

UNIVERSITY *of* WASHINGTON

eScience Institute

ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

April 2022

Sarah Stone, Executive Director, sstone3@uw.edu

Vision for today's talk

- History, mission, and governance
- Growth and impact
- **Core programs on campus and beyond**
- **Plans and next steps**

Institutional History

“All across our campus, the process of discovery will increasingly rely on researchers’ ability to extract knowledge from vast amounts of data... In order to remain at the forefront, UW must be a leader in advancing these techniques and technologies, and in making [them] accessible to researchers in the broadest imaginable range of fields.”

2008 - Institute founding, support from state budget line

2013 - Moore-Sloan Data Science Environment

2014 - WRF award

2018 - Permanent UW support from the Provost

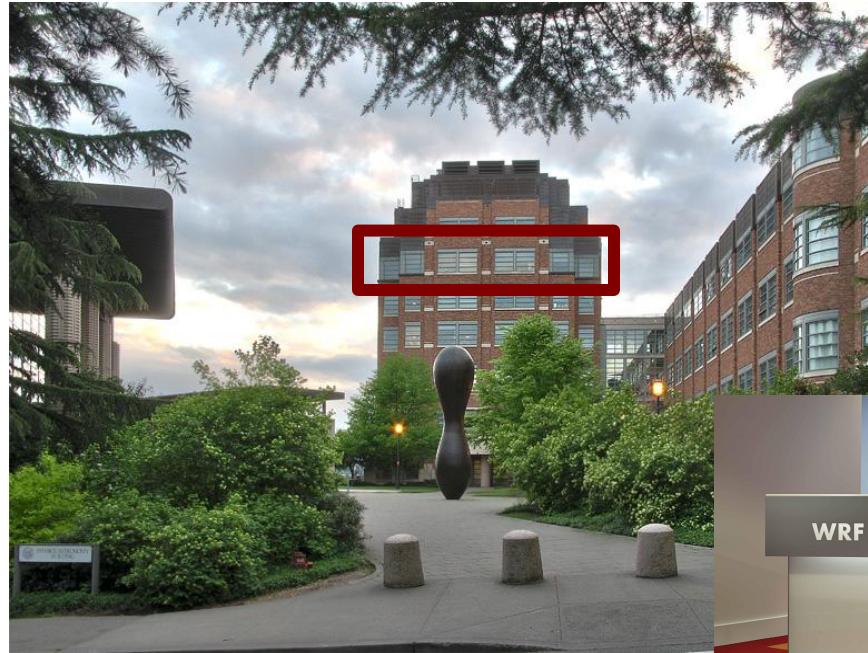
Our Mission

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



WRF Data Science Studio

Thanks to partnership with UW Libraries



UNIVERSITY
LIBRARIES

W

WRF Data Science Studio



Governance - Executive Committee



Andy Connolly
eScience Director;
Professor Astronomy



David Beck
Director of Research;
Research Associate Prof.,
Chemical Engineering



Sarah Stone
Executive Director



Ed Lazowska
Professor, Paul G. Allen
School of Computer
Science & Engineering,
Founding Director



Cecilia Aragon
Director, Human-Centered
DS Lab; Prof., HCDE



Ginger Armbrust
Professor, School of
Oceanography



Magdalena Balazinska
Director Paul G. Allen
School of Computer
Science & Engineering,
Former Director (2017-
2019)



Bingni Brunton
Associate Professor,
Biology



Rob Fatland
UW-IT Director of Cloud
and Data Solutions



Maryam Fazel
Professor, Electrical &
Computer Engineering



Bill Howe
Associate Professor,
iSchool; Former Associate
Director



Ben Marwick
Associate Professor,
Anthropology



Tyler McCormick
Associate Prof., Statistics,
Sociology



Colette Moore
Associate Professor,
English

[A&S Natural Sciences, A&S Social Sciences, Engineering, Environment, iSchool, UW IT]

