practicum

FAIR Data for AI
University of Florida

Parent Project title: GatorSTAR: A New MARC U*STAR Program at the University of Florida
Working title: Adding a FAIR Data Practices Curriculum to UF's Practicum AI AI/ML training workshops

PI: David Julian
Additional team members: Matt Gitzendanner, Hao Ye, Yulia Strekalova

Primary Audience

- Undergraduate students participating in biomedical research
- Small discussion groups
- Experiential learning

practicum

Three 1-hour modules

1. Data organization in spreadsheets
   - Challenges with spreadsheets
2. Data availability in repositories
   - Journal article data accessibility, metadata, and ontologies
3. Data repositories
   - Finding shared data

https://practicumai.org/courses/FAIR/

practicum

Student Activities

1. Data organization in spreadsheets
   - Examine sample data in spreadsheet files
   - Discuss issues and share possible solutions
2. Data availability in repositories
   - Look for data shared alongside publications in PubMed
   - Check for DOI associated with the data
3. Data repositories
   - Look for data in a specialist data repository
   - Discuss ease of finding shared data to reuse in a secondary analysis in particular, metadata that would be effective

Module 1

Background: A professor was interested in measuring the effect of exercise on heart rate. For this project, students were assigned to record their resting pulse rate and then either run for one minute or sit for one minute and then record their pulse rate again.

...

This experiment was performed over three years and the data files for each year are available to you here:

