# Artificial Intelligence and Machine Learning Fairness and Bias with Applications to Environmental Justice

**Phil Brown, PhD & Kimberly Garrett, PhD**

Northeastern University

Social Science Environmental Health Research Institute

19 October 2022

Northeastern University
Social Science Environmental Health Research Institute

# Transdisciplinary Training at the Intersection of Environmental Health and Social Science

- T32 Training Program co-directed by Phil Brown at Northeastern University and Julia Brody at Silent Spring Institute
- Interdisciplinary training that prepares graduate students and postdocs for successful research careers at the nexus of environmental health, social science,  and policy

<u>Current Postdoctoral Fellows</u>

- Abby Bline
- Elissia Franklin
- Vivian Underhill
- Kayleigh Ward

<u>Doctoral students</u>

- Alaina Boyle
- Allison Deese
- Jamie Hanna
- Xena Itzkowitz

# Stackable trainings in the FAIRification and AI/ML readiness of data with applications to environmental health and justice

Phil Brown-PI. Co-Is Justin Manjourides, David Kaeli, Jennifer Dy

- Addressing knowledge gaps and connecting environmental, health, and social science research to emerging technologies
  - Highlighting ethical considerations and best practices for the use of artificial intelligence and machine learning (AI/ML)
- Collaboration between:
  - Northeastern University Social Science Environmental Health Research Institute (SSEHRI)
  - Puerto Rico Testsite for Exploring Contamination Threats (PROTECT) / Superfund Research Program Center - Data Management and Analysis Core (DMAC )
  - Data Science @ Northeastern
  - Observational Health Data Sciences and Informatics (OHDSI) Center @ The Roux Institute
  - Experiential AI Institute @ Northeastern

# Guiding Principles for Data Justice

Wilkinson et al.
*Scientific Data*, 2016
www.go-fair.org

Global Indigenous
Data Alliance
www.gida-global.org

# AI/ML Fairness & Bias with Applications to Environmental Justice Module

- Our objectives:
  - Describe the state of environmental justice research, particularly in regards to issues of race and class in environmental health and the environmental justice movement
  - Explain community-based participatory research principles
  - Evaluate data sharing  and ownership agreement issues and concerns between researchers and community members
  - Generate research questions and recognize the ethical issues involved in addressing community research needs
  - Examine ethical and technical issues in data accessibility from state and federal environmental agencies

# Webinar 1: AI/ML 101, Ethics Ecosystems, and Applications to Environmental Science

# Webinar 2: Datafying Justice & Accessibility in Design

# AI/ML Webinars & Current Events on SSEHRI website

www.northeastern.edu/environmentalhealth

## Webinar 1: Intro to Artificial Intelligence & Machine Learning, Creating an Ethics Ecosystem, and Applications to Environmental Science


The inaugural webinar features introductions to AI/ML, big data, and ethical frameworks for addressing technology and data in research. We also hear from two researchers using AI and ML in the field, sharing their perspectives on data justice, exposure science, and their connections to social justice.
This session features:
Justin Manjourides, Northeastern University Department of Health Sciences, OHDSI Center, and the Roux Institute
John Basl, Northeastern University Department of Philosophy and Religion, Northeastern Ethics Institute
Lourdes Vera, University of Buffalo Department of Sociology
Abhishek Viswanathan, University of Pittsburgh School of Computer and Information Science

*Transcript forthcoming*

## Webinar 2: Datafying Justice and Accessibility in Design


## Additional Resources

Additional reading and current events on AI/ML, human and environmental health research, and social justice.

### NIH Launches Bridge2AI Program to Expand the use of AI in Biomedical and Behavioral Research

Updated October 2022

"The National Institutes of Health will invest $130 million over four years, pending the availability of funds, to accelerate the widespread use of artificial intelligence (AI) by the biomedical and behavioral research communities. The NIH Common Fund's Bridge to Artificial Intelligence (Bridge2AI) program is assembling team members from diverse disciplines and backgrounds to generate tools, resources, and richly detailed data that are responsive to AI approaches. At the same time, the program will ensure its tools and data do not perpetuate inequities or ethical problems that may occur during data collection and analysis." See full article here.