eScience Research Team

Director of Research



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

Data Scientists



Noah Benson
Ph.D. Biomedical & Health Informatics



Bernease Herman
B.S. Statistics
Formerly SE at Amazon



Valentina Staneva
Ph.D. Applied Mathematics and Statistics



Anissa Tanweer
Ph.D. Communication

Technical Education Specialist



Naomi Alterman
M.S. Electrical Engineering

Research Scientists



Anthony Arendt
Ph.D. Geophysics
APL



Nicoleta Cristea
Ph.D. Environmental Engineering



Bryna Hazelton
Ph.D. Astrophysics Physics



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



Scott Henderson
Ph.D. Geological Sciences

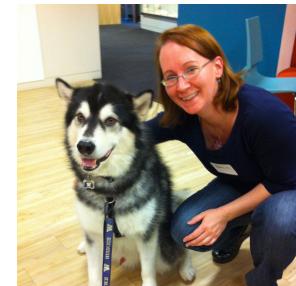


Vaughn Iverson
Ph.D. Oceanography



Spencer Wood
Ph.D. Zoology

eScience Administrative Team



Sarah Stone
Executive Director

Louisa Gaylord
Communications
Specialist

Jenny Holcomb
Office Coordinator

Janet Kim
Fiscal Specialist

Jane Koh
Program Specialist

Ann Nykamp
Grants Manager

Expanded administrative team allows for:

- complex event support and communications
- full proposal and award support
- *NEW* - remote event consultation



