Summer 2018 Data+ Proposal Guide, for Faculty Sponsors

September 8, 2017

1 Overview

Data+ is a ten-week summer research experience that welcomes Duke undergraduates interested in exploring new data-driven approaches to interdisciplinary challenges.

Students join small project teams, working alongside other teams in a communal environment. They learn how to marshal, analyze, and visualize data, while gaining broad exposure to the modern world of data science. In Summer 2017 there were 25 such teams, and they all worked together in Gross Hall, sitting in dedicated workspace provided by the Information Initiative at Duke (iiD), the Social Science Research Institute (SSRI), the Energy Initiative, and the Foundry. Each undergraduate participant receives a 5,000 dollar stipend. Each summer Data+ runs from mid-May until the end of July. As the communal atmosphere is essential for student success, **please note that Data+ projects only run during these ten weeks, and that all student participants are required to contribute full-time efforts (no employment, no other classes)**.

This document is a call-for-proposals for faculty-sponsored Data+ projects in the Summer 2018 edition of Data+. We are especially interested in proposals that involve a partner from outside the academy, or a faculty member from a different discipline. We also encourage proposals that involve previously untested ideas or un-analyzed datasets, and we hope that the Data+ team can make a contribution with important proof-of-principle work that may lead to more substantial faculty work and/or connections in the future. We also welcome proposals that will lead to the undergraduates creating tools that might be used in the classroom or that might facilitate community engagement with data and data-driven questions.

Proposals should be emailed to Ariel Dawn (ariel.dawn@duke.edu) by **November 1st, 2017**.

If you would like help in developing your proposal, please contact Paul Bendich (bendich@math.duke.edu).

2 Proposal Specifics

Please limit your proposal to three pages. Every proposal will be different, but here are some issues that it would be good to address in separate sections. It's also probably best to begin with a short background description for general context.

2.1 Project Goals

Give a description of the goals for the summer project. If possible, this should come in three parts:

- An entirely reachable goal that you fully expect the students to achieve: this could be an answer to a question, an exploration of a hypothesis, things of that nature. It'd be best to give time-based specifics here: for example, by week 3 they should have learned X and produced Y, by week 6 they should extend this to Z, and so forth.
- A tangible product the students will create in the course of their research, which ideally will be of use both to further researchers at the university and to the students as something they can show off to future employers or graduate schools. This could be, for example, a good piece of well-commented software, or a visualization device, or a detailed curation of previously raw data.
- A more outrageous goal that you would be quite (pleasantly!) surprised to see the students achieve, along with a plan for them to build a potential roadmap towards that goal. For example, this goal might only be reachable if you had data that you currently do not have, and the students might build a speculative roadmap towards acquiring that data (write a mock grant, a mock IRB application, and so forth).

2.2 Dataset Description

Please give a brief description of the dataset(s) you wish to have the undergraduate students work with. You should make sure to cover the following points:

- What the data are about: a high-level description
- What's actually in the dataset: a lower-level description
- Access issues: what is the plan to ensure that the students will be able to access the data before the beginning of Data+?
- Privacy/Ownership issues: are there data sensitivity issues at play? Is an IRB protocol needed, and, if so, what is the plan to obtain one? If the dataset is owned by an outside party, how will a data use agreement (DUA) be negotiated?

2.3 Partner Description

Some of the best Data+ projects have a partner from outside of the university, or at least from outside the traditionally academic parts of the university. This might be someone who is invested in the data or the questions, and to whom the students will in essence "deliver a product." Ideally, this partner will be able to come two or three times during the summer to hear updates from your students and provide feedback. If this paradigm makes sense in your project, please give:

- a short description of the partner
- a short description of their interest in the problem
- an estimate about how much funding the partner might be able to contribute
- a plan for student-partner engagement:
  - Name and title of main point-of-contact
  - how often and in what context they might meet with the students

2.4 Day-to-day mentoring

Day-to-day faculty involvement in Data+ is not expected. Instead, each Data+ project has a mentor, usually a graduate student or post-doc, who is on hand to give the student team more focused guidance. The time commitment tends to be 5-7 hours per week, and funding is generally available to cover this person's time.

- If you would like to involve a student or post-doc from your own group, please give this person's name and contact information.
- If you would like us to provide a mentor, please list skills you would like this person to have.

2.5 Software Needs

Please describe what types of software the students will need to complete the summer project. Bear in mind that much of their work will take place on their laptops, but that there are of course many remote login options.

2.6 Skills Needed

In order to help our recruiting efforts, please list skills that students will need to make reasonable progress on this project. You may want to divide these up into essential and desirable. Bear in mind that you have $2-3$ students working together in your group, and your group will itself be in a working environment with around 15 other groups, and so they will be motivated and able to learn skills from each other.