### US Artificial Intelligence Bill of Rights

Updated October 2022
The White House: Blueprint for an AI Bill of Rights (10/4/2022)
The Associated Press: White House unveils artificial intelligence 'bill of rights' (10/4/2022)
Wired: Biden's AI Bill of Rights is Toothless Against Big Tech (10/4/2022)

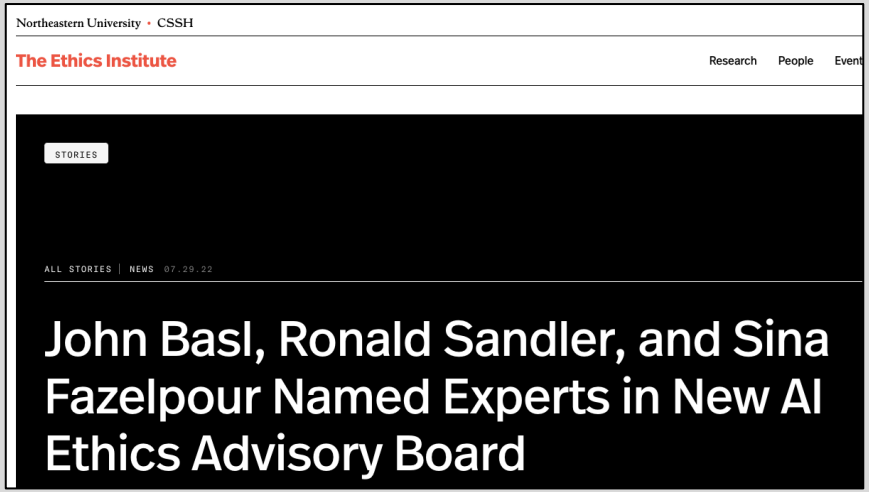
### National Science Foundation's Open Knowledge Network Roadmap

Updated September 2022

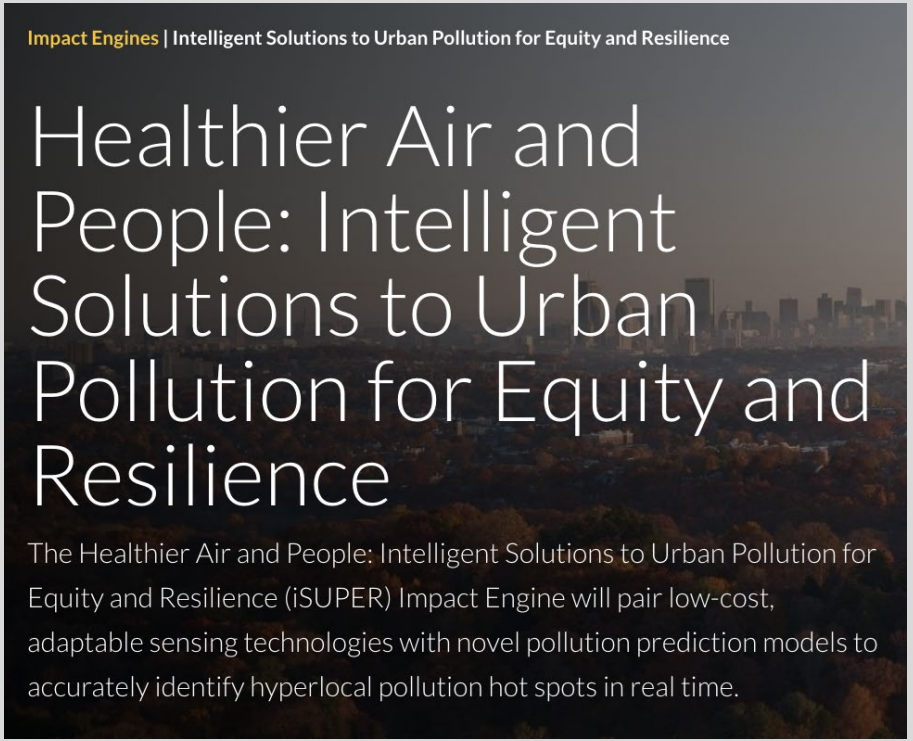
The NSF's Open Knowledge Network Roadmap report "outlines a strategy for establishing an open and accessible national resource to power 21st century data science and next-generation artificial intelligence. Establishing such a knowledge infrastructure would integrate the diverse data needed to sustain strong economic growth, expand opportunities to engage in data analysis, and address complex national challenges such as climate change, misinformation, disruptions from pandemics, economic equity and diversity."(1)