P. "Charley" Haley

**Consultant -
Remote Event
Facilitation**

Data Science Faculty Hires



Zack Almquist
Sociology
NEW MARCH 2020



Cecilia Aragon
HCDE



Sasha Aravkin
Applied Mathematics



Bingni Brunton
Biology



Steve Brunton
Mechanical Engineering



Marine Denolle
Earth & Space Sciences
STARTED JAN 2021



Zaid Harchaoui
Statistics



Mario Juric
Astronomy

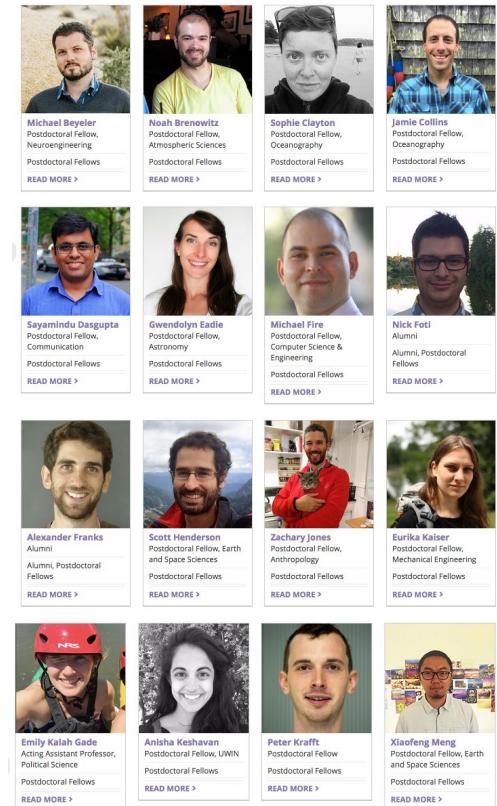


Sham Kakade
CSE + Statistics

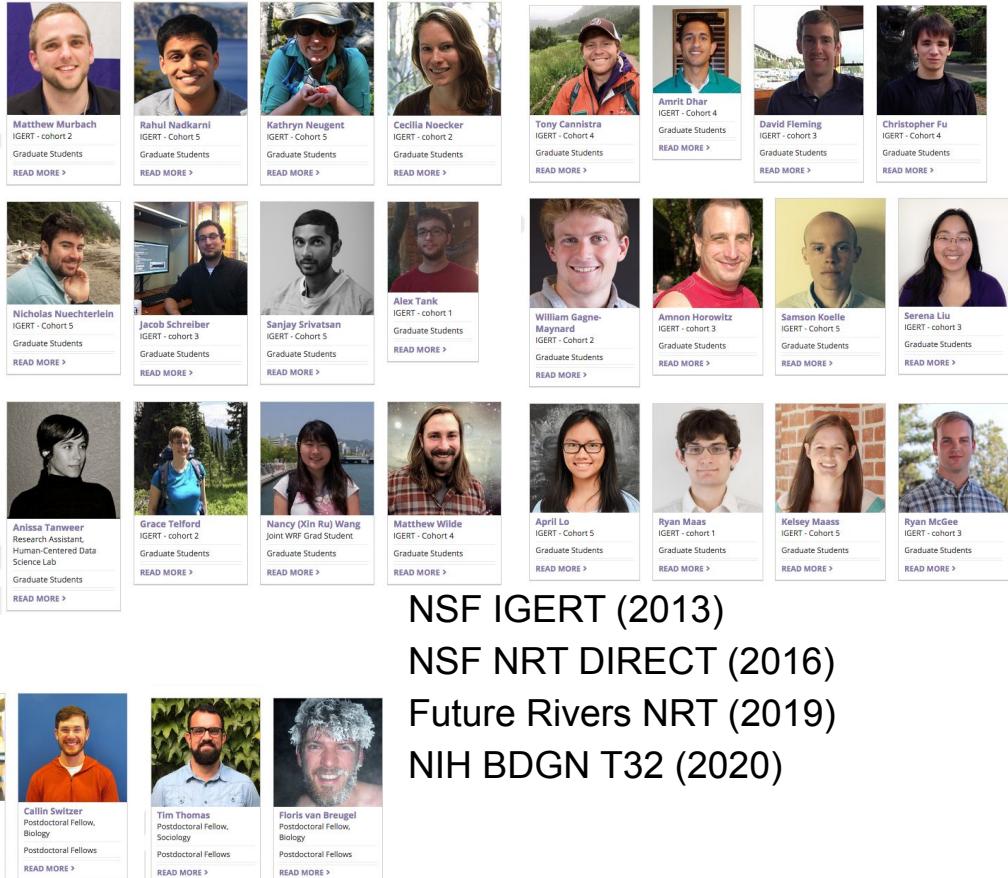


Eli Shlizerman
Applied Mathematics

...plus more than 120 affiliated faculty



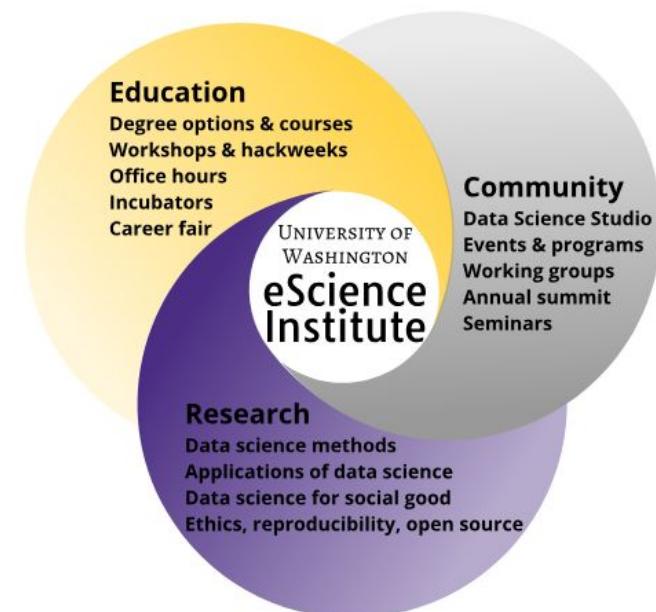
Postdocs and PhD Students



UW Data Science Postdoctoral Fellow Titling Program - 60+ postdocs to date

eScience Institute Role at the U of Washington and Beyond

- **Education**
 - We have built formal & informal data science education programs.
- **Research**
 - Through our programs & partnerships in funded research, we continually look for ways to support UW researchers.
- **Community**
 - Partnerships extend our reach within & beyond UW.



Formal Data Science Education Programs

Launched

PhD

Advanced Data Science Option
Data Science Option

funded - 2013
option - 2016
2017

Undergrad

Data Science Option
Data Science Minor

2016
2020 NEW!

Professional

Data Science Master's
Professional Certificates

2016
various

Together with Many Departments/Schools/Colleges

Participating Departments

(others pending approval)

undergraduate only
graduate only
both

ACMS

Applied
Math

Astronomy

Atmospheric
Sciences

BIME

Bioeng.

Biology

Chem.

Chem. E.

CEE

Comp. Sci.
& Eng.

Genome
Sciences

Geography

HCDE

iSchool

Math

ME

Mol. & Cell.
Biology

MoES

Oceano.

