

DATA JUSTICE AND THE RISKS OF DATA SHARING

MARIKA PFEFFERKORN

Many data scientists, activists, and others have begun to raise the alarm that big data, algorithms, and a lack of transparency in AI development will compound and exacerbate racial inequality. Often these alarms are accompanied by calls for community involvement. But it is crucial that community engagement is not just an add-on or window dressing for programs long in the making. That is exactly what happened in Minnesota's Saint Paul Public Schools. An unexpected twist in what was supposed to be a community-engaged process to improve services for "at risk" students resulted in a proposal for data sharing with the potential for discrimination and worse. Twin Cities Innovation Alliance (TCIA) brings together people, institutions, organizations, and communities to generate an educate-engage-equip model to activate community members more broadly on big data, predictive analytics and algorithms, and the engineered complexity of data-centric technology. Two of our biggest priorities are building partnerships to promote data literacy and agency and facilitating community action against harmful practices like data entrapment. Here is how it worked in Saint Paul, and what happens next.

A DATA SHARING STORY

At first, it seemed like a great idea. In the winter of February 2015, the Ramsey County prosecutor's office pulled together Saint Paul Schools and the city of Saint Paul to discuss how agencies could better coordinate resources. They then contracted with community hosts to hold engagement sessions on how to more efficiently and effectively deliver services. The response from Ramsey County residents participating in the year-long community engagement sessions was clear; they wanted to implement restorative practices rather than punitive models; they wanted reduced dependence on school resource officers and police intervention, and they wanted systems to proactively engage in solutions with the community. But when the official community engagement process report was released, many who participated were shocked. The report focused on using technology and data sharing between schools, counties, and municipal entities. Many community members had never heard a word about data sharing or other technological interventions raised in the community engagement sessions. Instead of getting a report that summarized community ideas and contributions to re-thinking school discipline, they got a plan for data sharing that seemed completely disconnected from the topics they all had discussed during the engagement sessions.

Soon after, Ramsey County, the city of Saint Paul, Saint Paul Police, and Saint Paul Public Schools announced their plans for a joint powers agreement (JPA) to begin a data sharing process. Their stated goal was to improve communication between schools, juvenile justice, prosecutors, public health, and child protection agencies through data sharing. The agreement included use of artificial intelligence and predictive analytics to identify students "at risk."

Community members were skeptical that the JPA would have benevolent impacts. While they acknowledged the need for better coordination between social welfare agencies and schools, having predictive analytics and law enforcement agencies in the mix was disturbing given the long history of racism in the criminal justice system. Our concerns increased when we got our hands on a copy of the JPA. It was full of technical jargon, lacked clarity about who was responsible for student data, and left many questions unanswered about how data would be used ethically to drive predictions about "risk." We also were alarmed at how quickly the county was moving toward this technical fix without consulting the community about the proposed data sharing practices. It felt like a bait and switch.

A COMMUNITY SUMMIT

TCIA gathered with other concerned community partners to organize. In contrast to the city and county's opaque decision making and lack of community inclusion around data practices, we drew on a host of authentic relationships and community-centered practices to engage folks in a deep dive into what the JPA was and how it could impact students and their families. The full story is outlined in the report, *Defeating the JPA: A Story of Community Empowerment through Education & Coalition Building*,¹ but here I want to take a moment to describe the culminating event of our process, the Cradle to Prison Algorithm Community Summit, which took place on November 10, 2018.