# Other AI work on campus

EAI The Institute for Experiential AI
Northeastern University




https://cssh.northeastern.edu/ethics/john-basl-ronald-sandler-and-sina-fazelpour-named-as-experts-in-new-ai-ethics-advisory-board/




https://impactengines.northeastern.edu/ie/isuper/

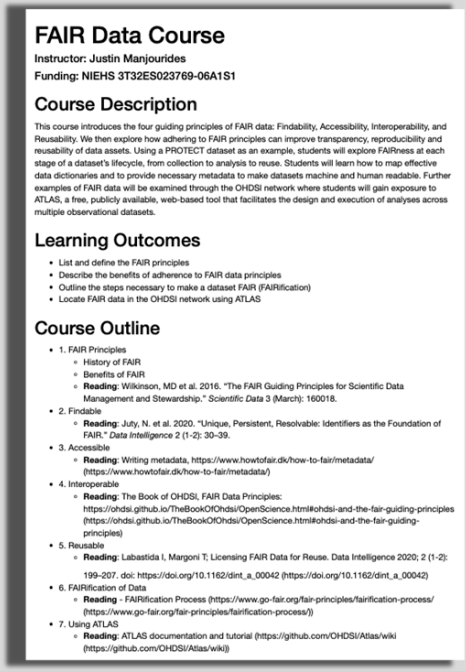
# Other AI work on campus: Superfund Research Program

Northeastern University

PROTECT

Slide provided by Justin Manjourides

## FAIR Data Module



### FAIR Data Course

**Instructor: Justin Manjourides**

**Funding: NIEHS 3T32ES023769-06A1S1**

### Course Description

This course introduces the four guiding principles of FAIR data: Findability, Accessibility, Interoperability, and Reusability. We then explore how adhering to FAIR principles can improve transparency, reproducibility and reusability of data assets. Using a PROTECT dataset as an example, students will explore FAIRness at each stage of a dataset's lifecycle, from collection to analysis to reuse. Students will learn how to map effective data dictionaries and to provide necessary metadata to make datasets machine and human readable. Further examples of FAIR data will be examined through the OHDSI network where students will gain exposure to ATLAS, a free, publicly available, web-based tool that facilitates the design and execution of analyses across multiple observational datasets.

### Learning Outcomes

- List and define the FAIR principles
- Describe the benefits of adherence to FAIR data principles
- Outline the steps necessary to make a dataset FAIR (FAIRification)
- Locate FAIR data in the OHDSI network using ATLAS

### Course Outline

- 1. FAIR Principles
  - History of FAIR
  - Benefits of FAIR
  - **Reading**: Wilkinson, MD et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (March): 160018.
- 2. Findable
  - **Reading**: Juty, N. et al. 2020. "Unique, Persistent, Resolvable: Identifiers as the Foundation of FAIR." *Data Intelligence* 2 (1-2): 30-39.
- 3. Accessible
  - **Reading**: Writing metadata, https://www.howtofair.dk/how-to-fair/metadata/ (https://www.howtofair.dk/how-to-fair/metadata/)
- 4. Interoperable
  - **Reading**: The Book of OHDSI, FAIR Data Principles: https://ohdsi.github.io/TheBookOfOhdsi/OpenScience.html#ohdsi-and-the-fair-guiding-principles (https://ohdsi.github.io/TheBookOfOhdsi/OpenScience.html#ohdsi-and-the-fair-guiding-principles)
- 5. Reusable
  - **Reading**: Labastida I, Margoni T; Licensing FAIR Data for Reuse. Data Intelligence 2020; 2 (1-2): 199–207. doi: https://doi.org/10.1162/dint_a_00042 (https://doi.org/10.1162/dint_a_00042)
- 6. FAIRification of Data
  - **Reading** - FAIRification Process (https://www.go-fair.org/fair-principles/fairification-process/ (https://www.go-fair.org/fair-principles/fairification-process/))
- 7. Using ATLAS
  - **Reading**: ATLAS documentation and tutorial (https://github.com/OHDSI/Atlas/wiki (https://github.com/OHDSI/Atlas/wiki))

This course introduces the four guiding principles of FAIR data: Findability, Accessibility, Interoperability, and Reusability.