Psychology

Statistics

Data Science Minor

NEW in Fall 2020

271 students

currently declared

- Help students to **leverage familiarity** with data science in fields outside of data science.
- Help students **gain skills and fluency** to work with data in their major domain of study.
- Set students up for success in emergent “**translator**” roles...poised to provide expertise in policy, design, social theory, ethics on a data science team.
- Serve as a **stepping stone** toward acquiring more advanced skills and degrees in data science.
- Currently 20+ depts/units involved
- eScience serves as the institutional home for the DS Minor

We Disseminate Data Science Expertise and Best Practices

Open Office Hours

**UW Data Science Seminar & Data
Science Coast to Coast Seminar**

Tutorials, bootcamps, workshops

- Software Carpentry
- Cloud 101
- ...

Hack Weeks

- Astrohackweek, Geohackweek,
Oceanhackweek, Waterhackweek,
Neurohackademy



Office hours

(*free* Drop-In Consulting - now via Zoom)



Bryna Hazelton
brynah@phys.washington.edu

Hours
Will resume in the fall.

- Expertise
- ➊ Fourier and image analysis
 - ➋ Statistical and mathematical modeling
 - ➋ Monte Carlo simulations
 - ➋ Python
 - ➋ SQL
 - ➋ IDL
 - ➋ Java



Bernease Herman
bernease@uw.edu

Hours
Will resume in the fall.

- Expertise
- ➊ Python, R, Scala
 - ➋ Machine learning
 - ➋ Interpretable models
 - ➋ Data visualization (D3.js, Leaflet, Shiny)
 - ➋ Version control (Git/GitHub)
 - ➋ Reproducibility and open science



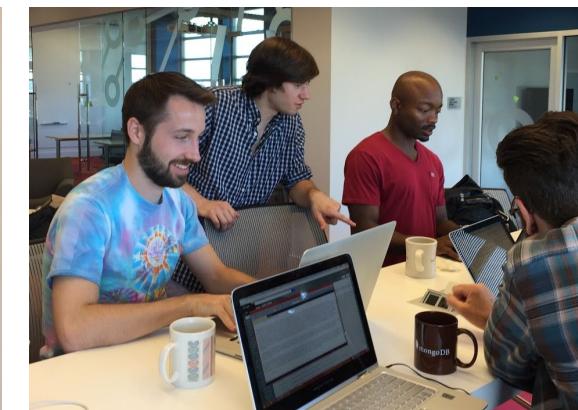
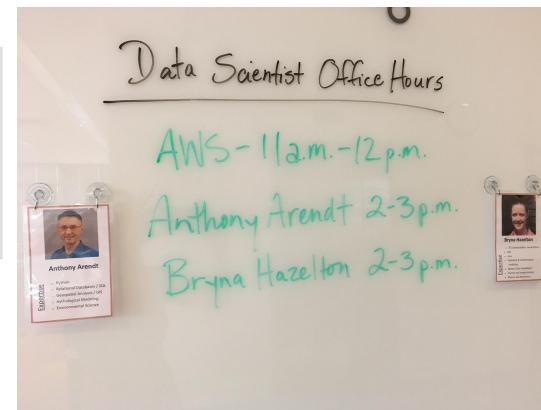
Joseph Hellerstein
joseph.hellerstein@gmail.com

Hours
Will resume in the fall.

- Expertise
- ➊ C
 - ➋ R
 - ➋ Python
 - ➋ Bash
 - ➋ SQL
 - ➋ Software design
 - ➋ Web server design
 - ➋ Reproducibility and open science



UNIVERSITY LIBRARIES



UW Data Science Seminar

*...thought-leading
speakers and researchers
who showcase data
science in the arts,
engineering, humanities
and sciences...*

