Oregon State University

National Institute of Environmental Health Sciences
Your Environment. Your Health.

DISCOVER ASSESS TRANSLATE APPLY
OSU/PNNL SUPERFUND RESEARCH PROJECT

T32 Supplement: NOT-OD-21-079 (AI-Workforce Workforce)

Training for Making data FAIR and compatible with Machine Learning (ML) and Artificial Intelligence (AI) Applications

Parent Grant # 5T32ES007060:
"*Integrated Regional Training Program in Environmental Health Sciences*"

Co-PI's: Craig Marcus, Ph.D. & Siva Kolluri, Ph.D. - T32 Co-Directors
Investigator: Martin Zanaj, M.S. - Center for Quantitative Life Sciences
Editor: Diana Rohlman, Ph.D. - OSU Superfund Center

NOT-OD-21-079– Virtual Closeout PI Meetings

October 24, 2022 and Nov 1, 2022

Oregon State University

Overall Objective:

Develop and disseminate on-line training materials for scientists and trainees to generate research data which is **FAIR** (**F**indable, **A**ccessible, **I**nteroperable, and **R**eusable) and **A**rtificial **I**ntelligence (**AI**) and **M**achine Learning(**ML**) compliant.

Overall Goals:

- **Create asynchronous, trainee-centered on-line learning modules.**
- **Make available at no cost to trainees and PI's via the internet.**
- **Generate materials compatible with a wide array of internet compatible electronic devices.**
- ***Workforce development* – generate training materials appropriate for and accessible for trainees at undergraduate level through post-doctoral fellows as well as for professional development for established investigators and faculty.**
- ***Broad target audience expanded to include not only T32 Fellows but SRP RETCC trainees, undergraduates, graduate students, post docs and faculty as well.***

Hosting Platform: canvas by INSTRUCTURE

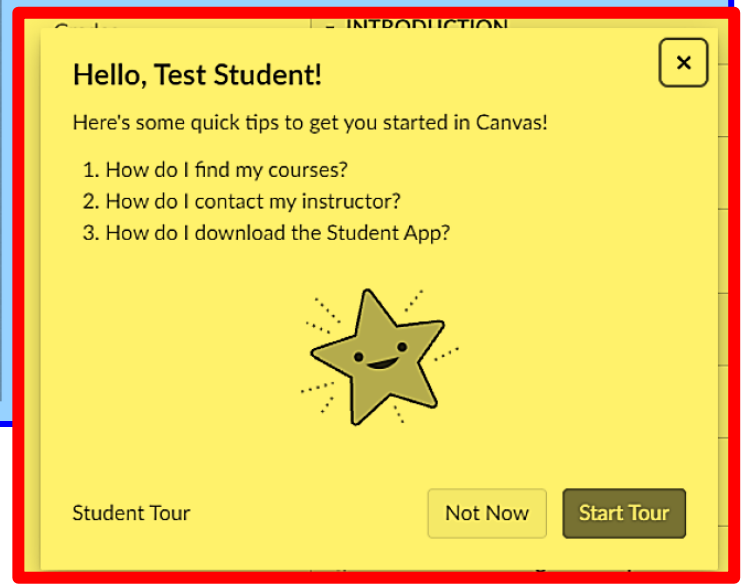
A Canvas Free-for-Teacher account will serve as the platform for the training modules.

