See Yourself in Data: Building a framework for data-based community engagement events

by

Jay Dev

B.A. Economics and African and Afro-American Studies
Brandeis University, 2013

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Master in City Planning
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Author_____
Department of Urban Studies and Planning
May 22, 2019

Certified by_____
Professor Ceasar McDowell
Department of Urban Studies and Planning
Thesis Supervisor

Accepted by_____
Professor of the Practice, Ceasar McDowell
Co-Chair, MCP Committee
Department of Urban Studies and Planning

See Yourself in Data: Building a framework for data-based community engagement events

by

Jay Dev

Submitted to the Department of Urban Studies and Planning on May 22, 2019 in Partial Fulfillment of the Requirements for the Degree of Master in City Planning

Abstract

Over the past decade or so, government data has been released through open data portals to improve efficiency, enable data-driven policy research and decision-making, increase transparency, and open a new avenue by which citizens may engage with the public sector. While open data has been a boon for researchers, journalists, technologists, and entrepreneurs, benefits from their publication have not necessarily flowed down to community organizations and residents. As unequal access to open data threatens to widen information gaps, models of citizen participation in the data-driven city have not fully developed. This thesis reviews possibilities and barriers of several forms of data-based participation, focusing particularly on participatory data interpretation as a liberating process and its pre-requisites of data awareness and literacy. It synthesizes a general framework for community-based data events, based on insights from Public Participatory GIS, Data Feminism, Data Activism, and Data and Digital Justice, and compares that framework to open data awareness and literacy-raising events in Pittsburgh and Los Angeles. Compared to the choices and achievements of these two cases, the framework holds as a guide for meaningful considerations that future community-based data events may take into account.

Thesis Supervisor: Ceasar McDowell
Title: Professor of the Practice of Community Development

Acknowledgements

Table of Contents

Chapter 1. Introduction to Open Data-Based Governance

In the past decade or so, governments at all levels have opened up their servers. As they have embraced greater data-driven planning and policymaking, they have also pushed to unify and publish data on their residents, services, infrastructure, and environment. Publication of these data is meant to improve transparency and accountability of government functions and policies, as well as build new pathways for denizens to engage with their government. Within this *open data movement*, actors ranging from government offices, technology companies, and watchdog organizations have worked to make government data available and useable to those interested in accessing it. At the municipal-level, local data intermediaries—including local non-profits and foundations, university research centers, and some government agencies—within the National Neighborhood Indicators Partnership (NNIP) network are translating and disseminating data to local stakeholders.

While the open data movement in general has attributed openness with participation, the definition of participation in this context is often quite vague. Many of the same stakeholders who were previously signing data sharing agreements and filing Freedom of Information Act (FOIA) requests—including researchers, journalists, and entrepreneurs—have benefitted greatly from more easily accessible and machine-readable formats of data. Yet, in many cases, open data has not yielded increased engagement by the average community resident. Beyond simply making civic data available through online portals, data publishers have had to develop engagement

techniques to improve awareness, literacy, and accessibility to a broad audience. This has included forming strategic partnerships and starting data user groups, creating educational material on accessing and using data, and hosting in-person trainings and workshops to build interest.

This thesis focuses on two examples of local data intermediaries within the NNIP network hosting series of community-based events to raise awareness and accessibility of open data among city residents using significantly different models of engagement. In particular, it suggests a general framework of data-based engagement and compares that framework to efforts in Pittsburgh (the Western Pennsylvania Regional Data Center and Carnegie Library's Data Days) and Los Angeles (the Sol Price Center for Social Innovation's Neighborhood Data for Social Change community workshops and data stories) to draw out key lessons for other organizations interested in putting on their own community-based data workshop.

To do this, it is critical to begin with an understanding of the open data movement broadly and the role of NNIP partners within it, which is presented below. The second chapter defines different types of participation within the data-driven cities, as models of participation have been largely underdeveloped within the open data movement, and establishes the clear need for awareness- and literacy-raising events. The third chapter surveys a range of other fields related to participatory data use to synthesize their tenants into a general framework for data-based community engagement. The fourth and fifth chapters review the efforts conducted in Pittsburgh and Los Angeles, respectively, and compare each of those initiatives to the framework proposed above.

The final chapter provides a broader analysis of the strategies and limitations identified in each of the two case studies and suggests pathways to address the more difficult aspects of the framework.

The Rise of Data in Policymaking and Planning

We currently stand at a moment where data and technology have promised to re-make the city: according to some 'from the internet up'[1]; to others, into the form of a "digital poorhouse."[2] The aspiration for technology-enabled utopias not new, but emergent technologies promise a greater scale of data-driven policymaking and planning. Historically, data has come from population censuses, case management systems recording the administration of public services, and the observation of landscapes and markets, and has been used to understand and govern the city through abstraction. This has shifted in newer visions of *smart cities*, which refers to the connection of existing infrastructure through communication networks, like the Internet, but has come to refer to much of the recent technological innovation in cities.[3] The current conversation around smart cities has been predicated on the idea that collection of larger data sets from more everyday objects and phenomena will create abstractions that more closely resemble reality. Outside of these promises, however, we mostly see data and technology used in three broad ways—monitoring and communication, prediction and modeling, and the creation of nascent connected systems.

---

[1] Doctoroff, "Sidewalk Labs | Reimagining Cities from the Internet Up."
[2] Eubanks, *Automating Inequality*.
[3] Townsend, *Smart Cities*.

One of the most prevalent applications of data in cities thus far have been in the creation of mobile and web applications and visualizations to monitor municipal processes and communicate that information to a broader public. Many cities have created data dashboards (and accompanying 'city health' indices) to curate and visualize growing streams of information.[4] In theory, these dashboards are meant to simplify the decision-making process by providing critical information through the lens of pre-defined metrics: are calls to the Public Works Department up or down this week? Are city welfare programs on track to achieve their outreach targets? What is the trend of monthly homicide rates, compared to a 10-year seasonally-adjusted average? Data are often paired together to help public officials "diagnose" the underlying ailments of the city. Open data has allowed such selective visualization to be strategically mirrored by private citizens and companies in the form of civic applications. For instance, the standardization of public transportation data into Google's General Transit Feed Specification (GTFS) has allowed application developers to create a range of dynamic, multi-city trip planning applications. Such applications have simplified timetables for a broader audience, synchronized modes of travel to enable easy multi-modal routes, and made real-time service updates and estimated time of arrival calculations possible.

The emerging ubiquity of machine learning techniques has also created opportunities for cities, as well as the organizations and companies who contract for them, to use their data for modeling and prediction. These applications have largely been done in the name of increasing efficiency of city budgets and resources. The

---

[4] Mattern, "Mission Control."

largest category of such applications has been to develop insights for public safety. For instance, the New Orleans city analytics team worked with Enigma, a data analytics startup, to predict the likelihood that homes in the city had smoke detectors using publicly available data.[5] In Atlanta, the fire department worked with volunteer data scientists to identify uninspected commercial buildings that were likely fire hazards.[6] Policing has long used predictive models, beginning with CompStat in New York in 1994, to strategically dispatch officers to places that are targeted as "crime hot-spots" (one current example is Los Angeles's PredPol).[7] And more recently, technology giants such as Amazon Web Services have offered their facial recognition software to police departments to identify suspects (though research has drawn into question the efficacy of such services to accurately match faces from those with darker skin tones and from those other than cisgender males)[8,9]. In a few cases, machine learning algorithms have been mobilized for other purposes unrelated to public safety, such as the prediction of homelessness,[10] yet here too it is done in the name of improving deployment (and thereby efficiency) of a limited set of city employees.

