UNIVERSITY of WASHINGTON

# UW Data Science for Social Good

## Information session for prospective project lead applicants

Anissa Tanweer, Program Director
Emily Keller, Program Manager

UNIVERSITY of WASHINGTON
eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

# Agenda

- Brief introduction to the eScience Institute
- Data Science for Social Good (UW DSSG)
  - Program overview
  - Proposal process
  - Program logistics
  - Previous projects
- Questions?

***Note: This session is for prospective project leads (our info session for students will be posted soon)***

# Introduction to eScience

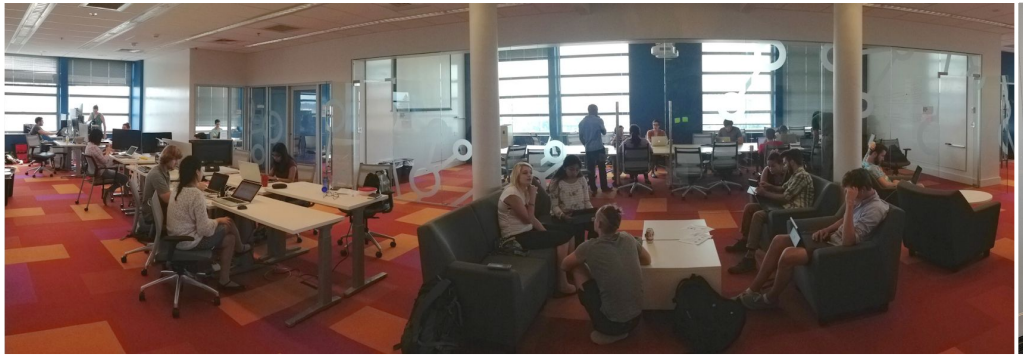
# eScience Mission

**The eScience Institute empowers researchers and students in all fields to answer fundamental questions through the use of large, complex, and/or noisy data.**

**As the hub of data-intensive discovery on campus, we lead a community of innovators in the techniques, technologies, and best practices of data science and the fields that depend on them.**

UNIVERSITY of WASHINGTON
eScience Institute
DATA SCIENCE FOR SOCIAL GOOD


UNIVERSITY
LIBRARIES



# eScience Research Team

## Director of Research

David Beck
Ph.D. Medicinal Chemistry,
Biomolecular Struct. & Design

## Data Scientists

Anthony Arendt
Ph.D. Geophysics

Bernease Herman
B.S. Statistics
Formerly SE at Amazon

Valentina Staneva
Ph.D. Applied Mathematics and Statistics

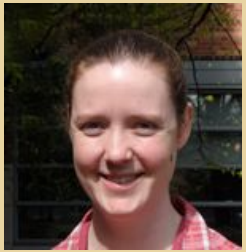
Anissa Tanweer
Ph.D. Communication

Noah Benson
Ph.D. Biomedical & Health Informatics

Joe Hellerstein
Ph.D. Computer Science
IBM Research, Microsoft Research, Google (ret.)

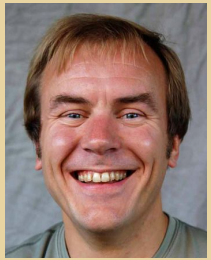
## Research Scientists

Nicoleta Cristea
Ph.D. Environmental Engineering



Bryna Hazelton
Ph.D. Astrophysics
Physics

Scott Henderson
Ph.D. Geological Sciences



Vaughn Iverson
Ph.D. Oceanography

Spencer Wood
Ph.D. Zoology

## Technical Education Specialist

Naomi Alterman
M.S. Electrical Engineering

# We Disseminate Data Science Expertise & Best Practices

- Open Office Hours
- UW Data Science Seminar & Community Seminar
- Tutorials, bootcamps, workshops, and hack weeks
  - Astrohack, neurohack, geohack
  - Software carpentry (> 400 participants since we started counting in 2015)
- Winter Incubator
- Summer DSSG

