Data and Donuts
Data organization
Notes

Key concept
Demo
Exercise

Slide 1: Hi, and welcome to Data and Donuts. I'm Tobin Magle, the Cyberinfrastructure facilitator at Colorado State University. Today we're going to be discussing good collaborative data management practices in the context of a useful tool called the Open Science Framework.

Slide 2: The open science framework is a rare tool that is useful in pretty much every part of the research data lifecycle. Data good data management practices initially important during the collection phase and through analysis, but will also make analysis archiving and sharing easier. You can share your data at a push of a button with OSF.

Slide 3: We'll be presenting this material in the context of the Open Science Framework. The OSF is a free, open source service of the Center for Open Science that was created as a public good to aid research. We're going to be talking about 5 data management topics and features of the Open Science Framework that assist with these practices:

1. You can store descriptive information about your project in the wiki
2. You can organize your research hierarchically using components
3. You can store your data on OSF storage or attach cloud drives as Add ons
4. You can allow collaborators that have OSF accounts access to your projects using the contributor feature
5. And finally, you can keep track of revisions using the built-in version control

Let's start by setting up our account and creating a project.

Slide 4:
Demo 1:

- To create an OSF account, go to http://osf.io, and click sign up in the upper right.
- Then fill in your name, email and a password.
- That's it!

Slide 5:
Demo 2:

- Once you're logged in, you should see a dashboard page,
- click the green button that says "Create a new project" above the search box on the right.
- Enter your project's name and hit create.

Slide 6: Let's go through an example dataset from a fictional grad student named Lou. He's just started his PhD and has picked up a project from a former post doc.

Exercise 1:

- First, download Lou's files at the following link: http://tinyurl.com/hvna4mg
- Then create an OSF account
- Finally, create a project called "Lou's project"

Slide 7: Your project should look like this. We'll be focusing on the panels for the wiki, files and components.

Slide 8: Let's start with the wiki.

- This is a good place to present a broad overview of your project.
- It's meant to evolve over the course of your project, so you can also add things like where you are in the course of the project, and what your near-term goals are.
- As you collect the data, it's also useful to put things like variable and ID definitions in this section.

Slide 9: To access the wiki, either click on the wiki panel, or the wiki menu item at the top.

Slide 10: Right now, your wiki has only one section, because we haven't added any subprojects.

- To add content, but the "Edit" button in the upper right.
- You can use the tools above the right panel to add formatting. The wiki is written in Markdown language, so if you want to see the formatted text, look in the middle panel.
- When you're done, hit save in the lower right.

Demo 3:

- Open the wiki
- Hit edit

Slide 11:
Exercise 2:
It's time to add content to your wiki.

- Open Lou's README from the files you downloaded
- Paste information into the wiki
- Format it as you see fit
- Make sure to edit the wiki as you change things!

Slide 12: Now that we have some information describing our project, we can start to think about how to organize the files.

- On our computers, we typically organize things using folders. If you have subfolders in a folder, you're already organizing things hierarchically.

- Organizing your files and folders hierarchically makes it easier to find things, especially if you and your collaborators agree on a standard.
- To organize research data, we recommend having one folder per project, and subfolders for distinct aspects of that project.
- For example, you could have folders for data, notes, protocols and manuscripts. Then within manuscripts, you could have folders for each paper, and then sections of that paper.
- No matter how you organize it, you should agree on a standard format within your research group and with your collaborators to make your data interoperable.

Slide 13:
When thinking about how to organize your data, it's a good idea to go back to the data inventory you created when writing your data management plan. This document describes the type of data you're going to collect and lists any other research outputs. So, as you're looking through your data inventory, ask yourself

- What kinds of files do I have?
- How can I group these things?
- Of these attributes, what are the most important ones?

You'll base your folder structure on the answers to these questions.

Slide 14:
Exercise 3:
Let's apply these concepts to Lou's data. Consulting the contents of the readme, answer these questions:

- What are the attributes for this project?
- Which seem like the most important ones?

Pause the video and take a moment to answer these questions.

Slide 15:
One attribute of this experiment is time, as the data are being collected longitudinally over the course of the year.

Slide 16:
Another attribute is infection status, because some of the data are from infected mice, and some from uninfected

Slide 17:
Additionally, we have two types of data here: weight and cytokine data.

Slide 18: Of these attributes, I consider data type to be the most important. Thus, I am going to create components in my OSF project for weight and cytokine data.

Slide 19: Components are essentially "sub projects" or subfolders. Each component has separate privacy settings, wiki, add ons and files. Here are some examples of how to separate your content into components. Let's look at some examples

Demo 4:
In this first example,

- The project is for the MacDougall lab
- The components are for projects within that lab, like a barcoding project and a insect food web project, and a project about the effect of chocolate on grad student happiness.
- If we click on the **barcoding project**, we can see that there are sub-sub components for Photos, lab notebooks, and data collection sheets.
- The insect food web project has components for identification keys and field samples

Demo 5:
This next example is for a clinical research project. Thus, the components are for standard parts of a clinical trial, such as clinical protocol, consent forms, IRB forms and analysis. Notice that there is no data component because of privacy concerns. None of these components have sub components.

Demo 6:
This example is a project associated with a manuscript.

- You can see that a link in the wiki takes you to the full manuscript within OSF
- Each component represents a manuscript section, like Analysis scripts and output, data, tables, and figures.

Demo 7:
This last example is a project for a large collaborative effort by many contributors to replicate psychological research. The components are fairly eclectic, with sections for analysis, resources, presentations, commends, and papers that they are trying to replicate. Each of these study component has subcomponents for a final report, data, study materials and analysis audit

Slide 20: Let's add components based on the organizational structure we've chosen. To add a new component, I will click on add component in the components panel.