The third, and at this point most speculative, utilization of data and technology in cities is to create connected systems. Generally, this has been done through expansion of the Internet of Things—that is, the connection of objects and infrastructures to

---

[5] Flowers, "Looking for Smoke Alarms in New Orleans."
[6] Jay, "Can Algorithms Predict House Fires?"
[7] Goldsmith, "Predictive Tools for Public Safety."
[8] Buolamwini and Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification."
[9] Raji and Buolamwini, "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products."
[10] Johns et al., "Predicting Homelessness with Individual, Building, and Neighborhood Characteristics."

communication networks so that they are able to transmit data to inform other nodes in

the system or to a central repository. In some cases, smart city devices have been

deployed to alert officials of problems in systems, such as installation of water pressure

sensors across the distributed pipes network in Houston to flag leaks and bursts. Other

smart technologies have primarily served to passively collect large quantities of data for

opaque uses, such as user information collected by Google through LinkNYC internet

kiosks[11] or Intel's smart streetlights, which are "jam-packed with technology" to collect

HDR camera footage, GPS coordinates, sound, temperature and air pollution, and Wi-Fi

and Bluetooth connections.[12] Compared to traditional data collection methods, smart

city sensors promise to capture large amounts of data to benefit cities and citizens, but

also the technology companies that are uniquely prepared to sell and set them up at

scale.

The frontier of data and technology-driven governance has pushed cities around

the country (and the world) to set up new analytics departments and innovation teams,

but has also proven to be quite profitable for businesses. A number of cities have hired

outside consultants, particularly from technology companies such as IBM, Google,

Microsoft, and Cisco, to advise and build smart city infrastructure to collect data and set

up computational data processing systems.[13] IBM alone boasts eleven different fields

that it has provided cities insight into;[14] other companies like Cisco and Microsoft have

---

[11] Green, *The Smart Enough City*.
[12] Evans, "How Cloud Computing Is Evolving Street Lighting."
[13] Goh, "Who's Smart? Whose City? The Sociopolitics of Urban Intelligence," 2015.
[14] IBM, "Smart Cities Overview," accessed December 3, 2018,
https://www.ibm.com/smarterplanet/us/en/smarter_cities/overview/.

built significant subsidiaries (Smart+Connected Communities and Microsoft Research Labs) meant to advance their technologies into city governments. We are even seeing cases where cities have ceded significant control over land, governance, and data to technology giants, such as Alphabet's Sidewalk Labs rein over Quayside, Toronto. Some of the money funding these efforts are being swallowed by technology companies (they see more value in collecting and selling data and consulting services than simply charging for the installation of infrastructure)[15], while many private foundations (including Bloomberg Philanthropies, which kickstarted at least 20 innovation teams in cities around the world)[16] and public agencies (such as the US Department of Transportation's Smart City Challenge and the National Science Foundation's Smart and Connected Communities awards) have invested heavily in growing the scope of smart city technology.

Private profit models aside, a key question is why cities are learning more and more on data. On one hand, the ability to store and use larger troves of data has unlocked the promise of making "smarter" decisions based on more information. This includes analyses which take information across longer time horizons, as the data archive is growing by the day. These analyses can take into account a greater number and range of factors— both because data collection methods are able to track a greater number of variables, and because machine learning techniques have quantified large amounts of data that was previously thought to be qualitative in nature (such as photos through image analysis and text through natural language processing or sentiment

---

[15] Green, *The Smart Enough City*.
[16] Poon, "Giving Cities Room to Experiment and Innovate."

analysis). To process these longer and wider data sets, open source statistical analysis packages in languages such as R and Python have enabled more users to perform statistically complex methods to quantify, combine, automate, and simulate processes than ever before.

Technological advancements have also lowered costs for storage and processing, making data analysis within city governments more feasible. The cost of data storage hardware has dropped significantly, and maintenance of servers can now be outsourced thanks to cloud storage. At the same time, the computationally intensive processes described in the previous paragraph have become technically attainable due to the increased efficiency and ubiquity of computational processing power (and, in the case of particularly computationally-intense procedures, the advent of cloud computing). While these may contribute to lower computational infrastructure bills, governments may only be able to capture some of the efficiency surplus within the realm of open data, as they deal with unlocking data from legacy systems, hiring new employees to clean and standardize the data, and contract to external services to help publish the data.[17]

Beyond feasibility, the movement towards technocratic data-based governance and planning fits into a longer trajectory of these fields towards disciplines of science. Beginning in the late-1950s and early 1960s, there was optimism that cybernetics, a field developed by MIT mathematician Norbert Wiener, and its systems-analysis approach could be adapted from the field of security intelligence and applied to the problems of cities (first, in the name of defense, but later, as a means to better

---

[17] Newcombe, "The (Hidden) Cost of Open Data."

understand the control of a full range of processes associated with metropolitan areas).[18] In aspiring towards a science, planning has long strived to be methodologically rigorous. At the same time, such an approach would suggest that techniques could be more easily scaled from city to city. Utilizing data would also provide a broader view of social issues than human beings alone could understand, and the perceived neutrality of data would push processes away from human bias. However, many of these early cybernetics experiments yielded disappointing results.[19]

On the other hand, there is clear evidence and a growing group of scholars and activists who are pointing out the shortcomings of relying on data and algorithms in governance. Data can be biased or inadequate due to a number of factors. Surveys or instruments that collect self-reported responses can capture unintended information if questions are worded in a confusing or misleading way. Data can also be biased in its cleaning; critical information may be removed or incorrectly combined in the process of creating data ordered from the perspective of its handler.[20] Data may also be incomplete or misrepresentative. Sometimes, data is more representative of its collection technique rather than an accurate sample of reality (a common example of this is the fact that 311 calls generally represent likelihood to complain, rather than a complete picture of demand for city services).[21,22] If data is untimely or unrepresentative of more granular subpopulations or areas (particularly vulnerable ones), it threatens to inspire

---

[18] Light, *From Warfare to Welfare*.
[19] Light.
[20] Rawson and Muñoz, "Against Cleaning."
[21] Kontokosta, Hong, and Korsberg, "Equity in 311 Reporting."
[22] Green, *The Smart Enough City*.

exclusionary or biased policy. And as data is removed from context, it can lose information fundamental to understanding personal networks—social relationships, local power dynamics, nuanced understandings of past benefits and traumas—which is critical to effective policymaking and implementation.[23]

Algorithms, too, have come under fire for the ways in which they may allow bias to leak in. In their very construction, machine learning algorithms require some level of human decision-making. In supervised learning, coders must decide how to approach analysis and evaluate different levels of reliability. Even in unsupervised learning, algorithms tend to detect patterns that are reflective of social contexts rather than apolitical insights.[24] This is in part because machine learning algorithms are generally trained based on pre-existing training data set; an algorithm's predictive power can be constrained by its propensity to replicate the trends seen in the training set. One of the most notorious examples of this is in predictive policing. These algorithms are fed data on prior arrests to predict where future arrests are more likely. However, arrests are not necessarily a good proxy for crime, particularly in more affluent and white communities. Such algorithms have contributed to the over-policing of lower-income and minority neighborhoods.[25]

In adopting these practices, cities are also at risk of taking on the underlying logic, assumptions, and goals of the fields they attempt to graft on—whether that has been science, defense, or business. As shown in the examples above, many of the

---

[23] boyd and Crawford, "CRITICAL QUESTIONS FOR BIG DATA."
[24] Eubanks, *Automating Inequality*.
[25] Green, *The Smart Enough City*.

applications of data analytics in governance have looked to improve efficiency of

government services or budgets, likely borrowing from the business intelligence sector

which smart city consulting companies also serve. In focusing on technical criteria, cities

are at risk of losing sight of more critical issues in the sphere such as fostering equity

and democracy.[26] Instead, in the process of capturing a complex world in data, context

must be flattened; this flattening is often done by those with social power and in such a

way as to reinforce that power, contributing to the perpetuation of oppression.[27] This is

compounded by the process of data interpretation; generally, the process of meaning-

making is not done by those who have their data collected, nor in the context of

collection. If cities contribute to the inequitable access to data and tools of analysis

among their denizens, they risk creating silos of power reinforced by inequitable.[28]

Data Democracy through the Open Data Movement

The open data movement has sought to alleviate some of the concerns of

unequal access to information through the open publication of key public data sets. The

concept of open data is rooted in the 2005 introduction of the Open Knowledge

Foundation's Open Definition. According to the current Open Definition, version 2.1,

knowledge is open if, "anyone is free to access, use, modify, and share it — subject, at

most, to measures that preserve provenance and openness."[29] This follows from the

---

[26] Green.
[27] D'Ignazio and Klein, *Data Feminism*.
[28] Gleeson and Dyer, "Manifesto for Collaborative Urbanism."
[29] "Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge."