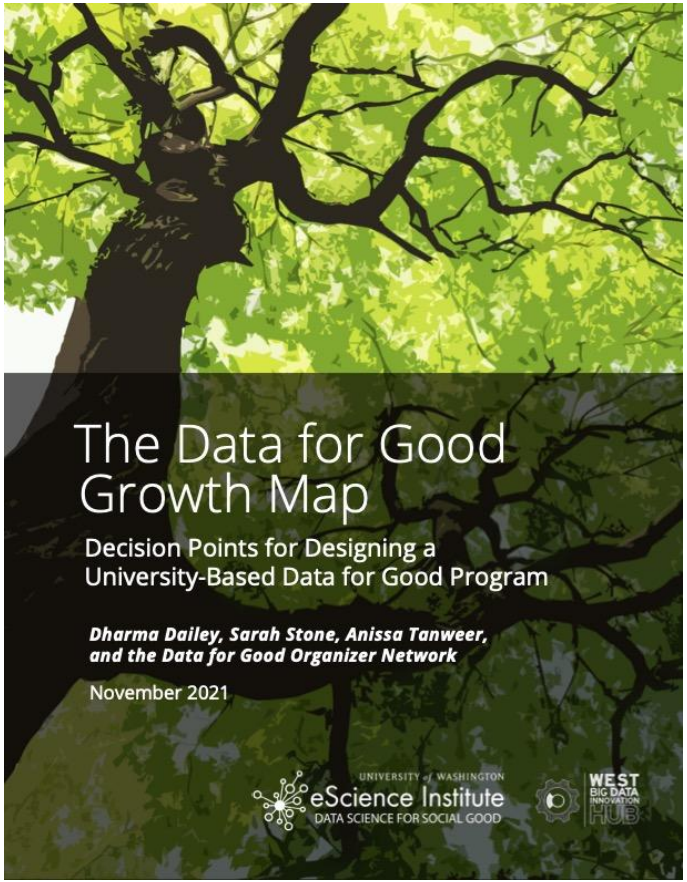
# Program Overview

UNIVERSITY of WASHINGTON

# eScience Institute

## DATA SCIENCE FOR SOCIAL GOOD


Modeled after similar programs with elements from our own Data Science Incubator.




# Data for Good Organizer Network

# Program Goals

- Figure out what it means to do “good” with data science
- Train students in data science methods
- Increase data science capacity across fields and organizations
- Positively impact society

# Mode of Participation

- Students are required to work in-person (with some flexibility)
- Project leads and data scientists may work **in-person, remote or hybrid**
- Program staff will be hybrid
- We will take on two projects this year

UNIVERSITY of WASHINGTON

# Team Composition

DSSG Student Fellows (4)
eScience Data Scientist Leads (1-2)
Project Leads (1-2)

# What Project Leads Get

- Intensive work on project
- Exposure to new methods and approaches
- Interdisciplinary teamwork
- Networking opportunities
- Publicity

# Examples of Project Lead Affiliations

- University of Washington (academia)
  - Disaster Data Science Lab
  - School of Social Work
  - Department of Astronomy
- Seattle Department of Transportation (gov)
- Bill & Melinda Gates Foundation (philanthropy)
- Conservation International (nonprofit)

# What we expect from Project Leads

- Scoping meetings in preparation
- Available to team 16 hours/wk on average
  - Probably more during first 2 weeks
  - For in-person teams, working in the Data Science Studio during this time
- Participation in program-wide sessions & meetings
- Domain expertise
- Stakeholder engagement
- Ability to discuss and promote work
- Open & reproducible when possible (Github)
- Description of project on our website
- Acknowledgement in publications

# What we expect from students

- 40 hours/wk ($9,000 stipend)
- Current student: grad or senior entering grad school
- Baseline programming and statistics knowledge
- Eligible to work in US (can’t support visas)
- Strong personal statement
- Team player

# What you can expect from us

- Data scientists highly experienced in cross-disciplinary collaboration
- Expertise in (non-exhaustive):

| | |
|---|---|
| Machine learning | Statistical inference |
| Databases | GIS |
| Modeling | Optimization |
| Visualization | Cloud computing |

- Best practices in version control, reproducibility, ethics, and human-centered design
- Help with team management

# Tutorials

- Intro to Git & GitHub*
- Git and Git Workflow*
- Team Management Processes
- Pandas, Geopandas, and SQL
- Python Coding Standards and Documentation
- Unit Tests
- Project Organization, Virtualization, Continuous Integration
- Pair Programming
- Object Oriented Python
- Software Design
- Machine Learning
- Web Design and Web Apps
- Cython/Dask/High Performance Python
- Vega/Altair
- Data Visualization with Tableau
- An Introduction to Visual Communication