Weekly seminar * 3 quarters

Typically, ~80-100 attendees

Feb 28		"Computer vision challenges in wildlife conservation" Dan Morris, Principal Researcher, Microsoft AI for Earth
Apr 3		"The battle of two cultures: statistics versus (?) data science" Bhramar Mukherjee, Ph.D., John D. Kalbfleisch Collegiate Professor of Biostatistics, University of Michigan
May 15		"Multi-group covariance estimation with applications to 'omics data" Alexander Franks, Assistant Professor, Department of Statistics and Applied Probability, University of California, Santa Barbara
May 22		"The life of data in the city" Paul Dourish, Chancellor's Professor of Informatics and Associate Dean for Research in the Donald Bren School of Information and Computer Sciences, University of California, Irvine
May 29		"Data is people: ethical considerations in data collection and use" Casey Fiesler, Assistant Professor, Department of Information Science, University of Colorado Boulder
Nov 19		"R and friends for better science in less time in big (and small) team collaborations" Julia Stewart Lowndes, Ph.D., marine ecologist, data scientist, and Mozilla Fellow at the National Center for Ecological Analysis and Synthesis (NCEAS), USA.
Dec 3		"Euclidean Neural Networks* for Emulating Ab Initio Calculations and Generating Atomic Geometries" Tess Smidt, Ph.D., Alvarez Postdoctoral Fellow in Computing Sciences at Lawrence Berkeley National Laboratory.

Various short courses and workshops



Python, your personal research assistant

WHEN Friday, Apr 8, 2022, 10 – 11:30 a.m.

EVENT TYPES Workshops

EVENT SPONSORS The [eScience Institute](#) and the [Open Scholarship Commons](#)

In this workshop, attendees will be introduced to the Python programming language as a tool to aid in their qualitative humanities work. Together, we will learn how to read and make use of Python code to perform repetitive tasks and free ourselves to do the meaningfully human parts of our research. Attendees will leave the workshop with tools to sort Twitter tweets by tone, create random project groups from lists of names, and resources for further learning. No prior technological experience is assumed -- this is a workshop for everybody!

Cloud Builders' Workshop: Zero to Jupyterhub

About this event

A hands-on workshop for computing directors and IT personnel to learn how to stand up and cost scalable multi-user computing environments using Kubernetes and Jupyter.

Date and time

Tue, May 31, 2022
10:00 AM – 1:00 PM PDT

Location

Online event
Already have a ticket?

Documentation for Cloud Computing and Data Science

Summary: Cloud computing tutorials, case studies, pathways into data science technology; presenting the public cloud as opportunity for innovation. This page is intended to accelerate research; developed in close association with the [UW eScience Institute](<http://escience.washington.edu>). We have two goals - Reduce friction, i.e. help with what you already do with computers, streamline your analysis and pathways to publication. Promote new ideas and methods, i.e. We want to make it easy for you to share data, collaborate, sandbox, and learn about new methods in data science.

Help us improve this website, take a short survey [here](#).

Announcements & News

- [Cloud 101 for Google Cloud Platform](#) will be held on Feb. 13, 2018 from 9am to 12pm. Registration now open.
- [Cloud 101 for Amazon Web Services, Microsoft Azure & Google Cloud Platform](#) will be held from Nov. 13 - Nov. 15, 2017 ** REGISTRATION IS NOW CLOSED. WE HAVE FILLED TO CAPACITY**
- [GeohackWeek](#) was held from Sept 11-15, 2017
- [Cloud 101 for the Foster School of Business](#) was presented on May 25, 2017.
- [Research Computing Cloud 101 Immersion Day](#) was April 4, 2017.

Hack weeks - Community based learning goes remote

NEUROHACKADEMY

ASTRO HACK WEEK 2020

AUG 31 — SEP 4, 2020

FLATIRON INSTITUTE @ SIMONS FOUNDATION IN NEW YORK, USA

WATERHACKWEEK 2020

WORKSHOP ON WATER DATA SCIENCE
UNIVERSITY OF WASHINGTON ESCIENCE INSTITUTE

OCEANHACKWEEK 2020

DATA SCIENCE + OCEANOGRAPHY
A HANDS-ON, INTERACTIVE VIRTUAL WORKSHOP

GEOHACKWEEK 2019

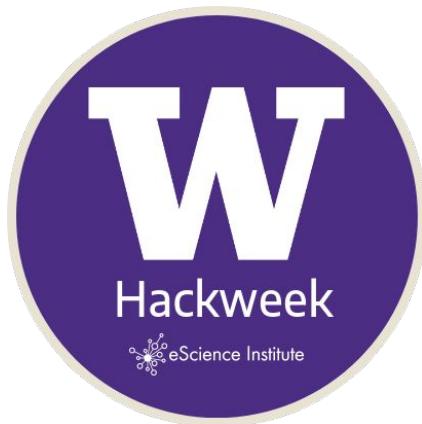
WORKSHOP ON GEOSPATIAL SCIENCE

- Quickly recruited and onboarded a remote facilitation expert in April 2020
- Organized a peer community of organizers of 10+ similar programs
 - Supported other departments in transition to remote programs
- Acquired funding from the Sloan Foundation in support of this work
- Building out a Hackweek Toolkit