concept of open-source software. More specifically, open data must be: 1) legally open, under an open license that allows access, reuse, and redistribution, 2) socially open, for collaboration, and 3) technologically open, in machine-readable and non-proprietary formats.[30]

The concept of open data took hold within United States governments several years later. Although governments in the 1990s and early 2000s had embraced data transparency, the Bush administration encouraged non-disclosure of data through the Ashcroft Memorandum, generally in the name of security.[31] On President Barack Obama's first full day in office, he issued the Memorandum on Transparency and Open Government, which reinstated disclosure to, "establish a system of transparency, public participation, and collaboration."[32] Shortly after, federal, state, and local governments began publishing data through "open data portals.[33]

The goals of open data outlined by the Obama White House have since been reinforced by a number of organizations. The Sunlight Foundation, a nonprofit organization devoted to advancing open data and civic technologies, outlines the range of claims often made by open data advocates; that open data can be used "as a tool to increase internal efficiency, to establish better data governance, to enable internal data-sharing, to keep government performance accountable to the public, or to enable better community data-sharing."[34] Beyond this, open data can leverage "civic hackers," or

---

[30] Shaw, "Making the Most of Open Data."
[31] McDermott, "Building Open Government."
[32] White House, "Memorandum on Transparency and Open Government."
[33] Shaw, "Making the Most of Open Data."
[34] Sunlight Foundation, "Open Data Policy Hub | Why Open Data?"

other volunteers with skills in programming, data analysis, and content-area expertise to develop applications to streamline or improve citizen interactions with government processes. Open data has helped drive forward the desire for more quantifiable public policy, or as written by former New York mayor Michael Bloomberg, "if you can't measure it, you can't manage it."[35]

Yet, as alluded to in this last point, concerns exist whether the open data movement is truly allowing universal access to information, or whether it is deepening the digital divide. An underlying assumption of the movement is that for citizens to make use of open data, they must also have access to the Internet and technology to read the data, the skills or data literacy to understand the data, and the organization and political power to make use of that data.[36] With those fundamental barriers, it is unsurprising that open data is not often used by those most oppressed by government; instead it is largely used by middle-class social movement leaders, researchers, non-profit staff, journalists, or civic technologists.[37] The Sunlight Foundation's original Open Data Guidelines, first published in 2012 and often cited as a standard for cities looking to draft their own open data policy, is missing any mention of how to support the use of public data once it is downloaded from a portal.[38] But in order to achieve its goal of raising civic participation, open data must be connected to democratizing digital infrastructure, capacity building efforts, and organizing capacity.

---

[35] Goldsmith and Crawford, *The Responsive City*.
[36] Gurstein, "Open Data."
[37] Meng, "The Social Impact of Open Government Data."
[38] Sunlight Foundation, "Open Data Policy Guidelines."

The National Neighborhood Indicators Partnership

Among the numerous actors within the space of publishing open government

data, local data intermediaries in the NNIP network have tried to serve as conduits

between local stakeholders and open data. The first local data intermediaries started in

the early 1990s in response to the gap in availability of neighborhood-level data. Up

until that point, data was difficult and costly to mobilize, even where it was available at

the city-level, so municipal policies rarely took data into account.[39] In 1992, the Center

on Urban Poverty and Social Change at Case Western University took on the role of

managing and publishing local-area data, and groups in Boston, Denver, Oakland,

Atlanta, and Providence, Rhode Island soon followed suit. In 1996, the six organizations

formed the NNIP network under the guidance of the Urban Institute, a nonprofit

research organization.

In the intervening years, the network has grown to over three dozen partners in

cities around the United States. Most partners are individual organizations, although

several are partnerships between organizations. Many of them are freestanding

nonprofits (or are located within freestanding nonprofits), but a quite a few, including the

two intermediaries covered in this thesis, are located within community-oriented

university departments or research centers (the Western Pennsylvania Regional Data

Center is located within the University of Pittsburgh's Center for Social and Urban

Research; the Sol Price Center for Social Innovation is located within USC's Price

School of Public Policy). Embedding within a larger institution can provide the

---

[39] Kingsley, Coulton, and Pettit, "Introduction to the Field."

advantage of program sustainability, both in terms of institution stability and possible

financial benefits.[40] At the same time, these partners may have to align their work with

the mission of the parent institution; in the case of university research centers, this

means their work often involves the participation of faculty and education of students.[41]

Although the network has grown alongside the ascension of the open data

movement, the core tenants of the network remain the same. Local data intermediaries

are expected to assemble, transform, and maintain data; disseminate information and

apply data to achieve impact; and use data to strengthen civic capacity and

governance. While governments at all levels have expanded their own open data

publication capacities, local data intermediaries remain a critical part of community

information systems. This is due, in part, to their commitment to distressed

neighborhoods and the outward-facing orientation of their work, both in providing

technical assistance for local stakeholders to access data and to focus their research on

topics of interest to communities (as opposed to the work that municipal analytics

departments are often mandated to do to improve internal capacity and efficiency within

the city hall).[42] This central tenant of local data intermediaries makes them well-

positioned to be considered as the hosts of participatory data efforts.

Chapters four and five describe community-based data engagement efforts run

by local data intermediaries in Pittsburgh and Los Angeles. Both series of efforts aim to

make open data published by the intermediaries more accessible, but take different

[40] Kingsley, Coulton, and Pettit, "Institutional Context."
[41] Hendey et al., "NNIP's Guide to Starting a Local Data Intermediary."
[42] Hendey et al.

approaches to do so. NNIP partners in other cities around the country have run similar events, including in Baltimore, Milwaukee, Denver, and Detroit. These two cases were chosen because the events were successfully run numerous times, were co-organized alongside community partners, and were conducted within the context of larger engagement efforts (in the case of Pittsburgh, that includes Data 101 trainings and office hours; in the case of Los Angeles, that includes data stories). Each organization focused on significantly different goals and approaches, which provides us with a broader spectrum of results to compare between. Other NNIP partners have implemented other types of activities within their events (for instance, Data You Can Use in Milwaukee have held a Data Dream competition to seed longer-term funded partnerships with community organizations that attend their annual events) and deserve to be studied in future endeavors.

Each case study presented in chapters four and five were informed by interviews and primary source material. In Pittsburgh, information was gathered from interviews with three of the central co-organizers and two community partners, while in Los Angeles, information was gathered from interviews with three of the central co-organizers and three community partners. In both cases, information published online and materials shared by the co-organizers was used to supplement insights from the interviews. While personal narratives were considered as a potential methodology (and a powerful one from the perspective of Data Feminism), perspectives from interviews have been combined to maintain privacy.

Chapter 2. Pathways to Citizen Participation and Data-Based Empowerment

The previous chapter highlighted the limitations and risks of proceeding with an unchecked data-based future. Data, without sufficient context, can mislead policymakers, recreate social inequalities, and fall short of answering critical questions about why social phenomena occur or how to address them. Instead, cities would be wise to consider opening up data collection, analysis, and mobilization to a broader set of stakeholders not just through open data publication, but also through more participatory processes.

As mentioned before, the open data movement has left notions of participation and engagement vague. This chapter attempts to define participation within a data-based governance framework. Participation is a key component of the democratic process and can provide clear benefits to data-driven policymaking and planning. There have been several attempts to theorize the role of citizens in smart cities; reviewing these models can prove instructive in identifying a spectrum of participation in the context of efforts that are already underway. Within this spectrum, the process of collective interpretation and meaning-making is analyzed as a particularly impactful form of broad-based empowerment. However, before city residents can be expected to do participate, there is a level of minimum awareness and literacy necessary (a gap which the cases in Pittsburgh and Los Angeles fill).