https://www.instructure.com/canvas/try-canvas

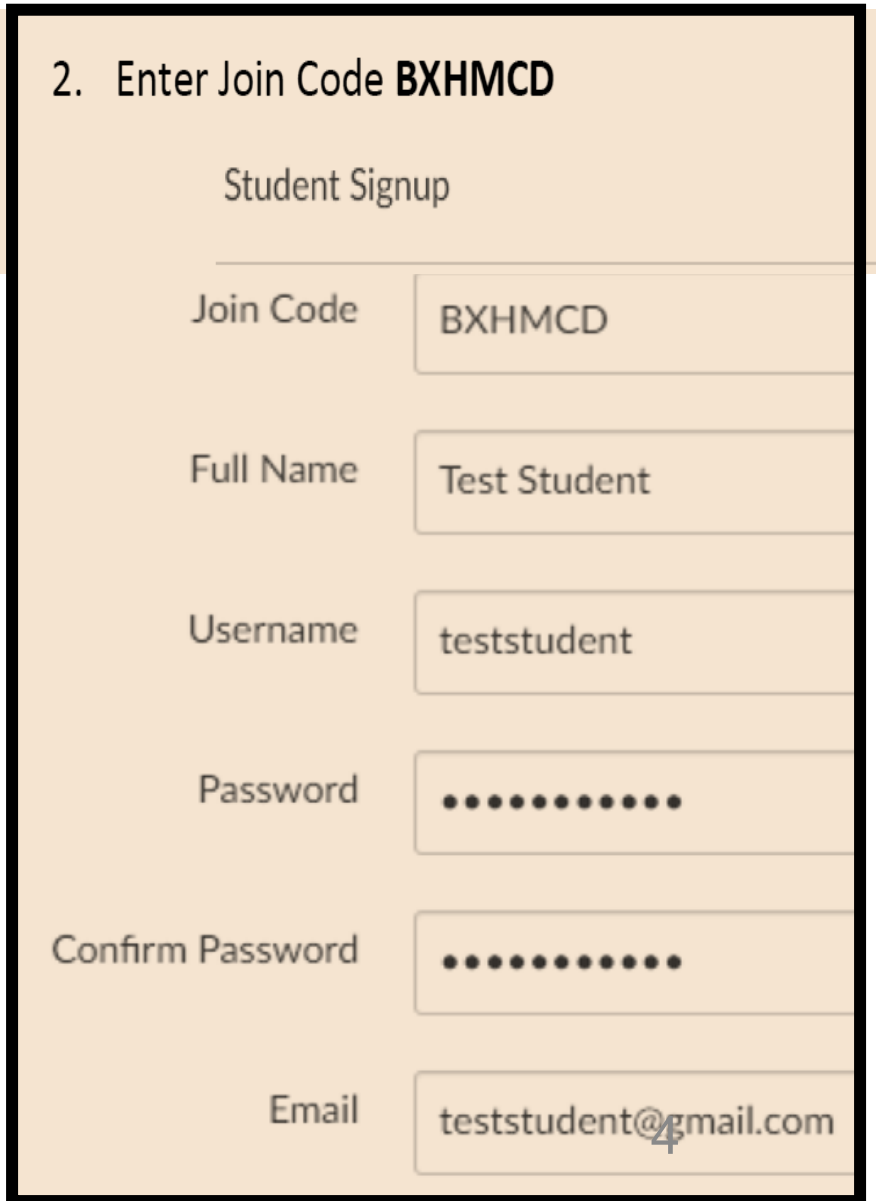
Self-Registration

New to Canvas?

https://canvas.instructure.com/register (STUDENT)



