Week 1: Tools for reproducible science I

ReproRehab POD 1, 10/2/2023

Agenda

- Metadata – what is it, things to consider, etc.
  - file format: JSON
- Git – what is it, and how much do you need to know?
  - commit changes to your repository
  - fork someone’s repository
  - make a pull request

Metadata

- **Metadata describe a data set.**
- **Metadata ARE data.**
- **Metadata can be applied to anything.**
- **Reproducibility / Reusability**

Ghiringhelli et al. (2023); https://doi.org/10.1038/s41597-023-02501-8
https://guides.lib.unc.edu/metadata/definition

Metadata

Youtube: Data Sharing and Management Snafu in 3 Short Acts

Metadata – File format standards

- Complete and open documentation
- Platform independence
- Non-propriatery (vendor/software independence)

Quantitative tabular data with extensive metadata top a dataset with variable labels, code labels, and defined missing values, in addition to the matrix of data	
Preferred Formats	**Other Acceptable Formats**

• Character delimited text (ASCII or Unicode preferred):
  ◦ Comma Separated Values (*.csv)
  ◦ Delimited Text (*.txt)
• SQL Data Definition Language
• Structured text or mark-up file containing metadata information, e.g. DDI XML or JSON	

https://researchguides.uoregon.edu/data-management/fileformats#tabnometa

JSON

- **J**ava**S**cript **O**bject **N**otation
- Syntax rules:

1. Data is in *name : value* pairs

2. Data is separated by commas

3. Curly braces hold objects

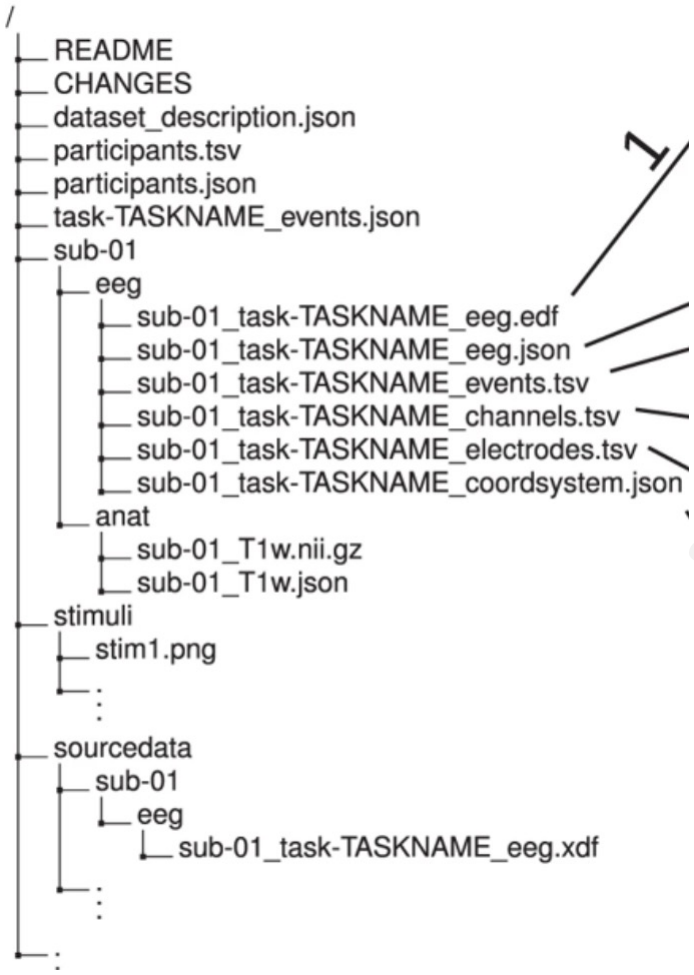
4. Square braces hold arrays

Ex. “Subject_Number” : 1 or “PID” : “C001”

Ex. {“PID” : “C001”, “age” : 34}

Ex. “Control” : [
{“PID” : “C001”, “age” : 34},
{“PID” : “C002”, “age” : 28}
]

