



## Week 3: (Meta)Data Management & MATLAB

ReproRehab POD 1, 10/20/2023

# Agenda

- Summary of this week's topics:
  - FAIR Data
- MATLAB
  - Data types (ex. Table, Structure, Cell...)
  - Work with Dr. Finley's MATLAB codes
- Activity
  - Change some parts, check results, and push your modified code to the shared repository.

# Quick check-in

- Depending on how far we go today...
- Week 4: tips and comments on Dr. Finley's code + learning more about the table data structure (useful functions + etc.)
- Week 5: Data visualization in MATLAB + interactive plots
- Week 6: Doing Statistics in MATLAB (we're not saying bye to R!)
- Week 7: Your topic, please! (and we need to reschedule it)

# The FAIR principles

---

- Make your (meta) data FAIR both for humans *and* machines.
- [Wilkinson et al. \(2016\)](#)
- [Translation](#) available



**F**indable



**A**ccessible



**I**nteroperable

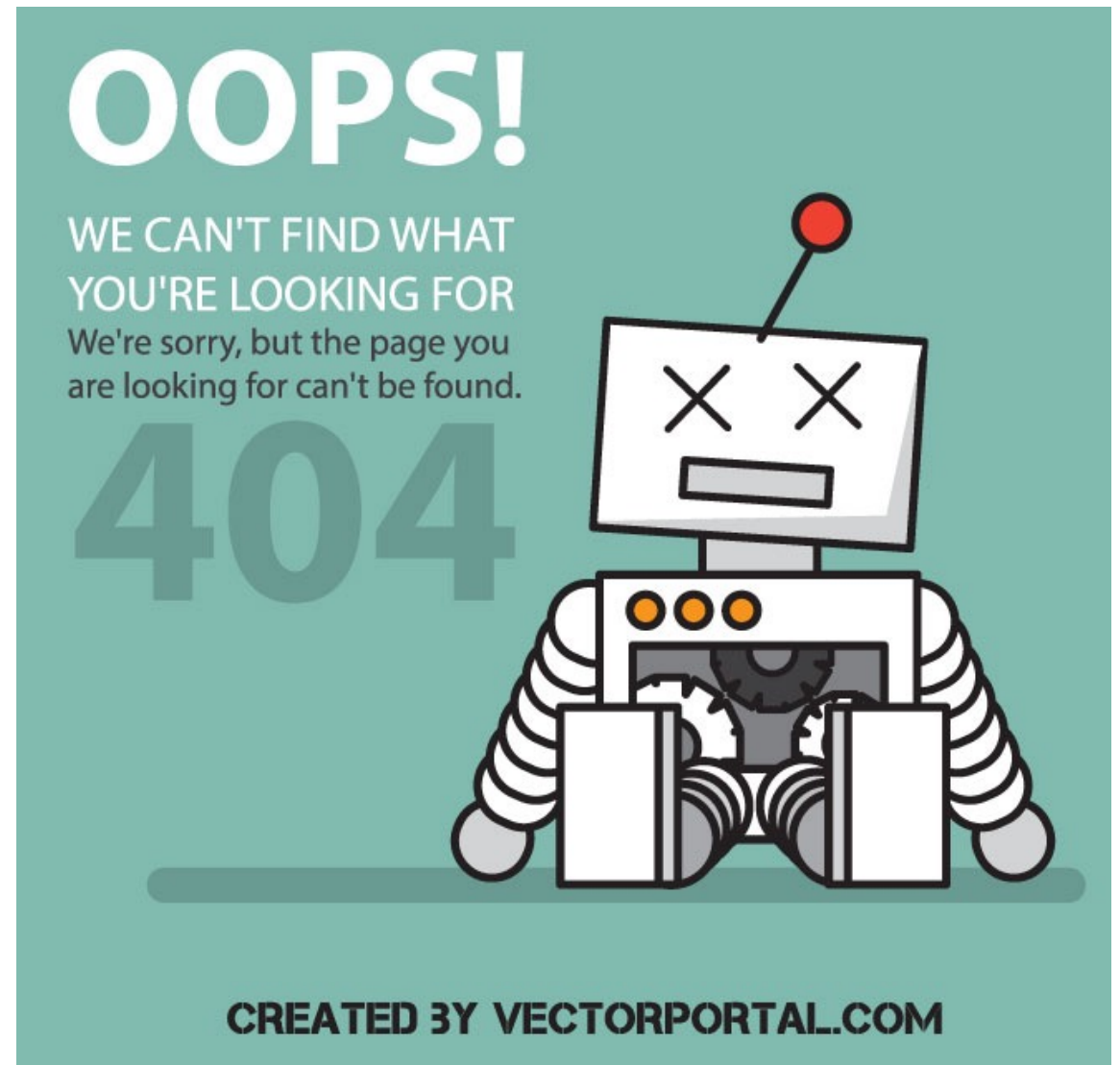


**R**eusable

# 1. Findable

---

- **F1.** (meta)data are assigned a *globally unique* and *persistent* identifier (PID)
  - Should never be *reused* (so no URL's)
  - DOI is one example (and many more!)



# 1. Findable

- **F2.** Data are described with rich metadata
  - Include necessary information to easily understand the data
  - Make it *structured*

- Unstructured: “subjects comprised male (N=6) and female (N =6) C57BL/6 mice, aged 25 days”

- ```
{ "subject_group": 1,  
  "Organism": "mouse",  
  "Age": 25,  
  "Age_unit": "days",  
  "Strain": "C57BL/6",  
  "Sex": "Male",  
  "Number": 6  
}
```

Ex. Metadata.json

- Subject Group: 2
- Organism: mouse
- Age: 25
- Age unit: days
- Strain: C57BL/6
- Sex: Female
- Number: 6

Ex. Structured vs. unstructured metadata  
(<https://www.youtube.com/watch?v=keH4Tc6mXMk>)

## 2. Accessible

---

- **A1.** (meta)data are retrievable by their identifier using a standardized communications protocol
