

# Project Template Structure

README, Data, Manifests, and SRC are required for all projects. Other directories may be used as needed.

**README** files are used at the project and directory levels to provide context. A project level README is a required component of the template.

**Data** directory holds all data collected and used in the research effort.

- Raw: original, unmodified data
- Interim: temporary location for outputs of data processing and analysis
- Endpoints: final location for outputs of data process and analysis. Once pipelines or analysis steps are cemented, final output data should be in endpoints
- Ref-data: any external or public data sets not supplied by CHOP resources or your team

**Manifests:** data inventory files that maintain participant mappings and context for all data files. If working within Arcus, you will not need to create these yourself

**Src:** contains information about access tools required to work with the research itself.

- Notebooks: computational notebooks such as Jupyter, Beaker, Zeppelin
- Rules: workflow summaries
- Scripts: custom scripts
- Tests: unit tests
- Further subdirectories can be customized

```
/home_dir/project_name/  
├── README.md  
├── configs  
│   └── README.md  
├── data  
│   ├── README.md  
│   ├── endpoints  
│   ├── interim  
│   ├── raw  
│   └── ref-data  
├── manifests  
│   ├── README.md  
│   ├── file_derivation.csv  
│   ├── file_manifest.csv  
│   ├── participant-crosswalk.txt  
│   ├── participant_family_role.csv  
│   └── participant_manifest.csv  
├── models  
│   └── README.md  
├── references  
│   ├── README.md  
│   └── protocols  
├── reports  
│   ├── README.md  
│   ├── figures  
│   ├── log.md  
│   ├── methods.md  
│   └── tables  
├── requirements  
│   ├── README.md  
│   └── project_requirements.txt  
└── src  
    ├── README.md  
    ├── notebooks  
    │   └── README.md  
    ├── rules  
    ├── scripts  
    └── tests
```

**Configs:** configuration files for workflows or applications

**Models:** machine learning models, model predictions, model summaries, datasheets for model training data: <https://arxiv.org/abs/1803.09010>

**References:** contains general information about the research effort such as IRB documents, reference papers, sample information, lab prep information, data dictionaries.

- Protocols: experimental protocols
- Further subdirectories can be customized.

**Requirements:** any module or library dependencies for workflows. Additional requirements files can be added as needed

**Reports:** content used for producing papers, presentations, websites, metrics, etc.

- Figures & tables: generated metrics and graphics for supporting reports
- Log.md: computational notebook (if one was used to create the content)
- Methods.md: version controlled methods section for the project
- Further subdirectories can be customized