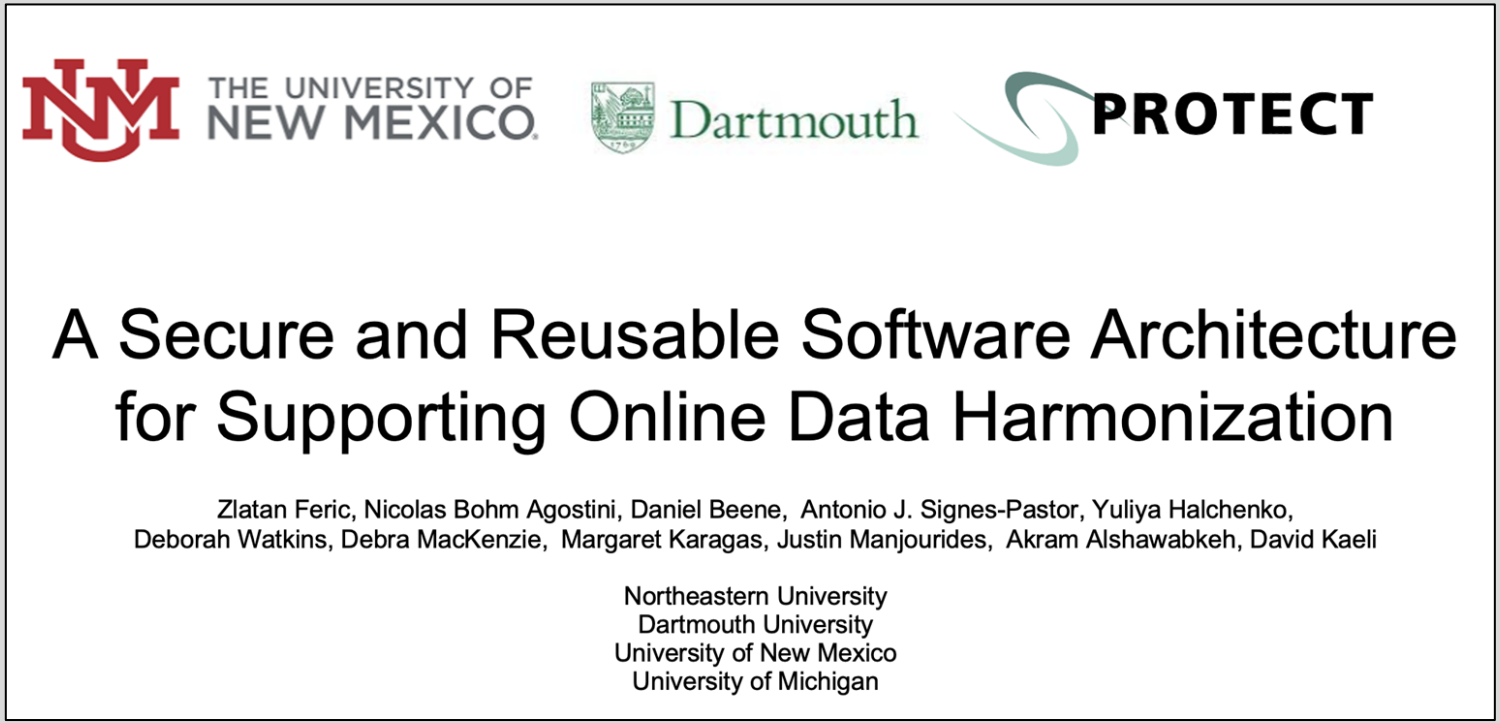
Examples of FAIR data will be examined through both the PROTECT data dictionary and the OHDSI network.

Targeted readings accompany each topic.

