Community Data: Designing workshops to support ethical data use

PIT-UN 2023
Final Report

Principal Investigators
Dr. Betsy DiSalvo
Interim Chair, School of Interactive Computing, Georgia Institute of Technology
bdisalvo@cc.gatech.edu

Dr. Carl DiSalvo
Faculty, School of Interactive Computing, Georgia Institute of Technology
cdisalvo@gatech.edu

In collaboration with the **DataWorks** program: http://dataworkforce.gatech.edu

Project Summary: *Provide a clear and concise statement summarizing the work your institution(s) completed during the duration of the grant period.*

During the grant period, we collaborated with novice data workers from the DataWorks program and multiple community-based organizations to prototype a process for responsible data work. This process is significant and distinctive because it is worker-centered: it situates data workers as essential contributors to public interest technology through data collection and cleaning.

Progress Towards Objectives: *Describe the specific objectives of the project supported by the Challenge, the activities you engaged in during the grant period to accomplish these objectives, and any progress towards the outcomes or impact that you were hoping to achieve. Where you were able to accomplish your objectives, please describe what you saw as the key drivers or enabling conditions of that success. If applicable, please share a specific instance or event that illustrates the impact of your project.*

The PIT-UN Challenge grant supported and amplified community engagement between DataWorks and local community-based organizations. More specifically, the grant provided funding that enabled community-based organizations to work with DataWorks free of charge, and by doing so, increase their capacities to use data and to do so in responsible ways.

We accomplished our objectives of working with multiple community-based organizations to facilitate and amplify their use of data, and through that, create processes for responsible data use. In doing so, we realized that the processes were more beneficial and impactful for the data workers than the community-based organizations. This is a change from our proposal, which focused on creating tools for community-based organizations. This change arose in response to the working conditions for the community-based organizations we

partnered with. They relied upon volunteers whose availability and commitment understandably fluctuated, and their data practices and capacities varied widely, even within an organization. Introducing standards and procedures into their already hectic workflow would have been unsuccessful and unreasonable. However, preparing and enabling the data workers to lead in responsible data use and proactively shape the relationship with the community-based organizations as clients requesting data work was successful and fair. In doing so, the workers' data literacy and leadership capacities increased, and collectively, we modeled a data work environment that heightens the status of the workers by developing and acknowledging their broad expertise. We believe this approach of situating basic data work such as data collection, cleaning, and formatting as a primary site of responsible data use, and data workers as actors with the capacity to establish and sustain more responsible data use is an innovative and compelling approach to developing a worker to do the labor needed to sustain the vision of public interest technology.

Realizing that processes were more beneficial and impactful for the data workers than the community-based organizations, we created a workplace curriculum for responsible data use. At the time of writing this report, we have gone through the curriculum once with an initial set of four (4) data workers. Based upon that, we have iterated the curriculum design and are currently going through the revised curriculum with a new set of four (4) data workers. Notably, two (2) of the workers who went through the first iteration of the curriculum are serving as mentors to the new workers. This is significant because in this process, the workers become educators of other workers regarding responsible data use. The curriculum is oriented around machine learning, the role of data in machine learning, and even more specifically, how data collection and cleaning effects machine learning and can be a site of responsible computing.

We have attached a copy of the revised curriculum. We are securing funds from the National Science Foundation and internal funds to hire a teacher and a designer over the Summer to iterate the curriculum again. In this next iteration, the teacher and designer will bring their expertise in instruction, curriculum, and visual design to solidify the curriculum, making it accessible to a broad audience. Once completed, we will share that material on our website.

A key driver of our success was centering worker perspectives. This was difficult, and we certainly cannot claim that we achieved this in all cases; many times, we fell short or only realized after the fact how our decisions were driven by research or our perspectives as academics. But when we made decisions with the workers, led by their interpretations and assertions, we could sustain critical and responsible engagements with data that model a bold and robust democratic work environment.

The one objective we have yet to accomplish is dissemination. Unfortunately, our website does not have the project description or materials. This is because we are overhauling our website this Summer with the support of the National Science Foundation Research Experience for Teachers program and internal funding. We will post the project description and materials to our new website in August or September of 2023. Although this is a delay,

the new website will increase the usefulness of these materials and make possible even broader dissemination.