2. Enter Join Code **BXHMCD**



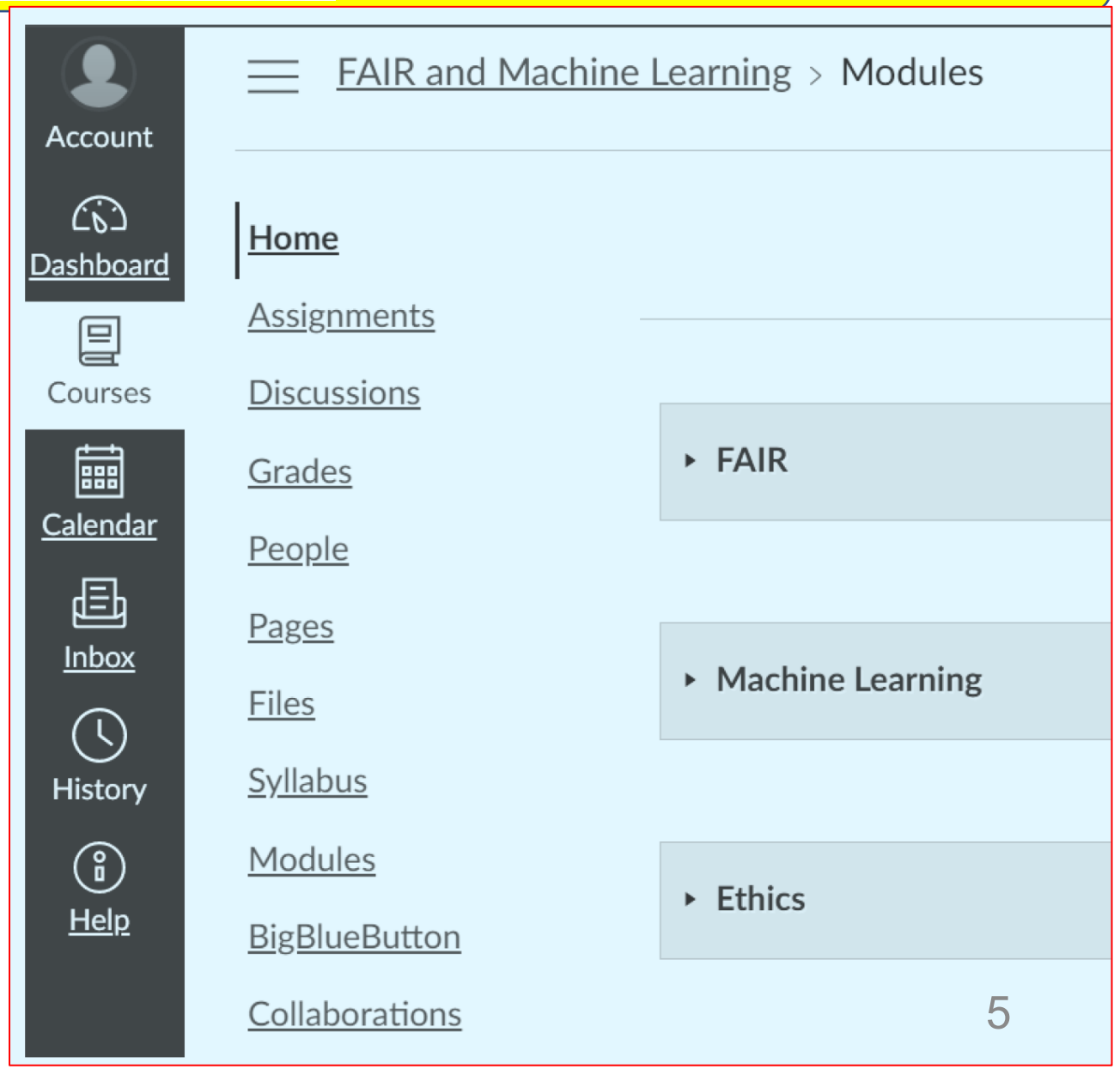
Hosting Platform: canvas by INSTRUCTURE

***Instructure Canvas* provides**

***Unlimited* AND *Free* environment for instructors and students.**

KEY FEATURES:

- **Content, assignments, assessments and discussions).**
- **Compatible with most internet enabled devices (Canvas *Mobile App Suite* and technical support).**
- **Integrate with third-party applications (Zoom, Panapto, Piazza, …)**
- **Self-enrollment; student paced; self-assessments with feedback.**
- **Course analytics for Instructors.**
- **Asynchronous access to learning materials for trainees.**



Dissemination of FAIR-ML-AI Training Materials

- **INSTRUCTURE – Free Canvas**
- YouTube
- All EHS T32 Grantees
- All Superfund Research Program Grantees
- Regional T32 and SRP Training Partners:
  - UC Davis, UC Berkeley, U. Washington, Univ New Mexico, Pacific NW National Labs
- OSU Websites:
  - EMT Dept; T32; SRP, CQLS

Oregon Big Data Research and Education Consortium:

Research and Graduate Degree Institutions:
Oregon State University
Portland State Uuniversity
University of Oregon.\
OHSU, Oregon Health & Science University

Four-year college partners
Heritage Unniversity
Lewis & Clark College
Linfield College
Reed College
Southern Oregon University
Washington State University Tri-cities
Washington State University Vancouver;
Western Oregon University

Community college partners
Blue Mountain CC
Chemeketa CC
Linn-Benton CC
Mt. Hood CC

Progress Report

Completed Tasks:

- All Training Modules Completed and Narrated.
- Instructure Free Course Site Created and Modules Uploaded.
- Initial Review/Evaluation Completed: (Professional Training Materials Evaluator, SRP DMAC Investigators; EHS Faculty Investigators/Mentors; EHS Graduate Student Trainees.
- Developed additional training materials on Ethical Concerns for FAIR, M/L and AI

No Major Challenges:

Remaining Tasks: (Institutional Resources)

- Revise Training Modules in Response to Reviewers
- Create independent You Tube videos from Instructure Modules

BASIC MODULE

FAIR PRINCIPLES & THEIR COMPONENTS

Data Preparation

What is Data Pre-Processing?