1993_data.xlsx

1996 Exercise Data.xlsx

Run_Sit_data_95.xlsx

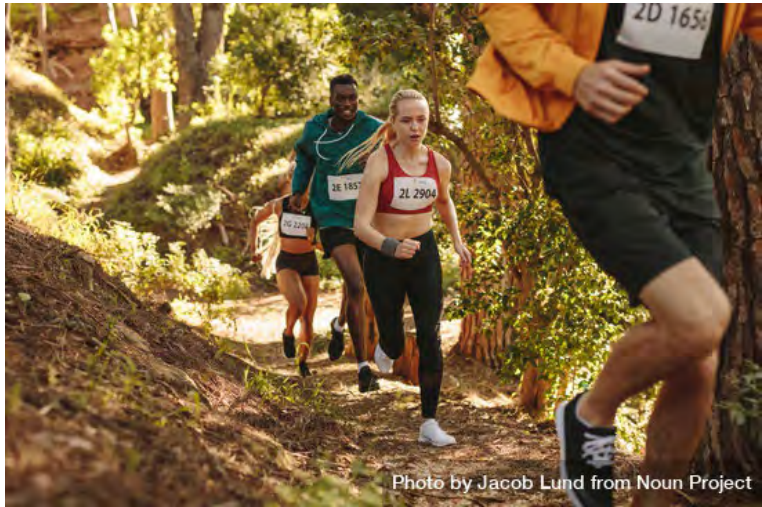


Photo by Jacob Lund from Noun Project

practicum

FAIR Data in AI/ML: Exercise 2

Student Instructions

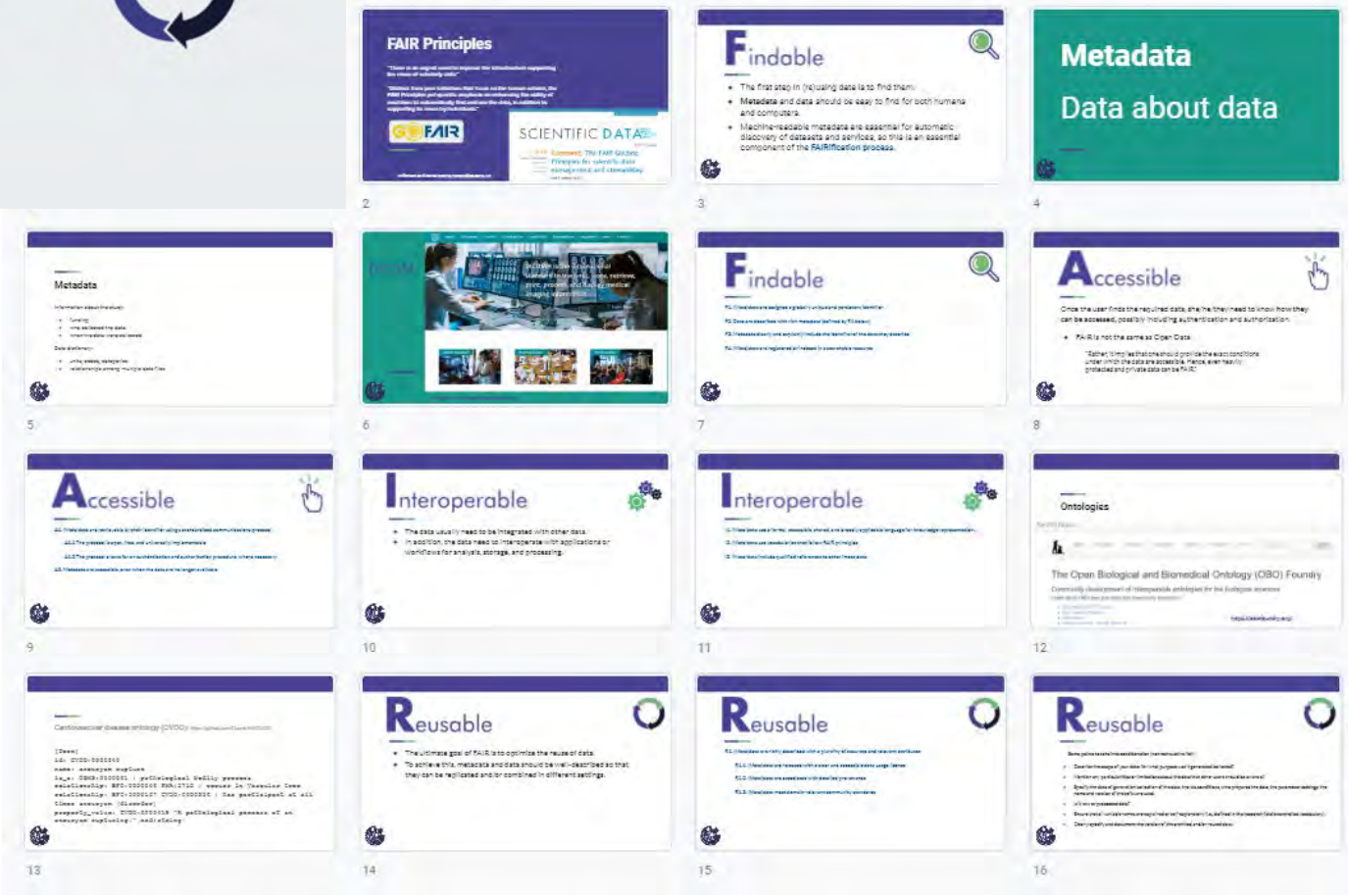
After the last exercise where you worked to compile data on the effects of running on heart rate, you have started to think that you might want to do more research on the overall effects of exercise on fitness. Perhaps, this research will lead you to launch your company's new line of fitness trackers!!

As such, you decide to start looking through the existing data on the effects of different exercises on fitness. There is already a fair bit of data published on this, so why not make use of those data?

Searching the literature for published datasets

practicum

Module 2
FAIR concepts



FAIR Data in AI/ML: Exercise 3

In exercise 1, we explored common issues with how data are organized in spreadsheets. We also provided a handout and the Broman and Woo 2018 paper with some best practices in organizing data in spreadsheets.

Then, in exercise 2, we explored the challenges with finding data associated with published literature and introduced the FAIR principles.

Now we turn to the process of making data available in data repositories.

Data repositories

About Biomedical Data Repositories and Knowledgebases

Accessible, well-maintained, and efficiently operated data resources are critical enablers of modern biomedical research. Data resources, through good data management practices, are the key to data and knowledge discovery, integration, and data reuse, as outlined by the FAIR Data Principles. To better support such a modern data resource ecosystem, NIH makes a distinction between data repositories and knowledgebases. While each activity is important for advancing biomedical research, data repositories and knowledgebases can have unique functions, metrics for success and sustainability needs.

Module 3
FAIR, RDM, Open