# Workshops

- Introduction to Data Science for Social Good
- Team Development
- Preparing for Stakeholder Engagement
- Stakeholder Analysis & Speculative Ethics
- Best Practices in Public Speaking
- Final Presentation Practice Talks

# Other Activities

- Project Spotlights
- Docathons
- Career Conversations
- Stakeholder Engagement
- Social Events

# Project Selection Process

# Call for Proposals is NOW OPEN!

**General Info:**
https://escience.washington.edu/using-data-science/data-science-for-social-good/submit-project/dssg-call-for-proposals/

**FAQ's:**
https://escience.washington.edu/using-data-science/data-science-for-social-good/submit-project/project-lead-faqs/

# We encourage you to reach out and meet with us before submitting a proposal

Office Hours: https://escience.washington.edu/office-hours/

# What we're looking for

- Argument in support of how project will lead to positive social impact
- Strong research, strong methods
- Availability and commitment
- Clarity and shovel-readiness (data in hand)
- Capacity for measurable outcomes
- Sustained engagement

# What we *don’t* do

- Build web portals
- App development as primary goal
- Data collection

# A non-exhaustive list of topical interests

- Affordability and livability
- City planning
- Environmental issues
- Equity and justice
- Hazards and resilience
- Health and wellness
- Housing
- Public education
- Sociodemographic disparities
- Transportation

# Technical areas of eScience expertise

- New platforms, new algorithms, new methods, new datasets
- Working with large, heterogeneous, and noisy datasets
- Scalable analytics and predictive models
- Interactive visualization
- Code review, publishing, and reproducibility
- Online teaching materials, tutorials








# Program structure

## Pre-Program

- Meetings with data scientists (2-3)
- Project Lead orientation

## First Two Weeks

- Mandatory team development workshops (may require more than 16 hrs total)
- Front-loaded tutorials

## Rest of Summer

- Weekly “project spotlight” meetings
- Regularly scheduled team check-ins
- Bi-weekly check-ins
- Occasional tutorials (can be on-demand)
- Visits and calls with stakeholders

## End of Summer

- Final presentations and reception

# Important Dates

Jan. 3 - Call for Proposals opened
Jan. 19 - Info Session for prospective project leads
Feb. 17 - Project proposals due
March 1 - Project short-list notifications
April 10 - Student offers of admission
Mar. - Jun. - Meetings with DS & PL
June 12th - First day of program
August 18th - Last day of program