"*Separating the wheat from the chaff.*"

Good Data

Pre-Processing

Bad Data

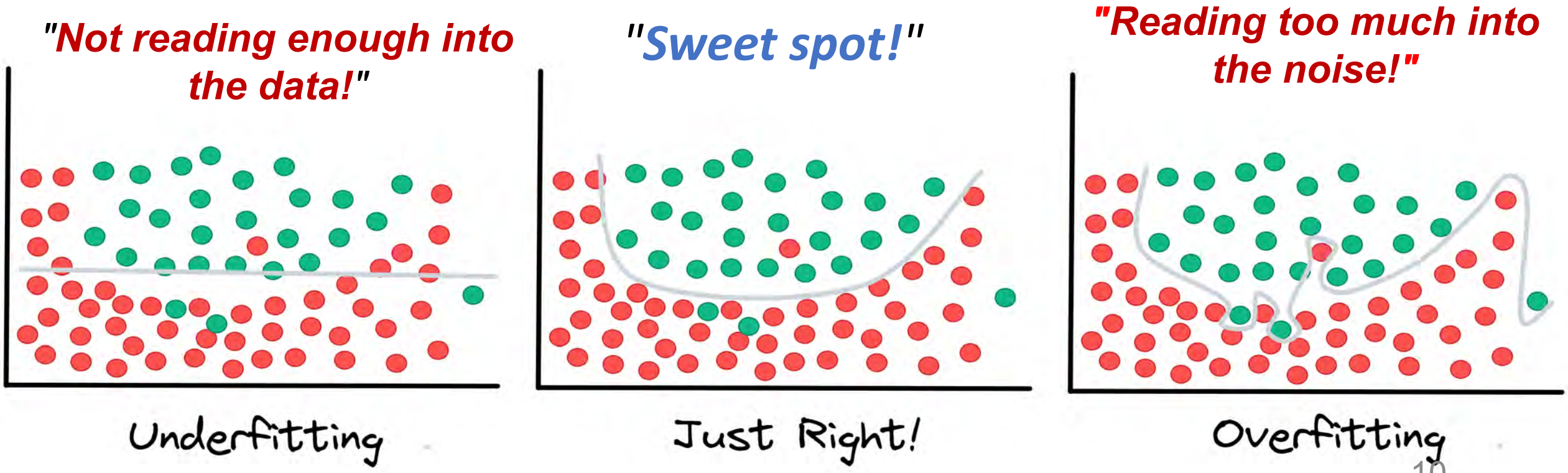
from YouTube

Data Preparation

Complex = less interpretable;
Simple = more interpretable;

Complexity, Bias, & Fit

- Complex models tend to have less bias (closer to truth), yet there are subject to overfitting.
- Simpler model tend to have more bias (further to truth), yet they are subject to underfitting.



The Genesis of FAIR

Everything Started in 2014 at a workshop held at Leiden University, Netherlands at the Lorentz Center.

- Fun fact- LU is the oldest university in the Netherlands ([18] Wikipedia, 2021).

What is the Lorentz Center?

"The Lorentz Center ([19] LC, 2021) is a workshop center that hosts international scientific meetings of typically one week. The workshops are characterized by an open and interactive atmosphere and their high scientific quality. We aim to bring scientific fields and minds together and we endorse diversity in the broad sense: scientific level, gender, culture and geography."

About the Workshop

The workshop was named "Jointly Designing a Data FAIRPORT" ([20] FairPort, 2014) and it brought together 25 leading academic and private sector experts ([21] FairPort 2014), where through "moderated plenary sessions a nd breakout groups" discussions, issues regarding data publishing, discovery, sharing and re-use were tackled. At the conclusion of the 4 day workshop, the consensus revolved around the idea of a global infrastructure for d ata publishing built upon a minimal set of community agreed standards and practices, where data providers and consumers could benefit. Successively, refined and improved by FORCE 11 ([11] Force11, 2016) members, this minimum set of principles was defined to be Findable, Accessible, Interoperable, and Reusable- the FAIR Guiding Principles as we know them today and published in 2016 ([17] Wilkinson, 2016).

Lorentz Center <https://www.lorentzcenter.nl/about-us.html>, Leiden University <https://en.wikipedia.org/wiki/Leiden_University>, Force 11 <https://www.force11.org/>, FAIR principles <https://www.nature.com/articles/sdata201618>

Collaborators:

Oregon State University