Benefits of Participation

Opening up these data-based processes not only fulfills ideals of democratic justice, it can also contribute to better policymaking. The utopian vision of a data- and technology-enabled, hyper-efficient state is one which matches the technocratic authoritarian state critiqued by Scott. Data-driven insights from within the government follow a quest for 'legibility' of its citizens as the authoritarian state seeks to fulfill its goal of social order. Yet data-driven governance is imperfect as it is "necessarily schematic" and requires a flattening of the real, complex nature of the world. In that simplification, the state loses an increasing level of its ability to generate and implement effective policy. Instead, Scott suggests, the well-functioning state remains cognizant of the role of "local knowledge and know-how."[43]

The practice of community-based participatory research is built on a similar underlying assumption of the power of local knowledge and accountability of the community to bolster data-based insights and can serve as an important example of the tangible benefits of a participatory process. Including community members and organizations at various points or throughout the process can help policymakers, researchers, or technologists better understand the nature of the issue at hand.[44] When local participants, or community researchers, are involved in the development of research questions, they can ensure that those questions are of relevance and interest to the community. In the data collection process, community members can help identify

---

[43] Scott, *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*.
[44] Minkler, "Enhancing Data Quality, Relevance, and Use through Community-Based Participatory Research."

and potentially gain access to key sources of information, ensure that survey instruments or interview questions are being asked in a culturally-competent way, and push back against data sources that may not be acceptable to use. In the data validation process, community members may be able to "ground-truth" existing data sets, particularly those accessed and published by the government. Community researchers likely have greater insight into the local culture and context reflected in the data, and can provide deeper interpretations than academic researchers.

Building connections with community members throughout a data-driven process can also produce better outcomes and smoother implementation. Community researchers can provide critical perspective through a local lens when co-designing data-driven interventions, and they can suggest ways for policies to not reinforce harm or distrust of government or researchers that already exists within the community. When the process is being documented and disseminated, community researchers can advise on the best formats and channels to reach key local stakeholders. The process of engaging a wide-range of stakeholders (from government officials, academic researchers, technologists, and community researchers) can build trust and help mobilize a broad base of supporters to carry-out the actions recommended by data insights. Finally, involving community researchers throughout the process can build consciousness of issues that residents may not be familiar with, raise awareness of existing laws and programs related to those issues, and build capacity to conduct future inquiries.

Models of Participation in Data-Driven Systems

At the current moment, citizens are already being engaged in data-driven cities. In reviewing various models of participation, the possibilities and constraints of future proposed efforts can be better anticipated. This section reviews four such models of participation, tied directly to efforts that have been implemented before in data-driven contexts, and compares them to similar categories within two frameworks of smart city citizenship (one by D'Ignazio and Zuckerman, the other by Cardullo and Kitchin). Each of these models present unique sets of powers and limitations.

At the lowest level of participation, citizens can be sources of data. Such data can be collected consciously—through surveys, interviews, or voluntary interactions with civic technologies or applications—or subconsciously—through sensors, or as a side-effect of public program or service utilization. Currently, data is being collected in these ways by the state and by private companies. Generally, this type of engagement can be typified as being a "non-participant" or a "consumer" in the smart cities context.[45] Neither of these roles provide much autonomy in decision-making power. Further along the spectrum of participation, data collection can take the form of citizens providing feedback to decisions made by those in power, for instance in attending public meetings on or electronically commenting on a proposed master plan or policy. However, without further organizing capacity, this information is at risk of facing tokenism by decision-makers.[46] Regardless of how it is dressed up, remaining a "data point" falls within the

---

[45] Cardullo and Kitchin, "Being a 'Citizen' in the Smart City."
[46] Cardullo and Kitchin.

traditional smart cities mold—where data is primarily in service of top-down paternalism.[47]

A level up from there is the model of citizens serving as data collection partners for cities. This can take the form of organized and voluntary collection of data (such as the Motor City Mapping project to map blight in Detroit)[48] or as decentralized citizen sensing projects (as is most often seen in the collection of environmental data).[49] This type of participation provides much greater levels of autonomy for citizens. In this form of partnership, city residents can ensure that their communities are being represented in data sets and can potentially help shape the way that data is being collected. Some have identified the potential for crowdsourcing data collection efforts, which have a relatively low barrier for entry and can scale to alleviate access constraints like time and geographic access, as a seed for community capacity building and bottom-up decision-making. In this vision, platforms are designed to recreate social networks and the most popular ideas are those that receive the most attention.[50] However, such aspirations may be too lofty to place on platforms themselves. Sensing projects, without requisite grounding in community or place, are in danger of extracting data from individuals without providing them access or ownership of that data.[51] For instance, when detailed

---

[47] D'Ignazio and Zuckerman, "Are We Citizen Scientists, Citizen Sensors, or Something Else Entirely? Popular Sensing and Citizenship for the Internet of Things."

[48] Wasacz, "How Motor City Mapping Became a Template for Citywide Collaboration."

[49] D'Ignazio and Zuckerman, "Are We Citizen Scientists, Citizen Sensors, or Something Else Entirely? Popular Sensing and Citizenship for the Internet of Things."

[50] Certomà and Rizzi, "Crowdsourcing Processes for Citizen-Driven Governance."

[51] Gabrys, "Programming Environments."

blight data were published on Detroit's Open Data Portal, residents had no way of preventing real estate speculators from accessing and profiting off of that information.

Another role that has been envisioned for citizens is that of advising open data organizations in their activities. That the grassroots level, citizens can be empowered to observe and organize around what they notice around them as citizen journalists or watchdogs. In a more top-down model, citizens can partner with city agencies to serve on open data advisory boards (which a number of cities including Detroit, Cambridge, Massachusetts, and Arlington, Virginia have instated). When imbued with a level of influence over decision-making power, citizens may be able to guide the process of publication and use of government data. Community interests on issues like research relevance and risk can be represented by delegated community members. At the same time, advisory boards with oversight power are still often constrained in their ability to influence day-to-day activities of those working with open data. This model of engagement is further constrained in imbalanced distribution of power and labor. In the top-down advisory board approach, only a few spots with considerable influence may open up for community members to engage with data-driven activities, and generally those spots are given to members of the community who are already deeply engaged with the data ecosystem.[52] As watchdogs, community members are required to be detectives and advocates for their power; both of which take considerable time and energy. Organizations like the Stop LAPD Spying Coalition have succeeded from such

---

[52] Cardullo and Kitchin, "Being a 'Citizen' in the Smart City."

a position, but their success is a testament to the deep engagement their members have had to the cause.

A fourth mode of participation is collective data interpretation, where community members are able to co-create data-based understandings of their lives to influence policymaking and planning. The process of meaning-making here is incredibly important, both as a instrumental exercise of aligning policy with local realities and as a liberating one (as Freire describes, one must declare their situation in reality in order to change it).[53] In most of the current applications of data-driven participation listed above, citizens are largely working within an existing framework of choices determined by those in power; in order deepen the political discourse, citizens must have the opportunity to reframe those choices to fit their perspectives of liberation.[54] This important method of participation is the one at the center of this thesis, and is worth unpacking further.

Striving for Collective Data Interpretation

While an underutilized form of participation compared to the other models reviewed above, community co-interpretation of data can bring lived experience in conversation with data insights. In addition to the benefits stated above, this can provide a more effective and equitable pathway to re-contextualizing data with social context, qualitative relationships, and power dynamics, which are generally lost in the data collection process. If it allows communities to speak directly to decisionmakers and the

---

[53] Freire, *Pedagogy of the Oppressed.*
[54] Cardullo and Kitchin, "Being a 'Citizen' in the Smart City."

processes they invent, participatory analysis can be one of the most powerful methods to reach a balance between efficacy and equitable democracy.[55] Healy (258) posits that the process of meaning-making within a social context is what creates that which is "known." In the context of policymaking and planning, this means that "expert 'planner'" knowledge and "lay knowledge" cannot be decoupled; instead their meanings and values grow together. It is therefore critical that policymaking processes remain open to capturing the full extent of both forms of knowledge.[56]

Yet, to be able to conduct collective data interpretation, communities must be aware and literate of the information available, and must have a platform with which they can express themselves. Building data literacy ensures that participants are able to read, work with, analyze, and develop arguments based on data as a part of the inquiry building process. This goes beyond technical trainings on using tools and numeracy; it requires that individuals are able to advocate for themselves using data.[57]

In many cases, the platform of creative expression not only helps speak to power in a nontraditional way, it can also unlock advocacy to those with creative skills. One format is data storytelling, which builds a narrative that is based around an insight from data. More generally, storytelling is a key cultural transmission process which embeds views of aspirations, actions, and motivations.[58] When paired with data, stories can capture a wide range of information, both observed and embodied. This can take the form of more traditional oral storytelling or, as has been conducted in several cases by

[55] Laird, "Participatory Analysis, Democracy, and Technological Decision Making."
[56] Healy, *Urban Complexity and Spatial Strategies*.
[57] D'Ignazio, "Creative Data Literacy."
[58] Little and Froggett, "Making Meaning in Muddy Waters."