- Your data become available by
  - HTTP + a browser

### HyperText Transfer Protocol

A sample DOI:

<https://doi.org/10.18112/openneuro.ds004808.v1.0.0>



## 2. Accessible

---

- **A1.** (meta)data are retrievable by their identifier using a standardized communications protocol
- Your data become available by
  - HTTP + a browser
  - FTP + a client application

### File Transfer Protocol





## 2. Accessible

- **A1.** (meta)data are retrievable by their identifier using a standardized communications protocol
- Your data become available by
  - HTTP + a browser
  - FTP + a client application
  - Application Programming Interface (API) + a browser

All articles about Tesla from the last month, pulled from newsapi.org

```
{
  "status": "ok",
  "totalResults": 11047,
  "articles": [
    {
      "source": {
        "id": null,
        "name": "Bringatrailer.com"
      },
      "author": "bringatrailer",
      "title": "2,800-Mile 2008 Tesla Roadster R80 3.0",
      "description": "This 2008 Tesla Roadster was acquired as a collector's item in September 2023. The car is finished in Electric Blue over a black leather interior.",
      "url": "https://bringatrailer.com/listing/2008-tesla-roadster-r80-3-0-2800-miles/",
      "urlToImage": "https://bringatrailer.com/wp-content/uploads/2023/10/2008-Tesla-Roadster-R80-3-0-2800-miles-1.jpg",
      "publishedAt": "2023-10-17T23:10:07Z",
      "content": "This 2008 Tesla Roadster was acquired by a collector in September 2023. The car is finished in Electric Blue over a black leather interior."
    },
    {
      "source": {
```

## 2. Accessible

---

- **A2.** metadata are accessible, even when the data are no longer available



### 3. Interoperable

- **I2.** (meta)data use vocabularies that follow FAIR principles
  - With respect to movement science / rehab science, this has a long way to go...