This summit met community members where they were in their understanding of the JPA. It featured experiential learning and included interactive, fun workshops like algorithmic improv and making a human algorithm poem. We hosted a tech talk featuring Yeshimabeit Milner, co-founder of Data for Black Lives; a "Dare to Data Clinic," and an activity called "One Mic" where parents could share their learning in bite-sized videos. All our workshops were centered around a restorative approach to ensure we did not perpetuate further harm, understanding that the legacy of systemic harm is real and long-standing surrounding the use and misuse of data in BIPOC communities. Examples of this are abundant in the Black community, from the infamous Tuskegee syphilis study to a more recent example in Pasco County, Florida, where the school district shared information about students with the sheriff's office without the knowledge of the students or their parents. The sheriff's office then used a computer algorithm to predict criminal behavior, ultimately using this "predictive policing" to label children as criminals "for crimes they have not committed and may never commit" as reported by the Institute for Justice.²

At the end of the summit, we debriefed with participants in restorative circles led by healing practitioners to unpack all they had learned, felt, seen, and heard over the course of the event. We gave ample time for this activity to ensure folks had an opportunity to process any tensions or discomfort brought up by discussions of historical and contemporary racial injustices. The outcome of the summit was a call to action: we would shift our focus from pausing the JPA process to dissolving the JPA altogether. And we did.

In the fall of 2018, in collaboration with In Equality and the Stop the Cradle to Prison Algorithm Coalition, we published *Improving Outcomes for*

Kids and Families: Beyond Predictive Analytics and Data Sharing, a policy brief we intended to use as a tool to better educate elected officials on the gaps and missteps embedded in the JPA. Below are the core messages from the brief.

- **Data sharing initiatives risk racially profiling children as “future criminals.”** Predictive analytical tools that draw from data influenced by systemic racial biases will continue to reinscribe inequalities and will not be accurate reflections of children’s individual strengths or challenges. For example, Minnesota is among the states with the largest racial disparities in suspensions and on-time graduation. Suspensions are correlated with race and law enforcement contact; therefore, BIPOC students would be seen as “higher risk” of becoming criminals.
- **Predicting behavior: “Risk” becomes “threat” when applied to children of color.** When preexisting racial biases that over-associate BIPOC folks with crime shape the data, then the “risk” score becomes a proxy for “threat to safety.”
- **Assigning risk scores, especially when there is lack of clarity about data chain of custody, will stigmatize children and families.** When children are flagged by the system for services, those scores are likely to leak throughout school communities, further exacerbating implicit and explicit racial biases.
- **Data sharing agreements may divert resources toward study and surveillance and away from services.** The JPA outlined an expensive, resource-intensive, and myopic study of individual children and family “weakness,” ignoring ways to address systemic injustice, bias, and harm.
- **By turning to big data to solve problems, local governments in this case obfuscated their own culpability in generating disparities and their responsibility to correct them.** We must stop jumping to making decisions via computer analysis rather than creating authentic, trusting human relationships.
- **Integrated data may be vulnerable to political agendas of those who want to criminalize segments of the community.**

Our coalition, activities, and report drew heavy media attention and public outcry against the JPA. With that pressure and a report of a data breach in the fall of 2018, the JPA was dissolved.

After the defeat of the JPA, our coalition celebrated with community, but we knew we had only scratched the surface. “What you have taken on here in St. Paul is 10 to 15 years ahead of the majority of places across the country and where you have succeeded provides a roadmap for others and should be replicated,” Yeshimabeit Milner, founder of Data for Black Lives, said. We catalogued our lessons. Many in the coalition returned to the primary focus of their advocacy, while the Twin Cities Innovation Alliance decided to go deeper into the work, launching the Data for Public Good campaign. The campaign’s goal is to educate and engage youth, parents, educators, administrators, superintendents, county officials, and elected officials on the lessons we have learned and to plot a new path forward—one centered on the public good as defined by the public.

WHY WE NEED DATA FOR PUBLIC GOOD

Data for Public Good (D4PG) is not an event but a milestone of a larger movement, a movement that defines shared leadership, vision, and responsibility for the good outcomes we want our data to drive. It is essential that data scientists and local, regional, national, and international governing bodies include the people who will be most impacted by big data and AI. Until these entities learn better practices of authentic community engagement, community organizers must remain vigilant. It is clear from our experience that governments are having difficulty keeping pace with technological change. Big data, new technologies, and new analytical approaches, if applied responsibly and in co-design with those most impacted, have tremendous potential to be used for the public good. But we need local, state, and federal agencies to work with communities to craft policies that establish a basis and expectation of trust that data and privacy is used for the common good, informed and determined by the people. D4PG uses a mix of research, networking, and public events to generate opportunities for community to be involved in data justice learning, activism, and policy shaping.