JSON: example



sub-01_task-TASKNAME_eeg.json

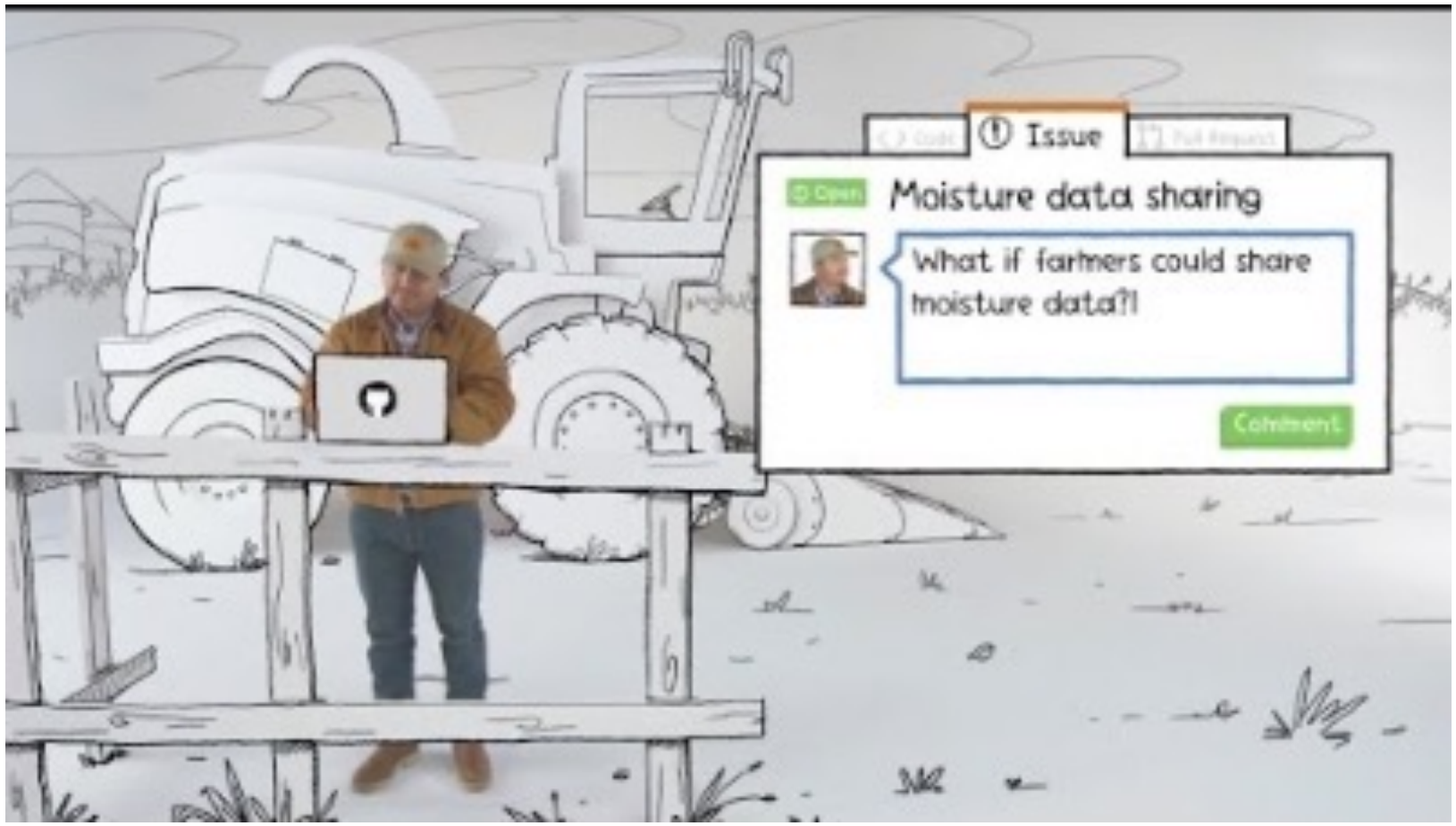
```
{
    "TaskName": "TASKNAME",
    "SamplingFrequency": 1000,
    "SoftwareFilters": "n/a",
    "EEGChannelCount": 4,
    "EOGChannelCount": 1,
    "EEGReference": "placed on Cz",
    "PowerLineFrequency": 50
}
```

sub-01_task-TASKNAME_coordsystem.json

```
{
    "EEGCoordinateSystem": "T1w",
    "EEGCoordinateUnits": "mm",
    "AnatomicalLandmarkCoordinates": {
        "LPA": [-0.067, 1.736e-09, -3.844e-09],
        "NAS": [-4.11e-09, 0.091, -4.541e-10],
        "RPA": [0.064, -6.435e-09, -4.566e-09]
    },
    "AnatomicalLandmarkCoordinateSystem": "T1w",
    "AnatomicalLandmarkCoordinateUnits": "mm",
    "IntendedFor": "sub-01_T1w.nii.gz"
}
```