Slide 21: Then, OSF will ask for a title, in this case cytokines. It asks you whether you'd like to add contributors for "lou's project". We will get to contributors later though. When you've named your component, click "Create".

Demo 8:

- Add component
- Call it cytokines

Slide 22: Now you have a new component. The layout of the component page looks similar to that of the entire project because all components have separate wikis, files, and can have their own components.

Slide 23: Now let's add the cytokine files to this component using the files pane.

Slide 24:
Demo 9:
To add files,

- click on "OSF Storage" in the file pane.
- Then click upload, which will open your computer's file browser.
- Then you can pick your files.
- Alternatively, you can drag and drop files into the Files pane.

Slide 25:
Exercise 4:
Do this for Lou's file.

- Create a cytokines component, and
- upload all the cytokine files onto OSF storage.
- If you have time, think about your data, and how you would organize it in OSF.

Slide 26: OSF Storage isn't the only option for adding files to OSF. You can also use storage add ons, which include many cloud services you might already be using. I'll show you how to use storage add ons with google drive as an example.

Slide 27: To connect an add-on service,

- select settings from the menu bar.
- Then Click Select add ons from the left side bar.
- Then click the check box next to the add on you want, in this case google drive.

Slide 28: Then,

- OSF will ask you to confirm that you want to connect this add on, given the conditions stated above.
- Agreeing to the terms will add google drive as an add on, and allow you to import your account.

Slide 29: After you import your account,

- OSF will ask you to pick a folder to associate with the component you want.
- This folder can have subfolders, but because your hierarchical structure is in the components, this shouldn't be necessary.
- After you Save this choice, you should be able to see your google drive in the files pane.

Demo 10:

- Go into settings
- Select add ons
- Pick google drive
- Confirm connection
- Log in
- Pick a folder

- Save
- Show google drive folder in files pane

Slide 30: If you change your mind about how things are organized, don't worry. You can reconfigure your add ons by going into settings and selecting Configure add-ons from the left panel.

Slide 31:
Exercise 5:
Now try this yourself.

- First, create a "weight" component
- Then Make a folder in your google drive or other supported cloud service.
- Next, add the weight files to this folder
- Then follow the instructions to link your Google Drive to OSF
- Finally, add the weight folder to the weight component
- Alternately, you can add weight to OSF storage, if you don't feel comfortable linking your google drive

Slide 32: Now that we have some data loaded, let's look at how to add collaborators to our project using the contributors feature.

- A contributor can be anyone with an OSF account
- When you add a contributor, you can control what parts of your project they can see
- Additionally, you can decide whether they can read content only, change content, or administer content, which gives them the rights to make components and add other contributors
- You can have different sets of contributors for every component.

Slide 33: To add a contributor,

- click on "Contributors" on the menu bar
- Then click the green box that says add to the right of Contributors heading

Slide 34: OSF will then create a search box where you can search for collaborators that have OSF accounts.

Slide 35: Click the green plus sign next to your collaborator's name. Then pick what type of access you'd like to grant them.

Slide 36: Finally, pick what components you'd like them to have access to.

Slide 37: At this point, you should see your collaborator's name in the contributors tab, and on the main page of your project.

Demo 11:

- Click on contributor tab
- Search for "Daniel Draper" – or pick someone from class

- Add the collaborator
- Give them read only access to the cytokines component

Slide 38:
Exercise 6:
- Now try this with a partner.
- Add each other as collaborators.
- Play with the access settings to prove to yourself that they work.

Slide 39: OSF also has a recent activity panel on each component. This lets you see what activity has been happening recently with each component of your project.

Slide 40: And all your work can be made public at a push of a button. You can control which parts are public at the component level.

Slide 41: The last feature we're going to talk about is version control. Have you ever gotten horribly confused about what the latest version of your document is after emailing things back and forth with collaborators? If so, OSF and its version control features are the thing for you.

Slide 42: OSF's version control feature tracks who did what when, and also interfaces nicely with more complex version control systems like git and github. Every time you make a change to a file in OSF, the version control features archives a hidden instance of the previous version, called a revision. But you only see one file in your directory.

Slide 43: Let's see how this works. I'm going to add Lou's readme file to his project.

Slide 44: If I click on the file name, the contents of the text file appear in a new window.

Slide 45: Now I can click edit above where the file is displayed. This opens an editing window. Does this look familiar? It's the same editing window as the wiki. I'm going to make a change here. Then click save

Slide 46: Now that the changes are saved, I can click on the Revisions button above the file. This opens a new pane where I can see that I have 2 revisions, with the newest version on top. If I want to look at an old revision, I can download it by hitting the blue button to the right of the user name.

Demo 12:
- Add the readme
- Edit the file
- Save
- Open revisions tab

Slide 47:

You can revert to older versions of a file if a mistake is made. To revert to an older version,

- Download a revision file by pressing the blue box with the downward arrow
- Then rename the file to the same name as the current revision
- Finally, drag and drop the file you just downloaded into the same folder as the file you want to revert.
- This will create a new revision.

Slide 48:
Exercise 6:
Let's practice:

- Edit one of Lou's excel files on your computer
- Re-upload the file with the same name and the same location
- What happened? Hint: Look at the revisions
- Bonus: edit a google drive file and look at revisions

Slide 49:

- Thanks for listening. I hope you found this data organization session to be helpful.
- Please email me at the address on this slide if you need help.
- Also, check out our data management pages for more information.