| Attribute | Data set 1                                                                     | Data set 2                                                       | Data set 3                                       |
|-----------|--------------------------------------------------------------------------------|------------------------------------------------------------------|--------------------------------------------------|
|           |                                                                                |                                                                  |                                                  |
| Age       | Age range: 15-30 days                                                          | Age range: 10-20 days                                            | Age: adult                                       |
| Organism  | <a href="#">B6J.Cg-Pax6<sup>em1</sup>Ems/EmsMmmh</a><br>(RRID:MMRRC_066963-MU) | B6-Pax6 <sup>^</sup> Sey-em1(3xF LAG)Ems<br>RRID:MMRRC_066963-MU | B6-Fey (FLAG-tagged Sey)<br>RRID:MMRRC_066963-MU |
| Gene name | Calretinin<br><a href="#">NCBIGene:794</a>                                     | Calb2<br><a href="#">NCBIGene:794</a>                            | CR<br><a href="#">NCBIGene:794</a>               |
| Location  | ACA<br>UBERON:0009835                                                          | Anterior Cingulate<br>UBERON:0009835                             | BA24<br>UBERON:0009835                           |

## 4. Reusable

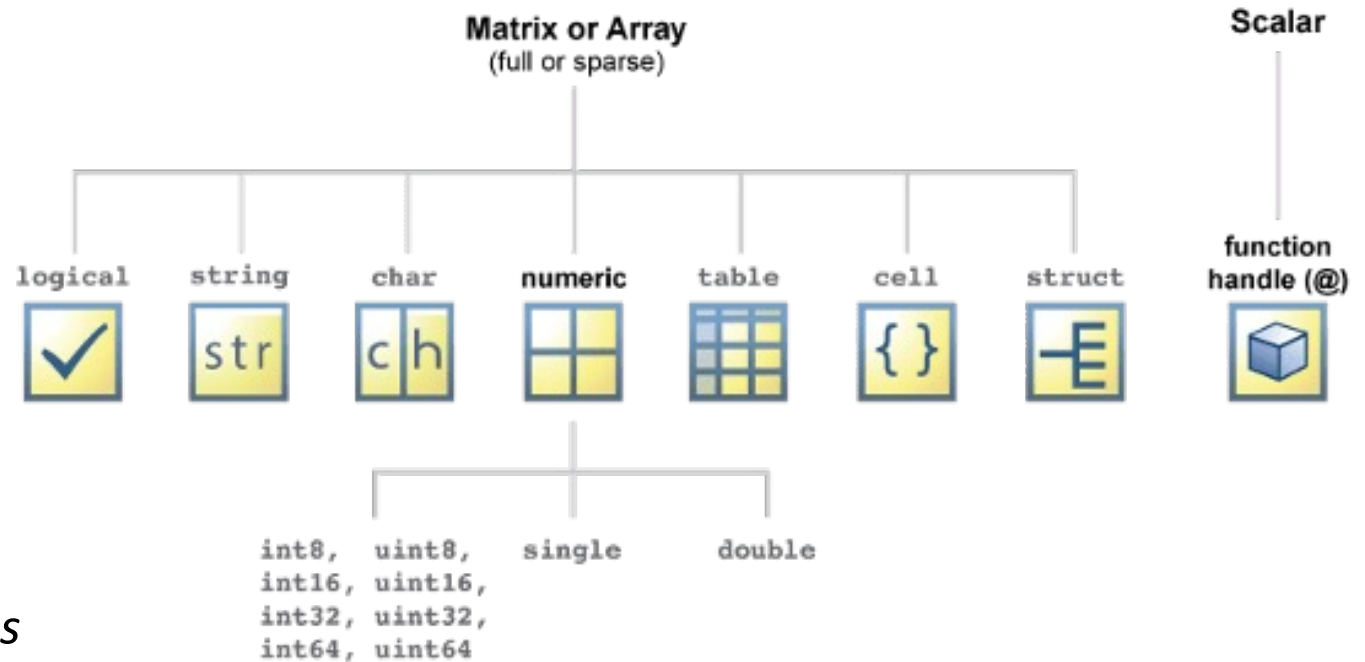
- **R1.** (meta)data are richly described with a plurality of accurate and relevant attributes
- **R1.3.** (meta)data meet domain-relevant community standards



```
└─ my_dataset-1/
   └─ sourcedata
   └─ ...
   └─ rawdata/
      └─ dataset_description.json
      └─ participants.tsv
      └─ sub-01/
         └─ func/
            └─ sub-01_task-rest_acq-default_bold.nii.gz
            └─ sub-01_task-rest_acq-longtr_bold.nii.gz
            └─ sub-01_task-rest_acq-longtr_bold.json
      └─ task-rest_bold.json
```

