Informatics team manual of procedures

Ania Tassinari & Caitlin Guccione

2019-08-02

Contents

1 About

This is a manual of operations for the Agios Informatics team. Its puspose is to:

- provide a resource on **best practices** (**??**)
- aid in setting up effective and reproducible **project workflows** (**??**)
- promote learning and sharing of **ideas** (**??**).

2 Best practices

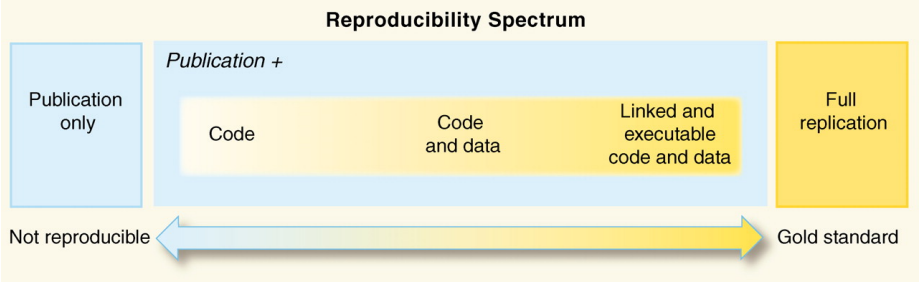
2.1 Reproducible research

2.1.1 Purpose

1. Improve collaborative analyses:

- make sharing easier
- enable retrieval and interpretation of results long after analysis ended

2. Simplify hand-off to Biostats
3. Improve confidence in our data and results



Source: Peng et al., *Reproducible Research in Computational Science*. Science 2011.

2.1.2 DO's and DON'T's of reproducible research

- DO start with good science
- DON'T do things by hand
  - Was any part of this analysis done by hand?
    - If so, are those parts precisely documented?
    - Does the documentation match reality?
- DON'T point and click
- DO teach a computer
- DO use version control
- DO keep track of your software environment
- DON'T save any output (until it's time to write a paper)
- DO set your seed

Source: Reproducible Research at Coursera

2.2 Version control

2.3 Code guidelines

2.3.1 R

- Tidyverse Style Guide
- Google's R Style Guide

3 Project Workflows

3.1 Using Workflowr

3.1.1 Quick Start

This section is a quick version setting up workflowr, for more clear or specific instructions skip to The Full Guide to Using Workflowr.

3.1.1.1 Set Up

In the `Console` tab of RStudio make sure you are in `(None)` project:

```
install.packages("workflowr")
library("workflowr")
wflow_git_config(user.name = "First Last", user.email = "first.last@agios.com")
```

Click here for more specfic details on **set up**

3.1.1.2 Creating Projects

In the `Console` tab,

```
wflow_start("PROJECT_NAME")
wflow_build()
wflow_publish(c("analysis/*.Rmd"), "Publish the initial files for PROJECT_NAME")
```

Click here for more specfic details on **creating projects**

3.1.1.3 Connecting to GitLab

In the `Console` tab,

```
wflow_use_gitlab(username = "first.last", repository = "PROJECT_NAME", domain = "ceres.agios
```

Go to your Agios GitLab and do the following:

- Create a project in GitLab with the same name as the project in RStudio
  - We called our project: PROJECT_NAME
- Scroll down to the push an existing Git repository option
  - Copy everything in the box besides the first line (`cd existing_repo`)
- Make sure you are in PROJECT_NAME directory
  - Paste what you just copied from Git into the `Terminal` tab in RStudio

Click here for more specfic details on connecting to **GitLab**

3.1.1.4 Creating a New File

In the `Console` tab,