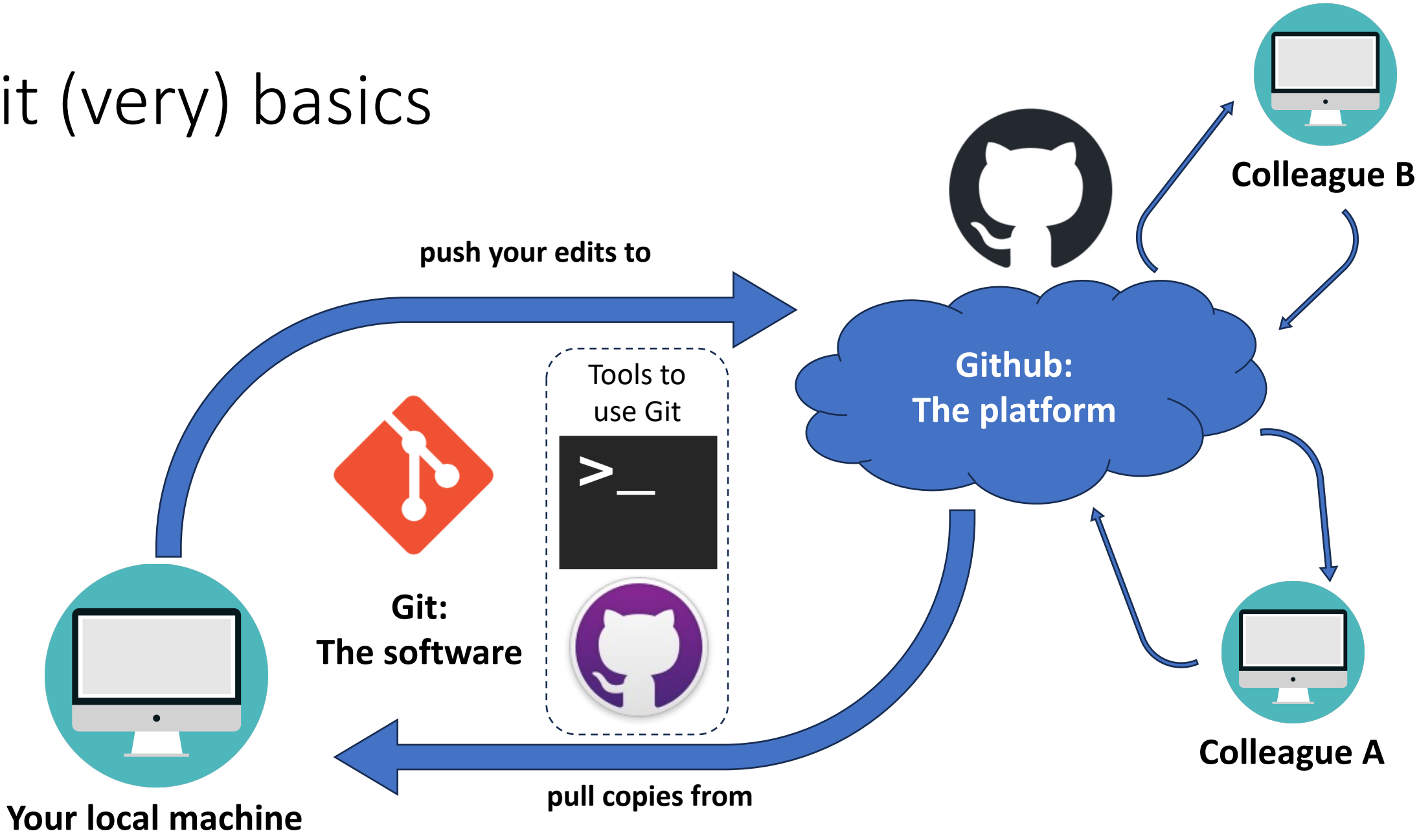
https://www.fieldtriptoolbox.org/example/bids/

Git



https://git-for-librarians.netlify.app/

Git (very) basics



https://git-for-librarians.netlify.app/

Git

- Why should a scientist learn how to use Git/Github?

1. Version control: every change you've ever made to your code or document will be saved, and you can always 'roll back'

2. Reproducibility: you can let someone who wants to use your code to copy your repository and move ahead!

https://www.gitscientist.com/#whats-the-point

Activity

- We will try out a very basic feature of Git/Github and edit a JSON file.