Challenges or Lessons Learned: *Highlight any challenges, expected or otherwise, or lessons learned throughout the grant period. Describe any adjustments or changes you made to your activities to address challenges as they arose.*

Despite having worked with many local community-based organizations for over a decade, we were significantly challenged by the capacities of those organizations and a misalignment between our expectations and what was reasonable for those organizations. As we started this project, many organizations were emerging from a massive contraction prompted by the COVID-19 pandemic: many organizations to lessen their staff and their programming. Although some of these organizations were in the process of increasing their staff and programming, they still needed more resources. Not recognizing this lack of resources was our oversight. There was a desire to collaborate, but the capacity still required to be in place to actually support collaborations. The limited capacity of the community-based organizations prompted us to pivot, and in the process, we began to focus more on the roles and activities of the data workers. This turned into an opportunity that shaped the project.

An important lesson learned was the extent to which the data workers could become partners with community-based organizations. DataWorks is founded on the idea that context matters in data science. When data workers and community-based organizations collaborate closely, novel insights from the data emerge, and the collaboration process sparks mutual learning. We witnessed that happen through work supported by the PIT-UN Challenge grant. For instance, we observed workers and Housing Justice League (a local housing activist organization) members collaborating in creating data categories, working together in an equal back-and-forth on how to best structure a data set and processes for cleaning data. By enabling data workers to make informed decisions about responsible data use, coupled with the agency and authority to lead data projects (rather than asking them to just unreflecting process data), we were able to create a demonstration of a robust and bold civic collaboration.

Equity, Diversity, and Inclusion: *Describe how your project meaningfully addressed the barriers to equity and access related to Public Interest Technology that you identified in your original grant proposal. If your project was not able to address said barriers as meaningfully as intended, please describe what challenges you experienced or lessons you learned.*

From the outset, DataWorks is a program to broaden participation in computing for minoritized communities in computing. All the data works are Black, and half of them are women. In addition, the community-based organizations we partnered with through the PIT-UN grant are also led by and primarily serve systematically oppressed communities, including communities of color and people in poverty. Through the PIT-UN grant, we were able to directly provide meaningful material resources to both the data workers and the community-based organizations.

Network Impact: *Describe how your project created shared resources or otherwise strengthened the community of educational institutions committed to Public Interest Technology.*

Once we complete production on our curriculum, our project will provide materials that can be adapted and used by other members of the Public Interest Technology community. We hope these materials will be helpful to both academics and community-based organizations as they partner with data workers to do the basics of data collection and cleaning needed for public interest projects. We also hope the materials are valuable to data workers directly. While few organizations or companies offer ethical or responsible data work, if such offerings grow, these materials can support those organizations and companies to develop fair, informed, and proactive work practices and environments.

To help achieve this impact, we welcome the opportunity to share our project broadly among the Public Interest Technology community. Please do reach out to us when those opportunities arise. We would happily present this work in larger events or small groups, in-person or virtually.

Institutionalization of Public Interest Technology: *Describe how, if at all, your project will contribute to Public Interest Technology becoming institutionalized within your university (i.e., through committed university funds, support from leadership, or collaboration between departments, faculty, or other groups). If you do not anticipate Public Interest Technology will become institutionalized in the short- or long-term, please explain why.*

Georgia Tech is committed to Public Interest Technology, particularly within the College of Computing. From the start, DataWorks has been funded, in part, by the Constellations Center for Equity in Computing at Georgia Tech. The PIT-UN grant enabled us to bolster that commitment by supporting a series of demonstration projects. These demonstration projects emphasize the role of data workers as decision-makers in responsible computing. We would not have been able to either create the materials or work with several of our community partners without the fund provided by the PIT-UN grant. The PIT-UN grant has been instrumental in enabling us to grow and broaden our DataWorks program and benefit community-based organizations. Through this, we will contribute to the vision and work of Public Interest Technology at Georgia Tech.

List Of Attached Files

DataWorks Responsible Computing v2.pdf

Financial report detailing final accounting of budgeted vs. actual expenditures of all grant funding, including the entire project budget and all sources of revenue and expenditures (including grassroots and direct lobbying expenditures, if applicable), in addition to this Grant.