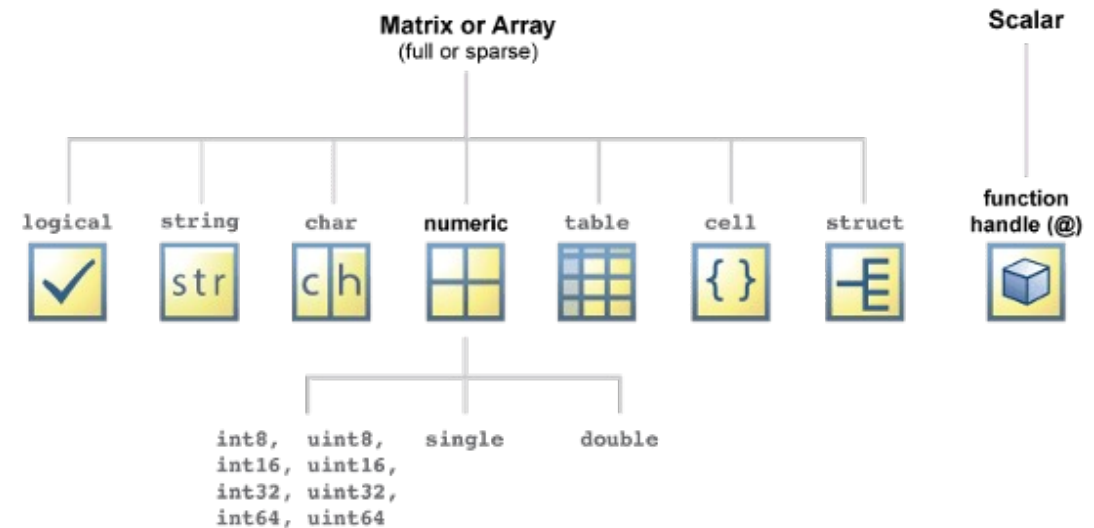
# MATLAB: Data types

- logical: **true** or **false**
- string: use `string()`
- char: 'char'
- numeric: 1, 1.0, -1.0,...
- categorical
- These are also called *homogeneous*



# MATLAB: Data types

- **Table**, cell, structure are *heterogeneous* data, meaning that each can contain data of different types.
- Ex) table with columns: id (char), knee joint angle (numeric), sex (char, numeric, or categorical)



# MATLAB: Table

---

- You often read output files in .csv, .txt, or .tsv
- Read them as **tables** using *readtable* function.
  - *csvread* – only reads csv files
  - *importdata* – if you have headers in your data file, it will prepare a structure.

|    | 4<br>lhipjoint_x | 5<br>lhipjoint_y | 6<br>lhipjoint_z | 7<br>lfemur_x |
|----|------------------|------------------|------------------|---------------|
| 1  | 9.2642           | 14.3610          | 34.5347          | 9.6028        |
| 2  | 9.2575           | 14.3631          | 34.5395          | 9.6002        |
| 3  | 9.2549           | 14.3663          | 34.5401          | 9.5993        |
| 4  | 9.2529           | 14.3658          | 34.5406          | 9.5986        |
| 5  | 9.2477           | 14.3648          | 34.5417          | 9.5959        |
| 6  | 9.2413           | 14.3647          | 34.5424          | 9.5925        |
| 7  | 9.2388           | 14.3624          | 34.5424          | 9.5895        |
| 8  | 9.2390           | 14.3608          | 34.5452          | 9.5865        |
| 9  | 9.2378           | 14.3616          | 34.5522          | 9.5842        |
| 10 | 9.2334           | 14.3613          | 34.5619          | 9.5832        |
| 11 | 9.2269           | 14.3602          | 34.5719          | 9.5800        |

# MATLAB: Table

---

- You can *dot index* a column variable

```
% read a csv file as a table
test = readtable('motion.csv');

% check the column names
% This is a cell array, so use {} for indexing
test.Properties.VariableNames
% ... like this
test.Properties.VariableNames{1,3}

% dot index a column variable
test.lhipjoint_x
```