Rahul Bhargava and Emily Bhargava, through the process of collectively finding a story in data and using it to build a data mural.[59]

More traditional visualizations are also a potential format for teaching data literacy. In part, visualizations are successful because they simplify the complexity of data into a visual form that can be easier to understand.[60] Yet the opposite is true too: data literacy is also critical to using data visualization as a civic planning tool. Visualization has been criticized as those in power have generally determined what data were presented, perpetuating top-down power structures. But as an interactive process to make open data legible, visualizations can be used as a powerful tool for exchanging knowledge, exploring and generating personal data stories, and seeding civic dialogue.[61,62] In such cases, it is not simply the resulting visualization that is of value; participants can benefit from the visualization creation process itself.

One illustrative example of the literacy building value of the data visualization process was conducted as a collaborative effort between the Center of Urban Pedagogy and a 10th grade English Language Arts class from the Academy of Urban Planning and Engineering in Brooklyn, New York. Students were led through the process of data-based argument, beginning with understanding the issue of New York City's new soda tax from both official sources and community voices. From there, the students were dispatched to collect quantitative and qualitative data from their neighborhoods. They then visualized the information using soda cans, sugar packets, and construction paper

---

[59] Bhargava et al., "Data Murals: Using the Arts to Build Data Literacy."
[60] Segel and Heer, "Narrative Visualization."
[61] Williams, *Data Action*.
[62] Williams, "Data Visualizations Break Down Knowledge Barriers in Public Engagement."

to identify and weigh different arguments. Finally, the students were able to write about their findings, which contributed to the ongoing debate, based on data and perspectives they collected. This effort not only built literacy skills (which came in handy as the students were learning to write persuasive essays based on infographics for upcoming standardized tests), but also taught literacy in the context of civic engagement.[63]

Without an awareness of the availability of open data and both the confidence and literacy to make use of it, communities are unable to lead collective data interpretation processes. At best, communities are then dependent on others to pre-process data, mapping outside assumptions onto it and potentially missing the opportunity to build a nuanced understanding of the data from the population in question. In addition, they would be limited to being invited to have a seat at the table, rather than being able to mobilize and act on their own. The following chapters look at how community-based data training workshops attempt to bridge the gap to reach these pre-conditions for collective data interpretation.

---

[63] Center for Urban Pedagogy, "Soda Census."

Chapter 3. Constructing a Framework for Community-Based Data Events

Instances of participatory data use have been initiated within the context of the open data movement, including the community-based data awareness and literacy events which this thesis focuses on, but similar activities and literature have been considered at-length in a range of academic and activist subdisciplines—including Public Participatory GIS (PPGIS), Data Feminism, Data Activism, and Data and Digital Justice. These fields have been used to describe a wide range of efforts, but they all generally attempt to re-contextualize and utilize data for the advancement of equity and democratic empowerment. They have evolved from many common predecessors—such as Critical Race and Feminist theories, Science and Technology Studies, Critical Geography and Data Studies, Critical Pedagogy and Pragmatism, and methodologies of Community-Based Participatory Research and Participatory Action Research—yet they have diverged in their approaches to study and action.

Just as the open data movement has only loosely defined the typologies and methods of participation, it has yet to synthesize a cohesive framework of participatory data use. This chapter seeks to develop one through a collective reading of the tenants and considerations of PPGIS, Data Feminism, Data Activism, and Data and Digital Justice, building off of the conclusion drawn in the previous chapters: 1) data provides both positive potential that must be distributed equitably and negative risks that must be managed proactively, and 2) various forms of knowledge provide greater potential when in dialogue with one another, rather ranked into a certain hierarchical order. The points

of this framework are based on common features of each field, although not every field

is reflected in each of the points of the framework. This framework is not meant to be

exhaustive. Instead, it attempts to connect several modern-day participatory data

movements to provide a basis with which to analyze the processes and goals of

community-based data events.

Four Analogous Disciplines

Before comparing these four disciplines, it is helpful to understand the context

within each originated and how they compare to one another. This section provides a

brief primer to each one, although a deeper reading of each is recommended.

Public Participatory GIS is the oldest of these four fields, but provides a useful

comparison to the current moment in critical data studies. PPGIS focuses on pragmatic

approaches to improve government transparency and broaden the public's involvement

in policymaking through the broadened use of GIS.[64] It originated in the field of

geography which, by the early-1990s, had grown into two competing factions of study

and practice: GIS, a field concerned with quantification and visual representation that

was borne out of the growing accessibility to desktop GIS software, and Critical

Geography, which surfaced questions on the impact of GIS technology on society. As a

critical geography research agenda began to form, one branch was interested in

building strategies for practical intervention. Many human geographers and GIS

---

[64] Sieber, "Public Participation Geographic Information Systems."

researchers began working together to put on workshops, starting in 1996, to introduce the budding field of 'Public Participation in GIS'. [65]

At its core, PPGIS was responding to the fact that GIS techniques were limiting the participation of the average citizen from policymaking processes. Critical geographers noticed that the maps produced by GIS held a level of persuasiveness that was difficult for community members or community groups to challenge.[66] Yet, rather than focusing on ontological theory, the field was made use of methods of participatory action research to include GIS researchers into the process of normative empowerment.

This does not leave PPGIS above reproach. PPGIS projects generally rely on data operators to interface with community organizations and community members, and therefore still hold power within the decision-making process.[67] At the same time, the PPGIS discipline has retained questions of the position of those operators, which can serve as an important guide to the local data intermediaries profiled in the subsequent chapters.[68]

On the other hand, Data Feminism, Data Activism, and Data and Digital Justice are more contemporary movements. They take slightly different approaches in response to smart city and open data movements: Data Feminism considers data from critique and analysis (predominantly from the perspective of feminism, but intersected with a range of perspectives, such as race, class, sexuality, ability, age, and geography, to name a few); Data Activism is based in the practice and theory behind grassroots

---

[65] Schuurman, "Trouble in the Heartland: GIS and Its Critics in the 1990s."
[66] Obermeyer, "The Evolution of Public Participation GIS."
[67] Brown and Kyttä, "Key Issues and Research Priorities for Public Participation GIS (PPGIS)."
[68] Williams, "Data Visualizations Break Down Knowledge Barriers in Public Engagement."

actions; Data and Digital Justice provide a rights-based framework of theory and practice.

Data Feminism has emerged as a unique field more recently, but feminist critique of science, technology, and data have existed for quite some time. Specifically, Data Feminism takes an approach to understand data and its communication that is informed by intersectional feminism. This means it is primarily focused not specifically on gender, but on power imbalances in data science caused by oppressive social structures. Therefore, it centers marginalized lives, experiences, and interpretations of data to strengthen objectivity and address injustice.[69] In practice, this involves surfacing and interrogating assumptions in what would otherwise be considered "apolitical" coding, analysis, and representation choices; as well as surfacing and interrogating the human beings who construct and are identified as consumers of data analytics and visualizations.[70]

Data activism has coalesced as a field over the past few years as a result of actions being taken in the name of data-driven contestation. Generally, the field is interested in studying actions, driven by both social and technological means, that are either affirmative engagement with data or resistance to mass data collection.[71] To do so, the field draws on notions from both Science and Technology Studies and Social Movement Studies. The field takes a purposefully twin-pronged definition in the name of agency to mobilize for or against datafication; as a result, the actions within its scope

---

[69] D'Ignazio and Klein, *Data Feminism*.
[70] D'Ignazio and Klein, "Feminist Data Visualization."
[71] Milan and Velden, "The Alternative Epistemologies of Data Activism."

are similarly broad to include both pro-active responses to data, including the use of

data for greater advocacy and improved pathways to engaging with the state,[72] as well

as re-active responses, such as actions taken to resist state and corporate monitoring.

