

Organization part 2: Methods, Protocols, & Code

Managing your Research for Reproducibility: Week 3

Ibraheem Ali, PhD

October 17, 2022

Hello!

I am Ibraheem Ali

Program Manager for Research
Informatics, UCLA

PhD, Biomedical Sciences UCSF
Post-Doc, Gladstone Institutes

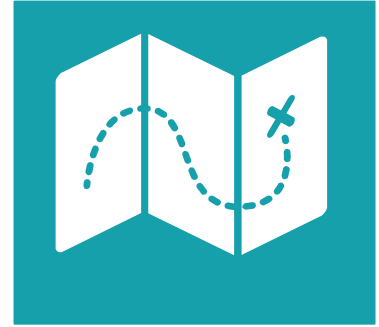


Course Overview

7 Sessions from Oct 3 - Nov 14:

- ▷ Oct 3: Introduction + Course Overview
- ▷ Oct 10: Organization Pt 1: Data and Documentation
- ▷ **Today: Organization Pt 2: Methods, Protocols + Code**
- ▷ Oct 24: Data Sharing & Repositories
- ▷ Oct 31: Publishing & Open Access
- ▷ Nov 7: Authorship & Peer Review
- ▷ Nov 14: Tracking and Getting Credit for your Work

Session Learning Outcomes



- ▷ Distinguish between protocols and methods
- ▷ Identify best-practices for managing & updating protocols
- ▷ Define computational reproducibility
- ▷ Recognize reusable published code with sufficient annotation
- ▷ Cite software or code you use in your research

The difference between methods and protocols

Methods are *what* you do,
Protocols are *how* you do it.

Hunting for research methods and analysis strategies can be a headache

Types of Headaches

Migraine



Hypertension



Stress



Method as previously described...



- ▶ Method as reported in “Not Methods et al. 2020”
- ▶ Method “as previously described”

... *was it really?*

- ▶ Using manufacturer’s specifications, with minor modifications

... *what modifications??*

Some helpful definitions

Research Method - the approach used to investigate a research question

- Next Gen Sequencing
- Drug Screening
- Overexpression
- Knock-out

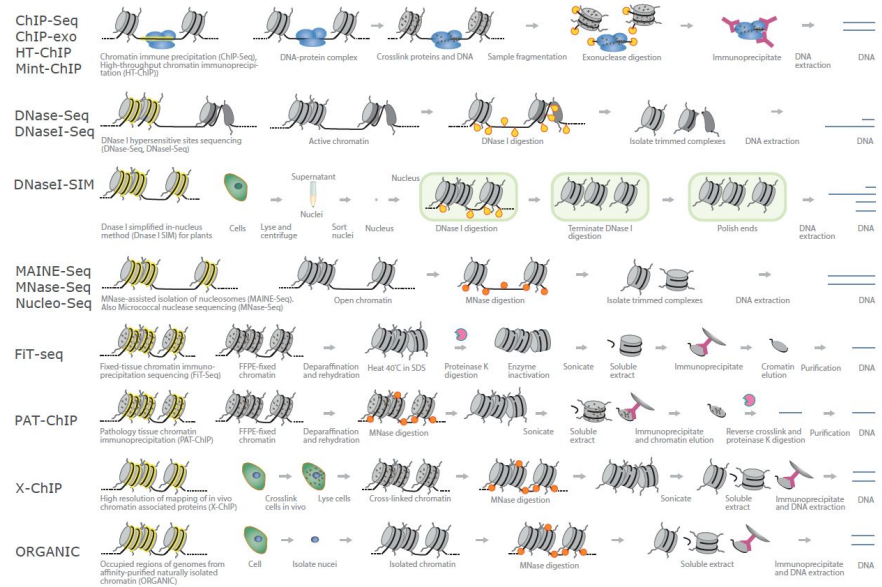
Research Protocol - the step-by-step process used to obtain experimental results

- Instruments used
- Incubation times/temps
- Constructs used
- Reagents used

You use Next Generation Sequencing (NGS), you say?

