date when project was started
Date modified: date of most recent revision
Spatial Coverage: Specify the geographic extent of your study. This may be a place name and link to a feature in a gazetteer like GeoNames or OpenStreetMap, or a well known text (WKT) representation of a bounding box.
Spatial Resolution: Specify the spatial resolution as a scale factor, description of the level of detail of each unit of observation (including administrative level of administrative areas), and/or or distance of a raster GRID size
Spatial Reference System: Specify the geographic or projected coordinate system for the study
Temporal Coverage: Specify the temporal extent of your study---i.e. the range of time represented by the data observations.
Temporal Resolution: Specify the temporal resolution of your study---i.e. the duration of time for which each observation represents or the revisit period for repeated observations
Funding Name: name of funding for the project
Funding Title: title of project grant
Award info URI: web address for award information
Award number: award number

Related to

OSF Project: doi links…
Pre-analysis Registration:
Post-analysis Report Registration:
Preprint:
Conference Presentation:
Publication:
Prior Study:
...:

Metadata for access

Rights: LICENSE: BSD 3-Clause "New" or "Revised"
Resource type: Collection
Resource language: English
Conforms to: Template for Reproducible and Replicable Research in Human-Environment and Geographical Sciences version 1.0, DOI:10.17605/OSF.IO/W29MQ

Data Metadata

- `Title`: Title of data source
- `Abstract`: Brief description of the data source
- `Spatial Coverage`: Specify the geographic extent of the data. This may be a place name and link to a feature in a gazetteer like GeoNames or OpenStreetMap, or a well known text (WKT) representation of a bounding box.
- `Spatial Resolution`: Specify the spatial resolution as a scale factor, description of the level of detail of each unit of observation (including administrative level of administrative areas), and/or or distance of a raster GRID size
- `Spatial Reference System`: Specify the geographic or projected coordinate system for the study
- `Temporal Coverage`: Specify the temporal extent of the data---i.e. the range of time represented by the data observations.
- `Temporal Resolution`: Specify the temporal resolution of the data---i.e. the duration of time for which each observation represents or the revisit period for repeated observations
- `Lineage`: Describe and/or cite data sources and/or methodological steps taken or planned to create this data source, e.g.:
    - sampling scheme, including spatial sampling
    - target sample size and method for determining sample size
    - stopping criteria for data collection and sampling (e.g. sample size, time elapsed)
    - de-identification / anonymization
    - experimental manipulation
- `Distribution`: Describe who will make the data available and how?
- `Constraints`: Legal constraints for *access* or *use* to protect *privacy* or *intellectual property rights*
- `Data Quality`: State any planned quality assessment
- `Variables`: For each variable, enter the following information. If you have two or more variables per data source, you may want to present this information in table form (shown below)
    - `Label`: variable name as used in the data or code
    - `Alias`: intuitive natural language name
    - `Definition`: Short description or definition of the variable. Include measurement units in description.
    - `Type`: data type, e.g. character string, integer, real
    - `Accuracy`: e.g. uncertainty of measurements
    - `Domain`: Expected range of Maximum and Minimum of numerical data, or codes or categories of nominal data, or reference to a standard codebook
    - `Missing Data Value(s)`: Values used to represent missing data and frequency of missing data observations
    - `Missing Data Frequency`: Frequency of missing data observations: not yet known for data to be collected