In its study, the field often centers the work of grassroots activists (as their work is left

out of traditional fields of data-based study, instead focusing on the work of institutions).

However, critical work within the field has looked at the role of agents within the state[73]

and corporate fields[74] to improve institutions from within.

  Finally, Data and Digital Justice are related fields which each draw a connection

between the principles of human rights and social justice to the use of data or digital

technologies in governance and capitalism (though they cover slightly different

technologies, concepts from both data and digital justice are helpful in understanding

justice in the realm of open data). Taylor draws upon three approaches to frame data

justice within a rights-based framework: 1) the nature of data collection and use to

reproduce power asymmetries[75], 2) the power of responsibly-owned, accessed, and

represented data to make the global poor visible in order to increase distributive

justice[76], and 3) the resistance against data-based surveillance through principles of

social justice.[77] This leads to a framework of *(in)visibility* in terms of both access to

representation and information privacy; *(dis)engagement* in terms of both fair access to

technology, as well as autonomy of choice to participate in or not participate in data

---

[72] Schrock, "Civic Hacking as Data Activism and Advocacy."
[73] Horgan and Dourish, "Ambiguity, Ambivalence, and Activism."
[74] Costanza-Chock et al., "#MoreThanCode."
[75] Johnson, "From Open Data to Information Justice."
[76] Heeks and Renken, "Data Justice for Development."
[77] Dencik, Hintz, and Cable, "Towards Data Justice?"

markets; and *antidiscrimination* in terms of the ability to point out and challenge bias in data use, as well as the freedom to not be discriminated against.[78]

This framework provides broad pillars of a rights-based approach to data-governance and capitalism. A more applied interpretation from the related field of digital justice comes from the Detroit Digital Justice Coalition, an action-based organization which organizes and advocates for community interests in the city of Detroit. They coalesce under the belief that communication is a fundamental human right, and that right is secured through activities grounded in access, participation, common ownership, and healthy communities.[79] Among their work, they have collected and published a set of recommendations for the city's open data portal, based on community-based research and the organizations rights framework.[80] These digital justice principles are described in greater detail below, within the pillars of the framework that this thesis puts forth.

A Framework for Community-Based Data Events

The four fields presented above cover various aspects of theory and practice of connecting citizens to trends in greater collection and use of data in sociocultural, economic, and political society. Despite their different approaches to issues of the use and representation of data, they overlap in meaningful ways. The following framework is

---

[78] Taylor, "What Is Data Justice?"
[79] Nucera and Bianchi, "How to DiscoTech."
[80] Detroit Digital Justice Coalition and Detroit Community Technology Project, "Recommendations for Equitable Open Data."

built from pulling out core issues that arise most frequently and centrally across PPGIS,

Data Feminism, Data Activism, and Data and Digital Justice and reforming them to fit

the context of community-based data events. It contains five primary pillars, which are

explained below: **Defining Primary and Secondary Audiences**; **Addressing**

Physical, Technological, and Personal Barriers; **Mitigating Risk**; **Providing**

Opportunities for Meaning-Making; and **Creating Direct and Indirect Impact**.

Defining Primary and Secondary Audiences

> *Events shall take into account who is planning and participating in*
>
> *them (primary audience), as well as the audience for any resulting*
>
> *activities (secondary audience).*

The first pillar of the framework generally asks the question, *"Who?"* regarding

both the organizers, participants, and secondary audience. As these events are

organized, it is critical to be conscious of which communities are represented at the

table and which groups are missing. From previous studies in PPGIS, it is important to

consider those who are impacted or can have an impact on the topic at hand (in the

latter group, this could include those who bring unique knowledge or power to the

process).[81] Data Feminism and Data and Digital Justice generally go further in their

goals, striving to center the most marginalized or historically excluded populations in the

---

[81] Schlossberg and Shuford, "Delineating 'Public' and 'Participation' in PPGIS."

process of collecting, analyzing, and understanding data.[82,83] Not only is this an opportunity to empower those who are at greatest risk due to greater datafication, it also ensures that their perspective is not spoken for by others (or, as often described from the activist perspective, "Nothing about us, without us").[84]

Going further, this is an opportunity to contest the expected roles and stakeholders in technological and data-based spaces. We must go beyond the traditional depictions of the data or technology worker, both because they can perpetuate exclusionary perceptions of what constitutes "work" (e.g., as only what happens in private companies or start-ups; as the contributions of white cisgender men), but also because such stereotypes erase the work that is being done by others in the ecosystem which may not immediately be termed as "data work."[85] Yet, this points to a challenge in diversifying the space: how do we ensure that communities are included when they are disconnected to existing networks, and when they themselves may not see their connection, role, or power in these conversations? The onus, then, is on convening organizers to go outside traditional networks and invite new members into the space.

Beyond ensuring a broad representation of groups in such an effort, organizers must also be aware of ensuring the space is open and accepting of marginalized groups, particularly to avoid tokenism or further marginalization within the context of the event. The creation of an accommodating space is discussed further in the following

---

[82] D'Ignazio and Klein, *Data Feminism*.
[83] Nucera, "Opening Data."
[84] Costanza-Chock et al., "#MoreThanCode."
[85] Costanza-Chock et al.

pillar, although two points are worth mentioning here. One is the importance of ensuring all members have equal access to the media and technologies used in planning and conducting the event, so that they may act as both producers and consumers.[86] In defining roles for participants, it is important that they are given access to positions which allow them to create, contribute to, and maintain a level control of each part of the planning process. The second is that organizers must be mindful of existing power dynamics between collaborators. Power dynamics tied to perceived skills and experience may be more malleable through this process (particularly if each stakeholder is given a chance to contribute from their place of expertise and learn from the organizing process, yet it can be more difficult to totally overcome structural differences in class, race, and gender that exist outside of the context of the project.[87]

As these spaces strive towards creation of new knowledge, representations, media, or lines of arguments, it is also critical that time and energy are devoted towards identifying who is the audience of those results. Defining a single, monolithic public may be less productive than acknowledging and developing engagement strategies to reach multiple levels of the public.[88] Here again arises the importance of building multiple messages and touchpoints to reach a broader swath of stakeholder groups with the products and activities resulting from these events. Utilizing a data ecosystem model rather than a single training meant to reach all stakeholders can allow for those with varying skill levels to benefit without going to slowly or quickly through material, while

---

[86] Nucera and Bianchi, "How to DiscoTech."
[87] Sieber, "Public Participation Geographic Information Systems."
[88] Schlossberg and Shuford, "Delineating 'Public' and 'Participation' in PPGIS."

also building organic steps into greater levels of engagement for those who are

interested. Data and Digital Justice, as well as Data Activism, also raise critical

questions of ownership of products. They suggest, in line with the Open Definition at the

core of the Open Data Movement, that the creation of knowledge, tools, and

technologies must be shared freely and openly with the public.[89] At the same time,

where ownership of a result is important to its long-term sustainability, these fields

suggest considering diverse business models of control, such as collective or

cooperative business models, membership organizations, or municipal ownership.[90,91]

Addressing Physical, Technological, and Personal Barriers

> *Events shall consider how their choices around physical location*
>
> *and layout, technological complexity and scalability, and personal*
>
> *approaches to pedagogy and depth of engagement promote*
>
> *access and create barriers to participation.*

Just as the first pillar asks, *"Who?"* the second pillar addresses critical questions

of *"How?"*; as in, how are the events attracting or dissuading the participation identified

in pillar 1 from being involved. The first aspect of this pillar is related to the physical

space that the event takes place in, both in terms of its geographic location layout, as

---

[89] Nucera, "Opening Data."
[90] Costanza-Chock et al., "#MoreThanCode."
[91] Nucera, "Opening Data."

well as the ways in which those things impact participants' perceptions of the event.