- NGS is actually a collection of research methods that assesses DNA or RNA based interactions
- All NGS methods require one or multiple protocols

DNA-Protein Interactions





Activity #1 (8 mins)

Breakout Room Activity:

In the class document, choose a method in your group that might use different or multiple protocols. Talk to your peers about what protocol differences could impact experimental results.

Optimizing and Maintaining Research Protocols

Electronic Research Notebooks
Protocol annotation

Reproducible protocols follow simple best-practices

- ▷ Include context relating to the purpose of the experiment
- ▷ Describe the step-by-step process of how the experiment is done
- ▷ Use an electronic lab notebook (ELN) to make printable protocol templates
- ▷ Record reagents, vendors, product IDs, and lot numbers
- ▷ Identify and record expected results (quality controls, control results)



Pro Tip

Using an electronic lab notebook helps with managing, executing & updating protocols

Electronic lab notebooks (ELNs) provide several advantages for keeping track of protocols & code



Searchable



Create Templates



Collaborate Easily



Track Changes

[How to pick an ELN](#)

[Highly utilized ELNs](#)

[Comprehensive list of ELNs and their features](#)

Other Considerations:

Check out [UCSF guidelines for electronic and paper lab notebooks](#)

To approve new tools, [UCSF IT can help with a security assessment](#)

Can you export the notebook data (as PDF or document)

Do you require a specialized ELN?

Electronic lab notebooks (ELNs) provide several advantages for keeping track of protocols & code



Use Templates



Unusual
Observations?



Record
Deviations



Record QCs

Crosslinking 30m

- 1 Fix cells adding formaldehyde to a final concentration of 1% and incubation at 37°C for 10 min.
- 2 Quench fixation adding glycine to a final concentration of 125 mM at cell culture plates and incubate at RT for 5 min.
- 3 After two washes with cold PBS, in the presence of complete protease inhibitor cocktail (Roche) and PMSF, scrape cells and centrifuge at 4°C for 5 min at 300 g.

Cell lysis and chromatin sonication 45m

- 4 Add 2,5 ml of 0.5% NP-40 lysis buffer to lyse cell pellet for nuclei isolation and incubate 10 min.
Let sit for 15 mins
- 5 Centrifuge 5 min 3000g 4°C
- 6 Resuspend cell pellet in 1 ml SDS 1% lysis buffer.
SDS at 0.5%
- 7 Sonicate cells in lysis buffer using Bioruptor (Diagenode, UCD-200) at high intensity for three cycles of 10 min with 30s sonication and 30s pause.
Hella Viscous... why?
- 8 Clarify chromatin by centrifugation at 17000g for 10 min at 4°C

Protocols.io offers a useful framework for managing and sharing research protocols.

- ▷ Create private or public experimental protocols
- ▷ Write out step-by-step processes for experiments
- ▷ Create templates that can be printed or modified
- ▷ Include detailed metadata to ensure reproducibility
- ▷ Publish protocols directly to share and cite your work

- ▷ For more info check out the [protocols.io tutorials](#)

Summary

- ▷ Writing protocols requires much more detail than writing about research methods
- ▷ Creating and managing protocols is made easier with electronic lab notebooks
 - Searchability, templates, ease of annotation are important factors in ease of use
- ▷ Sufficient context is essential in ensuring protocols are used correctly and provide informative results



Activity #2 (8 mins)

Solo Activity: Using the shared document, describe in one sentence how your lab maintains its protocols and research notebooks.

Computational Reproducibility

**The ability to obtain the same results
using the same data set**

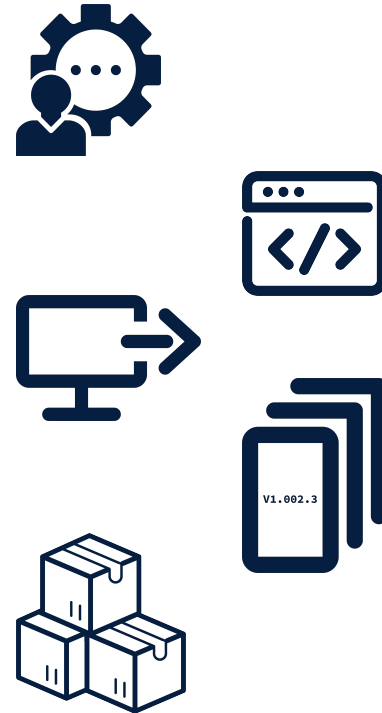


Three critical parts to computational reproducibility.