# UW DSSG Projects

# Summer 2022 Projects

**Cost of living with the Self-Sufficiency Standard**

Project Leads: Annie Kuclick and Lisa Manzer, UW Center for Women's Welfare

**Tracking family & intergenerational poverty**

Project Lead: Jennie Romich, UW School of Social Work; West Coast Poverty Center

**Satellite Streaks in Astronomical Images**

Project Leads: Meredith Rawls and Dino Bektešević and, UW Dept. of Astronomy

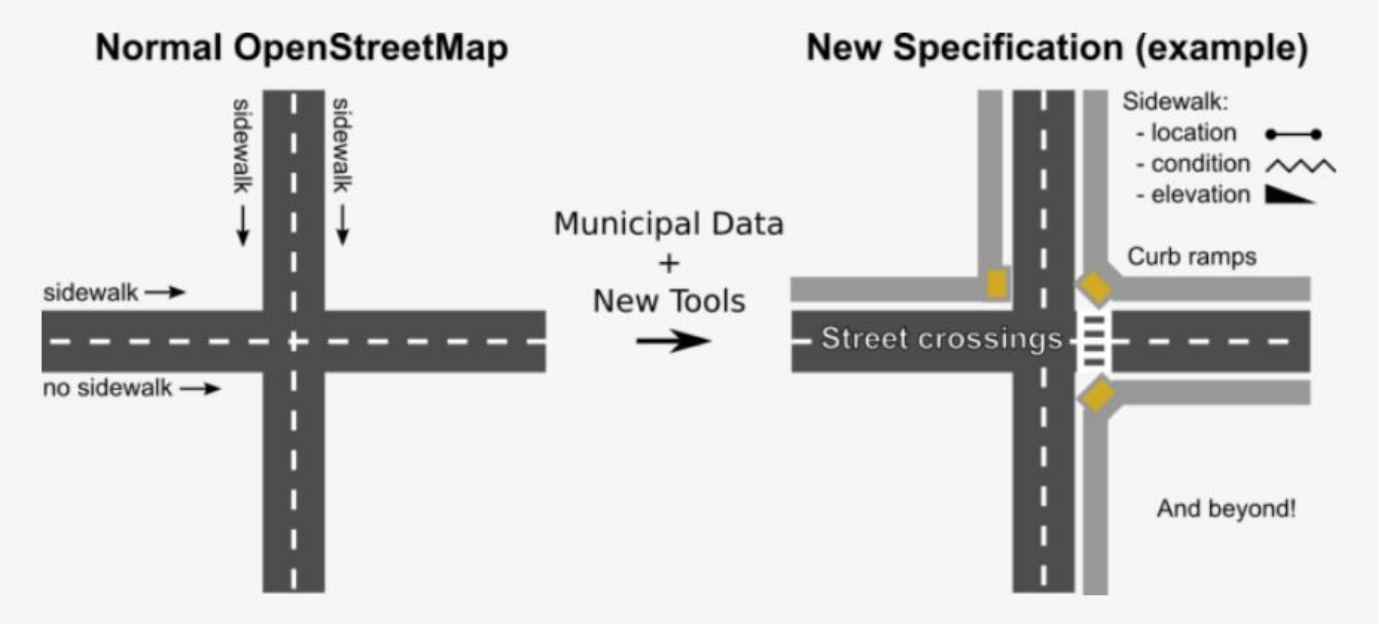
**Heating Loads in Alaska and Beyond**

Project Lead: Erin Trochim, Alaska Center for Energy and Power, Univ. of Alaska, Fairbanks

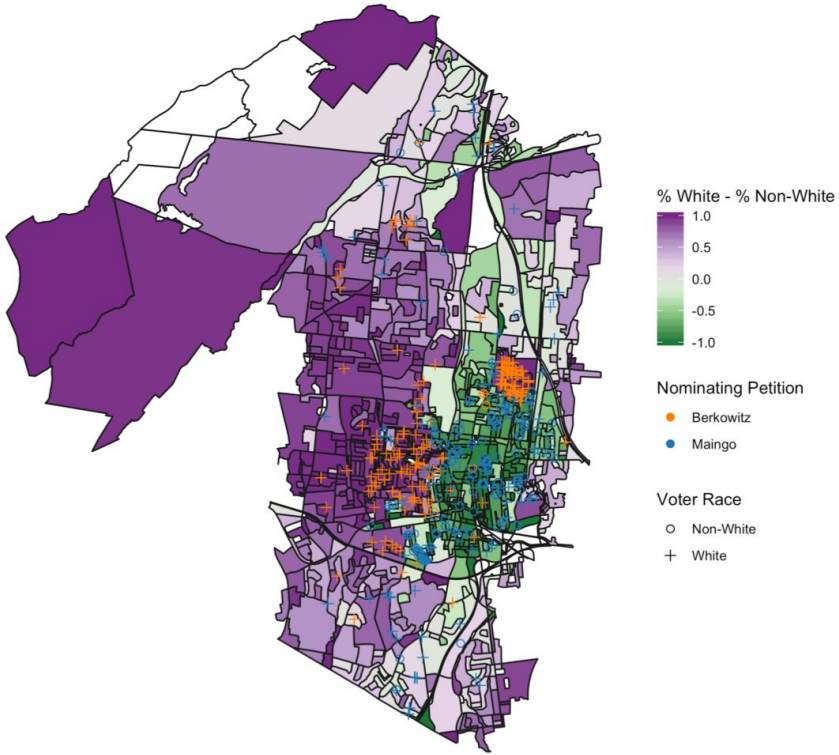
# We have a broad view of what counts as data science

# Mining Online Data for Early Identification of Unsafe Food Products

# Global Open Sidewalks: Creating a shared open data layer and an OpenStreetMap data standard for sidewalks



# Detection of Vote Dilution: New tools and methods for protecting voting rights




MDG


**"What surprised me about the program, I was not expecting for it to be as much fun as it was."**

**- Jennie Romich, Project Lead (2021 & 2022)**

# Questions?

Anissa Tanweer, Program Director: tanweer@uw.edu