Both of these factors impact the accessibility of spaces to those with physical

constraints or those who must travel far distances to attend, as well as the cultural

context within which topics are being discussed. For instance, it may be difficult to have

a critical conversation about a mayor's tactics for some participants (particularly public

employees) if an event is held at City Hall. Existing perceptions of a venue can lend to

the differential expectations of the session.[92]

At the same time, the layout of space has an impact on how co-organizers or

more general participants interact with one another and with any awareness or literacy-

raising processes. For instance, an event held in an auditorium or a lecture hall will

suggest a more explanatory session; an event held in a less formal setting, such as one

structured around booths, will facilitate greater engagement and exploration; an event

held around a single table may lend a sense of collaboration, yet will potentially yield

power to a moderator at the head of the table. Such choices must be made thoughtfully

to ensure spaces reflect the goals of the session, whether those goals are investigation

and organizing[93] or skill-building.

The second aspect of this pillar is related to the technology that is used at the

event. This is critical, as technological choices can reflect assumptions about the

prerequisite skills necessary to take part in this movement, as well as the replicability of

pieces of event in further workshops. While it may be appealing to strive to incorporate

the newest or trendiest technology, it is important to consider how much must be

---

[92] Sieber, "Public Participation Geographic Information Systems."
[93] Nucera and Bianchi, "How to DiscoTech."

learned by individual stakeholders in order to engage in the process (and which groups may benefit or be marginalized by using more advanced tools). Selecting a technology-intensive session may also produce barriers to reproducibility of the session: one that requires laptops or tablets and internet access may place constraints on how many participants can actively engage on the number of devices available, where additional events may be held based on power availability and bandwidth strength, and what types of organizations may be able to replicate the event on their own (likely those with greater technological resources or technically-proficient staff members).[94]

In other ways, technology can also facilitate greater access. When resources and activities are published online, they can possibly be shared with a wider range of participants. At the same time, providing analog versions of these resources can ensure that populations without equal access to the Internet may also be able to participate and learn from these efforts. Building multiple layers and forms of media and techniques can ensure just access.[95] As research from the Detroit Digital Justice Coalition shows, most community residents surveyed often learned about open data and online portals through in-person events, so these events remain relevant companions to online trainings and resources.[96]

The third aspect of this pillar is related to the ways in which choices around engagement methods allows for various ways of learning and various levels of participation. It is important to provide participants numerous touchpoints and strategies

---

[94] Sieber, "Public Participation Geographic Information Systems."
[95] Nucera and Bianchi, "How to DiscoTech."
[96] Detroit Digital Justice Coalition and Detroit Community Technology Project, "Recommendations for Equitable Open Data."

of getting involved; particularly to ensure that engagement doesn't necessarily benefit those who have the skills or resources to engage more deeply (as this may benefit younger, more wealthy, or technically able participants).[97] At the same time, when events are organized or resources are published, it is important to consider the languages, dialects, or forms of communication that are being preferred or excluded.[98] This flows down to the timing of certain publications (if an English-version is published before other translations, it may signal priority of access) or the availability of interpreters at events (those that advertise the presence of interpreters are more likely to seem inviting to those who may not prefer to engage in spoken English).

A leading example of inclusive engagement around data and technology is that of the Detroit Digital Justice Coalition and the Detroit Community Technology Project's DiscoTech (short for Discovering Technology) events. These events were designed to allow as many members of local communities to learn more about data and technology as possible through workshop stations with clear guidelines. To connect with audiences of all ages and skill levels and to allow participants to expend as much time or attention at any one station, they suggest keeping demonstrations short and interactive, documenting any step-by-step processes used in a guide, engaging participants to teach one another to encourage peer-to-peer learning, replacing overly-technical explanations or jargon with more familiar analogies, and conducting test runs of each station.[99]

---

[97] Sieber, "Public Participation Geographic Information Systems."
[98] Nucera and Bianchi, "How to DiscoTech."
[99] Nucera and Bianchi.

Mitigating Risk

> *Events shall be conscious of how they are presenting risks*
>
> *associated with datafication, putting participants at risk, and*
>
> *suggesting defensive strategies.*

The third pillar in this framework is related to the ways in which these events handle concerns of risk in both the material that they cover and the information that they consider collecting and storing about the participants. Although this concern is largely missing in previous PPGIS, it is at the core of not only Data and Digital Justice and Data Activism, but it is central in modern discussions of datafication.[100,101] Resistance tactics are starting to be published, although this is often done in such a way that marginalized communities are not exposed through publication of tactics.[102] These events are an excellent opportunity for activists to raise awareness of risks, share strategies, and ensure that participants can do so in a safe space.

One of the key issues of privacy is the risk of being personally identified and targeted in data. Despite anonymization processes, such as de-identification, randomization, or even aggregation, some data are unique enough or can be combined with other information to identify its origin.[103] This risk is inordinately borne by those who

---

[100] Milan and Velden, "The Alternative Epistemologies of Data Activism."

[101] Taylor, "What Is Data Justice?"

[102] Lewis et al., "Digital Defense Playbook."

[103] Detroit Digital Justice Coalition and Detroit Community Technology Project, "Recommendations for Equitable Open Data."

are already socially, economically, and politically marginalized. For instance, if they are

caught in certain databases, new migrants and returning citizens may face barriers in

accessing the rights and protections that most citizens enjoy because they risk being

flagged by the surveillance state. Other interviewees in the Our Data Bodies study

reported being held back in their access to job opportunities, social services, or even

internet content and media based on the way their personal data were being collected

and acted upon.[104]

At the same time, communities are at risk of being collectively profiled by opaque

systems and algorithms which are difficult to oversee and moderate. When algorithms

are hidden from the public, as many sentencing algorithms are, it makes it difficult to

audit their underlying assumptions or potential biasedness of the data that they are

trained on.[105] Furthermore, when data definitions and processing are unclear, it can

prevent all but content-area experts or insiders from using the data.[106]

Both of these things are being made possible in part due to implicit or obfuscated

methods of obtaining consent, or the un-anticipated repurposing of data. This is a risk

when extraneous data is collected on individuals in anticipation of it becoming useful at

some point in the future. Such data can be subject to accidental loss or unauthorized re-

use.[107] In addition, these efforts must be considerate of the history of datafication and

the extent to which they have been subject to research (and have lacked control over

---

[104] Petty et al., "Out Data Bodies: Reclaiming Our Data."
[105] Currie et al., "The Conundrum of Police Officer-Involved Homicides."
[106] Detroit Digital Justice Coalition and Detroit Community Technology Project, "Recommendations for Equitable Open Data."
[107] Detroit Digital Justice Coalition and Detroit Community Technology Project.

how their data is used).[108] These events serve as an excellent opportunity for

marginalized communities to build and mobilize defensive tactics against mass

surveillance and datafication, yet they must do so in a way that ensures participants are

not put at risk for simply attending and engaging.

Providing Opportunities for Meaning-Making

> *Events shall uncover the processes and people behind data*
>
> *collection/cleaning/use; provide community members an*
>
> *opportunity to point out gaps or inconsistencies; and empower*
>
> *community members to add or create individual or community*
>
> *interpretations of data that is collected on them.*

The fourth pillar is related to the opportunity space provided by these workshops

to empower active participation in data-driven systems. This is related to giving people

voice in the procedural aspects of data use, as well as the ontological aspects of

describing various realities through the use and augmentation of data-based insights

with lived experience. Oftentimes, these events focus primarily on the former point:

ensuring that data is being selected and used properly through community feedback.

On one hand, community organizations are going to have unique insight into local

circumstances or procedures and can ensure that the data that is being used is

---

[108] Petty et al., "Out Data Bodies: Reclaiming Our Data."

appropriate and accurate in the current context.[109] On the other hand, these events

provide an excellent opportunity for potential data users to connect with the authors of

those data. In elevating the authors of published data as people rather than sets of

abstract procedures, it can help demystify the choices made in how the data was

cleaned and aggregated (and can provide an opportunity for the author to learn from

users how the data can be re-published in a more helpful way).[110]

These events also provide an opportunity to interrogate the perspectives and

assumptions that underlie data analysis. This is related closely to the idea of

positionality, or the perspective from which an idea or theory is proposed. From the late

1980s, Haraway and other feminists within science and technology studies sought

theories of how meanings and bodies are created, in order to understand and critique

the impermeability of science, and landed on the notion of situated knowledges. To

achieve this, she suggests "politics and epistemologies of location, positioning, and

situating, where partiality and not universality is the condition of being heard to make

rational knowledge claims."[111] This strives to take a view from "a complex, contradictory,

structuring, and structured body" rather than a "view from above, from nowhere, from

simplicity." This concept has been extended to modern discussions as

technopositionality, or a positionality which is about technology and takes on the

discourses and technicalities of technology.[112]

---

[109] Sieber, "Public Participation Geographic Information Systems."
[110] D'Ignazio and Klein, *Data Feminism*.
[111] Haraway, "Situated Knowledges," 589.
[112] Wilson, "Towards a Genealogy of Qualitative GIS."