The Data for Public Good Campaign led a Community Participatory Co-Research National project in 2019 and 2020 with support from a multi-racial group of interns under the guidance of Dr. Catherine Squires and in cooperation with the Dignity in Schools Campaign and the Communities for Just Schools Fund, whose members and grantees represent more than 125 communities with whom we engaged across the United States, resulting

in the release of several reports and a toolkit at our national online conference in November 2020.

The toolkit provides information and resources regarding data-centric technology; student data sharing and privacy; infographics on districts; comparisons on district transparency about which agencies have access to their data; types of data collected; explanation of data shared; transparency about third parties; and clarity of the consent process. Also included are a case study with a comparison of the dos and don'ts of data sharing agreements for school districts; a frequently asked questions guide on FERPA and education technologies for families; PowerPoint presentations for sessions on student privacy; a data primer to break down jargon; a reader's guide for the Data for Public Good Book Club; and a compilation of videos and articles on data-centric technology, with current examples of immediate and secondary impacts, to share with communities.

In addition, we have created the No Data About Us Without Us Fellowship and community institutes. The fellowship is a six-to-nine-month cohort-based co-learning experience for parents, youth, educators, and community members. It is designed to build data literacy and data advocacy skills to empower the fellows to disrupt the ways big data, predictive analytics, and engineered consent are currently weaponized against marginalized and BIPOC communities, especially in education. Sessions are designed to be interactive and experiential, and the fellowship is grounded in relationships. We meet fellows where they are at and move at the speed of their understanding and trust. At our No Data About Us community institutes, participants learn about and deepen their understanding of big data, predictive analytics, algorithms, engineered consent and other terms; understand the historic arc of BIPOC communities and the misuse of data; review existing policies and laws meant to protect us and the problems and gaps that have been created; and learn to do research on the use of data in local communities and school districts and the existence or absence of transparency. They are supported in creating site-based campaigns and projects. In 2020, TCIA partnered with the Education Partnership Coalition to pilot the fellowship across six Minnesota communities, including Red Wing, Northfield, Farmington, Saint Cloud, Saint Paul, and Minneapolis. Sessions were simultaneously translated into Spanish, and materials were provided in Spanish as well. No Data About Us Without Us community institutes are delivered in partnership with government

systems, nonprofits, and funders to ensure program recipients and grantees have agency over the use of their personal data.

WHAT'S NEXT

TCIA is now co-developing and piloting a national and state policy tracking, analysis, and collaboration tool with Civic Eagle. As we roll out this tool, we will continue to collect and use the Data for Public Good Campaign survey data to design interactive heat maps that will provide a bird's eye view on this emerging policy trend. TCIA will work with local and national partners to introduce and track a constellation of policies that subsequently protect privacy, safeguard data, and ensure community trust.

All this is what authentic community engagement and data justice looks like. Our community-centered work is the opposite of what many governmental bodies, nonprofit agencies, private companies, and technical assistance providers put forth as “community engagement.” As the story of the JPA demonstrates, if community partners are not involved when technological solutions are brought into the mix, just data practices will not result. Data fixes generated by systems built on injustice will most likely replicate those injustices. Communities disproportionately injured by bad data practices need to be at the center of discussions and designing any use of technology that purports to address those injuries. We insist on authentic engagements and conversations between communities and data scientists, tech vendors, foundations, and government agencies that want to apply technology to solve inequality. When we say, “No data about us without us,” we mean it. Our well-being depends on it.

NOTES

1. Twin Cities Innovation Alliance, *Defeating the JPA: A Story of Community Empowerment through Education & Coalition Building*, 2020, www.tciamn.org/cpa-journey.
2. See Institute for Justice website, <https://ij.org/case/pasco-predictive-policing/>.