**Growing portfolio of hackweeks
450+ Participants in 2020**

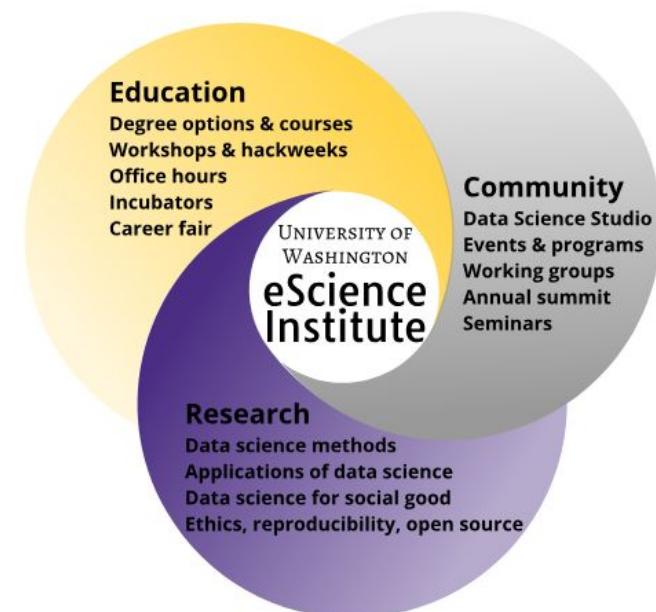
Hackweeks as a service



Sense of Belonging	Education	Immersive Learning
		
We foster a welcoming, diverse and inclusive environment.	We gain experience in conducting data-intensive research.	We encourage immersive, interactive and collaborative learning.
Co-creation	Networking	Code of Conduct
		
We incorporate elements of co-creation with participants.	We provide opportunities for people to build their networks.	We adhere to and enforce a shared Code of Conduct.

eScience Institute Role at the U of Washington and Beyond

- **Education**
 - We have built formal & informal data science education programs.
- **Research**
 - Through our programs & partnerships in funded research, we continually look for ways to support UW researchers.
- **Community**
 - Partnerships extend our reach within & beyond UW.



Data Science Incubator Program - *Scalable Research Impact*

- Move from “accidental” encounters to engineered quarter-long+ partnerships
- Provide a shared environment where researchers can learn from our team and each other



=> 55 projects over the past 7 years, 30 departments represented
=> 5-7 projects per year

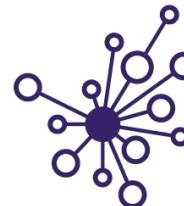
Example projects from Winter 2021 Incubator

- Using Gliders to Observe Submesoscale Flows – **Oceanography**
- The Universe of International Treaties – **Political Science**
- Dry Thunderstorm Forecast Using Machine Learning Techniques – **Atmospheric Sciences**
- Learning to See the Forest and the Trees: Using Computer Vision to Make Forest Stewardship More Accessible – **Environmental and Forest Sciences**
- Climate Adaptation for Future Maize – Novel Plant Traits and New Management –**Biology**
- Detecting Wildflowers in Spectral Imagery – **Biology**
- Using Social Media to Model Backcountry Use in Rainier National Park – **Environmental and Forest Sciences**

Incubator reflection

“The program is state-of-the-art with respect to mentoring, programming instruction, and machine learning didactic training...The combination of these regular one-on-one meetings and a weekly group meeting with other incubator mentees and mentors provided a rich environment where I could ask a programming or data science question and obtain an answer immediately...The UW eScience incubator has my full endorsement and I would highly recommend it to anyone who is starting out in programming or anyone who has intermediate experience in programming and data science.”