Learning Outcomes:

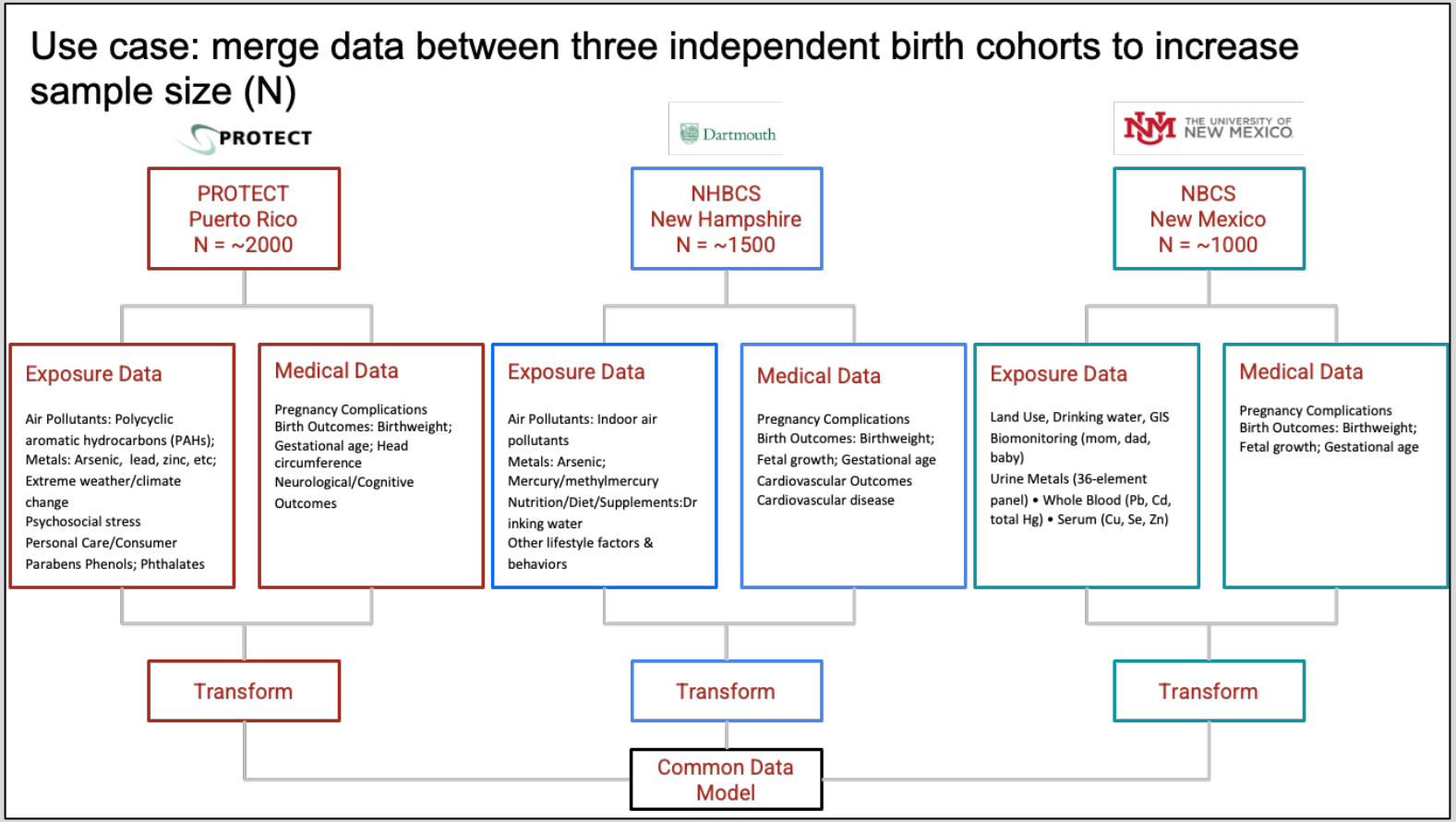
- List and define the FAIR principles
- Describe the benefits of adherence to FAIR data principles
- Outline the steps necessary to make a dataset FAIR (FAIRification)
- Locate FAIR data in the OHDSI network using ATLAS.

# Other AI work on campus: Superfund Research Program




Slide provided by Dave Kaeli

# Other AI work on campus: Superfund Research Program




Slide provided by Dave Kaeli

# Future Directions

- Webinar focused on data ethics and reproductive rights (Spring 2023)
- Future webinars with AI/ML leaders
- Maintain “additional resources” page through SSEHRI website with current events and relevant topics