- ▷ **Documentation** - include any necessary context for the experiment, how to run the analysis and interpret the results
- ▷ **Citation** - cite all code and software needed to process, clean and visualize the data; share your code so it can be cited!
- ▷ **Annotation** - use a consistent coding format, include human-readable descriptions, improve readability with indentation

Create a README file to **document** how to apply your code or analysis

- Include context relating to the purpose of the script
- Describe how to install or execute the script from the coding files
- Indicate what is the expected output
- Note the version of coding language used
- The packages used, data sets that are required, other dependencies.



Let's unpack some examples of README files

- Good Examples
 - [Tracking the global reduction of marine traffic during the COVID-19 pandemic](#). March, D. 2021. Zenodo.
 - [Protected Health Information filter \(Philter\): accurately and securely de-identifying free-text clinical notes](#). Norgeot et al. 2020. npj Digital Medicine 3, 57
- Weaker Examples
 - [Cell-specific responses to the cytokine TGF \$\beta\$ are determined by variability in protein levels](#). Jette et al. 2018. Mol Sys Biol 14:e7733



Pro Tip

Cite all the code and data that you use in your analysis, so that it can be reused in the future

Use a software **citation** checklist to be sure it can be found by a future reader

- ▶ Have I identified the software, including my own, which makes a significant and specialized contribution to my academic work, and therefore should be cited?
- ▶ Have I checked to see if the software has a recommended citation from the creators and used it if available? If this is to a paper, have I also cited the software directly?
- ▶ Have I created as complete a citation as possible? Does it include who created the software, when it was created, the title of the software, and where it can be accessed?
- ▶ Have I referenced the software appropriately in my academic work, in compliance with any citation formatting guidelines?

Some examples of software citations

- ▷ Examples in text:
 - “We performed statistical analysis using R 4.0.3 (R Core Team, 2020) and the dplyr (Wickham et al. 2020) package for analysis.”
- ▷ References:
 - R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org>
 - Wickham, H., François, R. Henry, L., and Müller, K. (2020). dplyr: A Grammar of Data Manipulation. R Package Version 1.0.2. <https://CRAN.R-project.org/package=dplyr>



Activity #3 (8 mins)

Solo Activity: Using the shared document, write down two pieces of software or software packages that you use for your research

Share your code using open, citable repositories.

- ▷ Github is a great place to collaborate on code, but not the best place for publishing
- ▷ Use Github for ongoing work and a code repository like Zenodo to create an archived copy with a citable link.
- ▷ It is easy to port Github repositories using the [Github to Zenodo pipeline](#)
- ▷ For software development, submit your software to a disciplinary repository, or consider licensing your code to establish limits on its appropriate use.

Good **annotation** is key to reproducibility.

- ▷ Code is not always easily readable
- ▷ Use indentation and spacing to improve readability
- ▷ Use nouns to describe variables
- ▷ Use verbs to describe functions
- ▷ Add comments to describe why specific lines of code are used.
- ▷ Output the results in self-documenting code



Reading my own code, Ibraheem Ali, 2021

Good **annotation** is key to reproducibility.

```
61 FeedbackSurveyRaw <- read.csv("InstructorData.csv")
62 LibInsightRaw <- read.csv("OurData.csv") %>%
63   filter(Entered.By != "")
```

Efficient, no extra lines
of text, hard to reproduce

```
48 ## Data Import Section for Instruction Feedback
49
50 # Here we pull the raw data from the feedback survey and from the survey
data collection platform (data found at https://doi.org/laa/di.daa)
51
52 FeedbackSurveyRaw <- read.csv("InstructorFeedbackSurvey.csv") # feedback
survey results
53
54 LibInsightRaw <- read.csv("SelfReportedData.csv") %>%
55   filter(Entered.By != "") # Self Reported Instruction Data, filter step
removes blank submissions
```

