Data Science for Social Good

Information session for prospective project lead applicants

Sarah Stone
Executive Director, eScience
Program Director, DSSG

Anissa Tanweer
Research Scientist, eScience
Program Chair, DSSG
Info session outline

- Introduction to the eScience Institute
- Data Science for Social Good (DSSG)
  - Program overview
  - Proposal process
  - Program logistics
  - Previous projects
- Questions
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.
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
- Jose Hernandez
  Ph.D. Measurement & Statistics
- Valentina Staneva
  Ph.D. Applied Mathematics and Statistics
- Amanda Tan
  Ph.D. Civil & Env. Engineering
- Anissa Tanweer
  Ph.D. Communication

**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
We disseminate data science expertise and best practices

- Open Office Hours
- UW Data Science Seminar
- Tutorials, bootcamps, workshops, and hack weeks
  - Neurohackademy, Geohack, Waterhack, Oceanhack, etc.
  - Software carpentry (> 400 participants since we started counting in 2015)
- Winter Incubator
- Summer DSSG
Modeled after similar programs with elements from our own Data Science Incubator.
DSSG 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
Team Composition

- Student Fellows (4-5)
- eScience Data Scientist Leads (1-2)
- Project Leads (1-2)
What Project Leads Get

- Intensive work on project
- Exposure to new methods, tools and approaches
- Interdisciplinary teamwork
- Networking opportunities
- Publicity
Examples of Project Lead Affiliations

- **Academia**: E.g. University of Washington
  - Washington State Transportation Center
  - Disaster Data Science Lab
  - Architecture Department
- **Government**: E.g. Seattle Department of Transportation
- **Nonprofits**: E.g. Conservation International
- **Industry**: E.g. Bell Labs
What we expect from Project Leads

- Scoping meetings in preparation
- Co-presence 16 hrs/wk on average
  * Probably more during first 2 weeks
- 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
- Acknowledgment of the program in publications & authorship for team members
What we expect from Student Fellows

- 40 hours/week ($7,000 stipend)
- Current student, grad and advanced undergrad
- Baseline programming and stats knowledge
- Eligible to work in US (can’t support visas)
- Strong personal statement
- Team player
What you can expect from eScience

- Selection of talented, hardworking students
- 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 and human-centered design
- Help with team management
- Support in promoting and disseminating your work
Call for Proposals is open now!

General Info:
https://escience.washington.edu/dssg-proposal

FAQ’s:
https://escience.washington.edu/dssg-pl-faq/

Application Form:
https://form.jotform.com/203355099665161

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

- Data Scientist Office Hours:
  https://escience.washington.edu/office-hours/#eScienceDataScientists
- Program Chair Anissa Tanweer:
  tanweer@uw.edu
What we look for in project proposals

- Strong argument in support of how project will lead to positive social impact
- Availability and commitment of Project Lead
- Strong research, strong methods
- Clarity and shovel-readiness
- Capacity for measurable outcomes
- Evidence of implementability and sustainability
Projects we CANNOT support

- Building web portals
- App development as primary goal
- Data collection
A non-exhaustive list of our interests

- Poverty, equity, income
- Housing and homelessness
- Public education
- City planning
- Transportation
- Hazards/resilience
- Utilities
- Environment
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
We take a broad view of what counts as data science
Proof of Concept  ->  Polished Product

Mining Online Data for Early Identification of Unsafe Food Products

- DSSG 2016

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

- DSSG 2020
Proof of Concept
Infrastructural

Polished Product
Analytical

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

- DSSG 2016

Strengthening Capacities, Knowledge and Data Sharing Platforms for Sustainable Development

- DSSG 2017
CrowdSensing Census: A heterogeneous-based tool for estimating poverty

ADUniverse: Evaluating the Feasibility of (Affordable) Accessory Dwelling Units in Seattle

- DSSG 2016

- DSSG 2019
Proof of Concept
Infrastructural
Novel Data
Inference

Polished Product
Analytical
Traditional Data
Automation

Access to Out-of-School Opportunities and Student Outcomes - DSSG 2018

Automatic Damage Annotation on Post-Hurricane Satellite Imagery - DSSG 2018
Proof of Concept
Infrastructural Novel Data Inference

Polished Product Analytical Traditional Data Automation
Overview of DSSG Program Structure

Pre-Program
- Meetings with data scientists (generally 2-3, sometimes more)
- Project Lead orientation

First Two Weeks
- Mandatory team development workshops (requires flexibility during this time)
- Front-loaded tutorials

Rest of Summer
- Occasional tutorials (can be on-demand)
- Weekly “project spotlight” meetings and program check-in
- Bi-weekly leadership meeting with all PL’s, DS’s and administrators
- Visits and calls with stakeholders
- Project work and regularly scheduled team meetings

End of Summer
- Final presentations and celebration
Running DSSG Remotely

- Zoom, Github, Slack, Google Calendar, and other technologies
- Schedule program-wide meetings with consideration for multiple time zones
- Flexibility in coordinating teamwork necessary
- Intentionality in cultivating social connectedness
- Support for students in challenging physical environments
- High degrees of satisfaction with last year’s remote offering
Important Dates

Now        Call for Proposals open
Now        Student applications open
Jan 26     Student Info Session
Feb 22     Project proposals due ***
Mar 2       Project shortlist notifications ***
Mar-Apr     Follow-up meetings re: shortlisted proposals
Apr 7      Project acceptance notifications ***
Apr 9      Student selection completed
Apr-Jun     Scoping Meetings with DS & PL
Jun 14     First day DSSG ***
Aug 20     Last day DSSG
Questions?

Anissa Tanweer
Research Scientist, eScience
Program Chair, DSSG
tanweer@uw.edu
Summer 2020 Projects

Identifying Coronavirus Disinformation Risk on News Websites

Maggie Engler, Lead Data Scientist, Global Disinformation Index (GDI)

Lucas Wright, Senior Researcher, Global Disinformation Index (GDI)

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

Matt A. Barreto, Professor of Political Science and Chicana/o Studies, UCLA

Loren Collingwood, Associate Professor in the Department of Political Science, UC Riverside