

11 Jan 2019

SIParCS Program at NCAR

To whom it may concern,

The robustness of a data management plan determines its utility. Just imagine designing a metadata scheme that fails to account for partial information, or setting up a database under the assumption that future users will only want a single set of hard-coded queries. My goal this summer, then, is to lead the project “Building a Historical Data Image Archive to Support Climate Research...” by designing and prototyping a robust repository for historical climate documents.

1. In terms of technical sophistication, my basic computing toolkit includes version control (git, duplicity), high level (R, Python, Haskell) and low level (C++) programming languages, and markup languages (ipynb, Rmd, XML, HTML, pandoc markdown, \LaTeX). I am accustomed to both Unix-like (MacOS, Ubuntu/Debian) and Windows operating systems.
2. From exposure to data management in healthcare, in personal records, and in open source development, I am motivated by two heuristics: to create *long content* with *low noise-to-signal* ratio. The utility of pre-1960 climate records is apparent—my project specific interest is rather in formatting and deploying a repository of such records. By which gateways will the data be accessed? Should we create additional interfaces for non-specialists? How can historical climate records be visualized? How can crowd-sourced data entry be verified? I look forward to collaborating with peers in SIParCS and faculty at NCAR to answer these questions.
3. Statistical and computational experience would support my career goal to do topological data analysis. There are limited opportunities in my department to train with sophisticated computational tools. I believe that SIParCS would complement my theoretical strengths, which are afforded to me by my enrollment in a *pure* mathematics department.
4. In terms of unique contributions to the program: After my undergrad, I took two years to perform stipended service work. For a year in Houston, I developed scalable resources for refugee case management, including a crowd-sourced map of clinics and languages spoken. I wrote bug reports for the implementation three SQL databases, and, when Texas cut funding for Refugee Medical Assistance, I contributed to a data management plan for refugees transitioning from state to federal medical care. For a year in Olympia, I served as a community organizer at a 24/7 homeless shelter. I relied on distributed version control, and became a staunch advocate for deploying “early, often, and with redundant backups”. In all, this background experience allows me to contribute to a inclusive research environment. I strive (i) to collaborate, e.g., to focus my effort on tasks where I have a comparative advantage, (ii) to make incremental contributions on the work of others, and (iii) to be transparent, so that others may work off of my contributions.

Thank you for your consideration,



Colton Grainger