Section Heading

Context, Data Source

Informative Names

Notes on cleaning steps

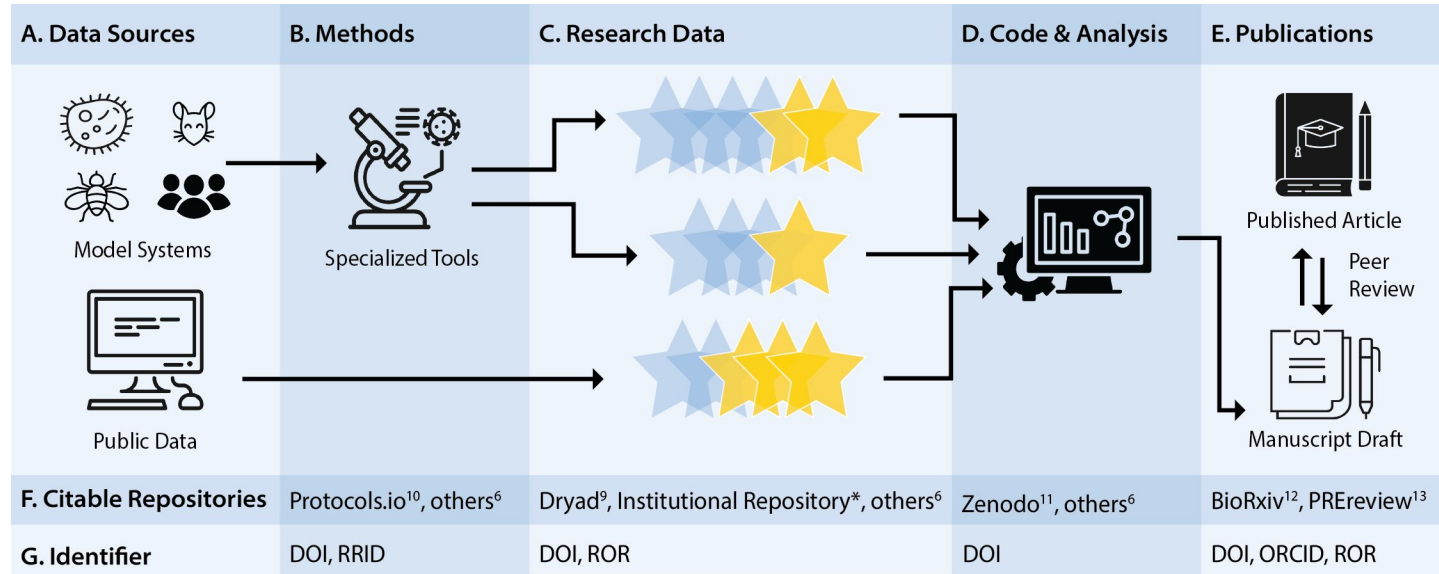
Summary

- ▷ Protocols and Methods include different levels of details
- ▷ Extensive annotation in protocols and code help with reproducibility
- ▷ Computational reproducibility is the ability to generate the same results with the same data
- ▷ Software citation gives authors credit, and helps when repeating data analysis
- ▷ The metadata (context) for protocols, code and data publications is essential for proper reuse

Persistent Identifiers bring it all together

**Connect your publications, data,
methods and code**

Research products are connected by their persistent identifiers.



Due Oct 24: Assignment #1

Short Project #1: Sharing your Organization System

The goal of this assignment is for you to apply what you have learned about designing reproducible organization systems to your own research materials. Please select an existing folder containing your documentation, data, code, or methods, and either design a better organization system, or describe why the current one is effective for your purposes. Submit a screenshot of the re-organized system, and answer the following questions in a short reflection document. Your reflection piece should be no longer than one page. Assignments submitted on time with both a screenshot and answers to the reflection questions will be graded as satisfactory.

Reflection Questions:

1. What folder/system did you select and why?
2. How did you decide to group the information?
3. What information did you include in your file names and why?
4. Are you the only person that accesses this system? If not, how will you ensure that the information design makes sense to others?
5. How will you maintain this system?



[Details and submission link in the CLE](#)

That's all folks!

- ▷ Questions?