- Charles Zhou, Anesthesiology & Pain Medicine,
Winter Incubator 2020



Data Science for Social Good - *Timely, publicly relevant projects*

- 8-16 Student Fellows from universities across the country work on collaborative projects
- 2-4 Projects sourced from academia, public agencies, & non-profits through an open proposal process
- Annually, since summer 2015



<https://escience.washington.edu/dssg/>

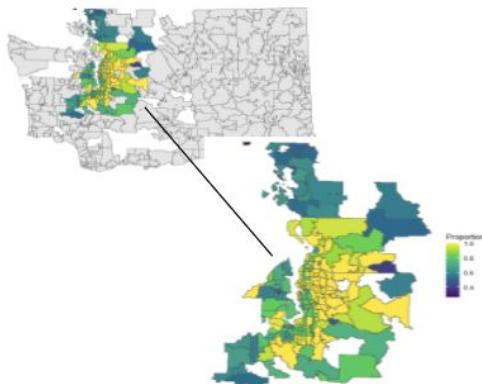
Summer 2021 DSSG Projects



Developing Ensemble Methods for Initial Districting Plan Evaluation

Daryl DeFord, Project Lead

Assistant Professor of Data Analytics in the Department of Mathematics and Statistics at Washington State University



Geography, Equity, and the Seattle \$15 Minimum Wage Ordinance

Jennie Romich, Project Lead

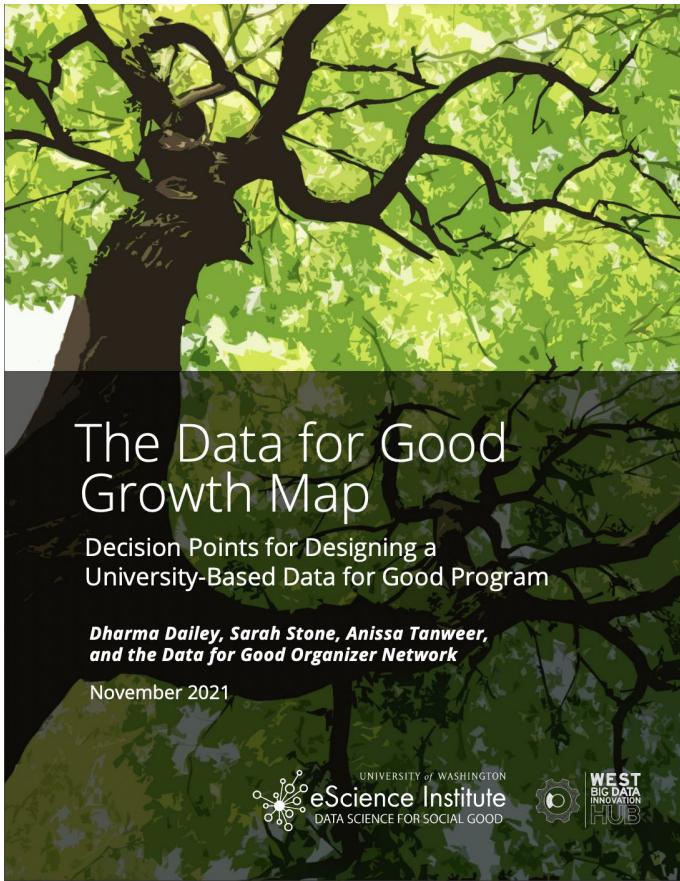
Professor of Social Welfare at the UW School of Social Work and Faculty Director of the West Coast Poverty Center

Data Science for Social Good success story

Taskar Center for Accessible Technology awarded Federal Highway Administration (FHWA), 5-year \$11.5M award

“OpenSidewalks wouldn’t be where it is without the DSSG summer 2016 program, and that’s also where I met Mark Hallenbeck who is a co-PI on this.”

- Anat Caspi, Director
Taskar Center
2016 Project Lead

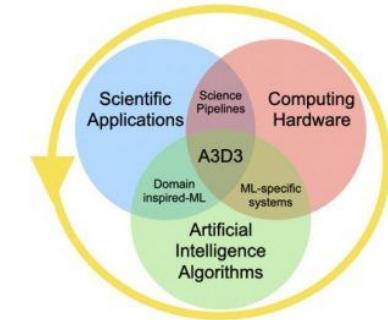


Data for Good Organizer Network

- Carnegie Mellon University (formerly University of Chicago)
- Iowa State University
- Stanford University
- University of British Columbia
- University of Massachusetts Amherst
- University of North Florida
- University of Virginia
- University of Washington
- University of Warwick

http://escience.washington.edu/wp-content/uploads/2021/11/Data4Good_GrowthMap.pdf

Ongoing projects & collaborations



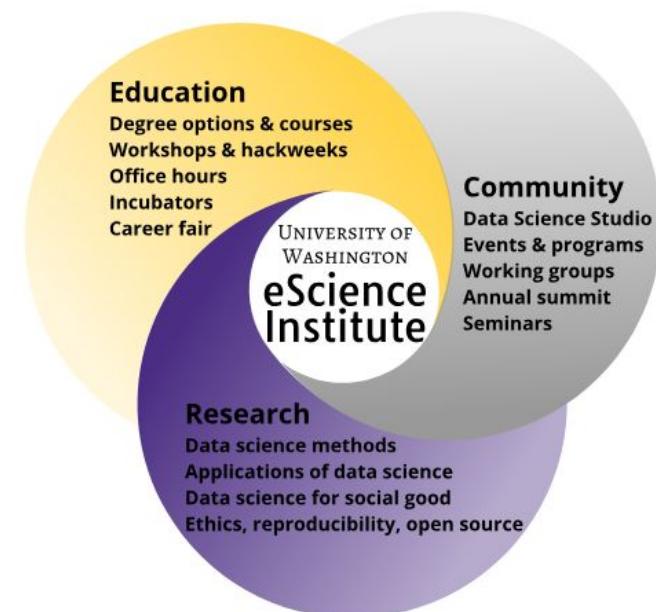
DiRAC

**AI Institute
in Dynamic
Systems**

FUTURE RIVERS

eScience Institute Role at the U of Washington and Beyond

- **Education**
 - We have built formal & informal data science education programs.
- **Research**
 - Through our programs & partnerships in funded research, we continually look for ways to support UW researchers.
- **Community**
 - Partnerships extend our reach within & beyond UW.



UW Scientific Software Engineering Center (SSEC)

Software engineering is vital to Seattle's emergence as a tech leader driving a local climate of innovation. The new Scientific Software Engineering Center (SSEC) at UW's eScience Institute will leverage local software engineering talent to advance scientific frontiers that will change the world around us and respond to the challenges that face humanity.

See recent media coverage in *Science* and *Geekwire*.



NOW HIRING – We are looking for experienced software engineers, including a lead software engineer, to work on impactful science research projects, infusing them with the software industry's best practices, and delivering reusable, open source software that will accelerate future research in areas like climate, health, energy and basic science.

- **SSEC Lead – ACCEPTING APPLICATIONS OPEN NOW!**
- SSEC Principal Research Software Engineers – opening soon!
- SSEC Senior Research Software Engineers – opening soon!



Discovering AI@UW Event - May 23rd, HUB Lyceum, UW Seattle campus

Be in conversation with AI experts from across the University of Washington campus

This inspirational education and networking opportunity for faculty, researchers, and students will feature:

- Talks by funded AI centers at UW
- Prominent experts discussing cutting edge AI methodologies
- Novel AI applications in areas such as environmental science, engineering, astronomy and sociology
- Breakout sessions featuring panel discussions
- Poster presentations highlighting AI research by postdocs and graduate students at UW
- Social gatherings including light breakfast, lunch, and a celebratory reception

UW Data Science Recruiting Event

October 12, 2020 12:00 PM to 4:00 PM Pacific Daylight Time (PDT)

Transitioned to fully remote platform - over 500 students signed up!



MAQ Software



Join our mailing list...

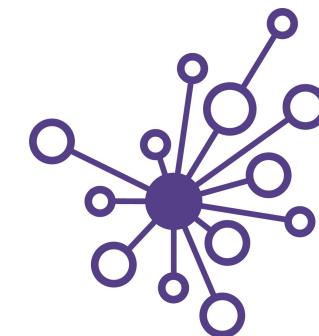
simply “opt-in” on escience.washington.edu

Follow us on social media...

Twitter, @uwescience

LinkedIn

YouTube, UW eScience Institute



Thank You!

GORDON AND BETTY
MOORE
FOUNDATION



Micron[®]



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