Once (techno)positionality is identified, the next step is to allow for multiple representations of data and knowledge. Combining data from multiple sources and forms of knowledge can provide a deeper, more meaningful, more timely representations and visualizations of current conditions; however, creators must be mindful not to reduce the complexities of traditional or local knowledge to fit a specific medium. The best model of representation should fit the needs of the communities being centralized in this work.[113] Outside of mixing sources and types of data, visualizations and representations should also find ways to show multiple perspectives and truths so as to remove the primacy of a single narrative.[114]

Furthermore, this serves as an opportunity to challenge prevailing narratives and contest the legitimacy of any singular truth that is being advanced to the detriment of marginalized groups, using data-driven insights and lived experience together. Data Activism argues that this is important for the efficacy of data, as we are better able to evaluate the qualities of data within the context of power struggles.[115] This process of challenging existing frameworks is one that allows communities to retain cultural agency. These events can incorporate the digital justice principle of the right to tell one's own stories by providing individuals and communities both the opportunity and platform to do so.

---

[113] Sieber, "Public Participation Geographic Information Systems."
[114] D'Ignazio and Klein, "Feminist Data Visualization."
[115] Currie et al., "The Conundrum of Police Officer-Involved Homicides."

Creating Direct and Indirect Impact

> *Events shall be driven by explicit goals, tied to measures of*
>
> *impact; these may include 'direct' effects of raised data*
>
> *awareness, literacy, and/or action coalescence, or 'indirect' effects*
>
> *of capacity building, resource sharing, and/or social network*
>
> *development.*

The fifth, and final, pillar is related to acknowledging various definitions and achievements of success. Numerous PPGIS projects have discovered that defining goals is not a one-size-fits-all process; instead, it requires negotiating the priorities of different stakeholders. It also requires managing the consequences of success on both potential beneficiaries, as well as on those who may be potentially harmed as a result.[116] Measuring impact can also be difficult to tangibly define, especially when in pursuit of causal results. Yet the ability to cite quantifiable outputs and outcomes, however simplified, often makes fundraising for further efforts more possible.[117] As a result, organizers benefit from being able to point to both 'direct' outcomes of these events, as well as 'indirect' outcomes.

What are some crucial 'indirect' outcomes to look for in these types of engagements? One is the capacity that is built among co-organizers and participants of the event. In the process of organizing around collectively-interpreting data about their

---

[116] Sieber, "Public Participation Geographic Information Systems."
[117] Sieber.

surroundings, communities may develop a greater sensitivity and consciousness of the issues around them. This benefit should not be understated as a positive result, as it allows communities to make progress towards Freire's concept of liberation through declaration of situation. [118] The products of these events can also help continue capacity-building processes into the future. For instance, media and technology that is either used, developed, or adapted for the workshop should be integrated into further educational opportunities to open up the learning process up to a wider audience.[119]

These workshops also serve as an excellent opportunity to build social networks around data. Data Activists have utilized such in-person workshops to "organize organizers," around issues of data and technology.[120] In the interest of building a broader community and transcending expectations of who is doing this work, organizers should seek out strategic partnerships with local organizations to collaborate with to develop methods and put on the event, as a way to build longer-standing relationships.[121] Even in those contexts where open data is not immediately valuable to community activists, partnerships with larger academic or non-profit organizations that are more directly active in the open data movement come with the benefits of legitimacy and resource mobilization that can be difficult to access otherwise.[122]

The next two chapters will describe community-based data engagement efforts run by local data intermediaries in Pittsburgh and Los Angeles, respectively. The final

---

[118] Meng and DiSalvo, "Grassroots Resource Mobilization through Counter-Data Action."
[119] Nucera, "Opening Data."
[120] Nucera and Bianchi, "How to DiscoTech."
[121] Detroit Digital Justice Coalition and Detroit Community Technology Project, "Recommendations for Equitable Open Data."
[122] Meng and DiSalvo, "Grassroots Resource Mobilization through Counter-Data Action."

chapter, chapter six, collectively reviews the choices made by each local data

intermediary along the way to put on the events, as well as the structures put in place to

continue engaging participants afterwards. Through this collective reading, we can

sharpen the efficacy of the framework developed in this chapter.

Chapter 4. Data Days: Bringing Data to the People of Pittsburgh

The Western Pennsylvania Regional Data Center (WPRDC) is a major local data intermediary in the city of Pittsburgh. It was launched in 2015 as a partnership between the city and the county to publish open data, and has received support from the Richard King Mellon Foundation, the Heinz Endowments, and the University of Pittsburgh.

WPRDC is hosted at the University of Pittsburgh's Center for Social and Urban Research (UCSUR). While the Regional Data Center has existed for only a few years, UCSUR has long strived to make community information accessible to residents. Prior to the launch of WPRDC in 2015, UCSUR had run the Pittsburgh Neighborhood and Community Information System, and is currently home to two other community indicator platforms on the city and the region (one of which, Southwestern Pennsylvania Community Profiles, shares staff with the Data Center).

The Data Center was formed, in part, to address gaps in legal and technological infrastructure that previously existed in the city's effort to enable broader data use. The Data Center has maintained the Alleghany County and City of Pittsburgh's open data portal. But beyond simply publishing data to enable research, analysis, and decision-making, the Data Center was founded to enhance civic engagement and build local capacity among community organizations and regional stakeholders to use data. This included partnering with the local "civic hacker" community that had grown around the burgeoning technology industry in the city, supporting a local data users group, and continuing to support other municipalities throughout their region. Community

engagement remains a central part WPRDC's mission, and a key driver of the Data Days program.

Data Days: A "Family-friendly Community Fair"

From 2009-2015, UCSUR hosted an Annual Data User Conference, attracting between 100 to 200 attendees and centered around panels, guest speakers, and breakout sessions. The conferences were intended to serve those who were already working with data, and would focus on issues of particular interest to data professionals. For instance, during the last conference in 2015, there was a debut of the Center's new open data tool, and the attendees were asked to give feedback on its design and the data sets that were published.

Around the same time, Eleanor Tutt started as the open data and knowledge manager at the Carnegie Library of Pittsburgh, and was interested in partnering with WPRDC. Eleanor and the WPRDC staff knew each other from previous work. They decided to partner together to redesign the annual conferences for a number of reasons, including to better build relationships.

In her work, Eleanor was familiar with the Data DiscoTech model created and run by the Detroit Digital Justice Coalition (DDJC) and was curious to learn more about it. She attended a Data DiscoTech in Detroit and noticed not only the logistical choices of the events, but also their look and feel. The space that was created was inviting, with both food and music, and the event looked more like a festival than a data training. While the library could not replicate quite that dynamic, particularly the music, Eleanor

felt that this model was a positive one to emulate as she navigated the particular unique opportunities and challenges that came with increasing data access from within a public library.

This model became the inspiration for a new 'Data Day' event that would be held in October 2015. WPRDC and the Carnegie Library of Pittsburgh partnered more formally with the Detroit Digital Justice Coalition to share resources and support in organizing Data Day in exchange for just compensation, and leaders from DDJC agreed to attend the Data Day event. The event was held on a Saturday morning from 11am to 2 pm in the main branch of the Carnegie Library of Pittsburgh and sought to reach and learn more about a broader audience than those who were coming to the previous User Conferences.

The library advertised the event through their paper and online calendars and placed fliers for it at all branch locations, while WPRDC fliered around the city. The fliers were colorful and inviting, describing the event as a "family-friendly community fair." Information was posted to social media (in the second year, a promotional video was produced and shared), an invitation was sent out through WPRDC's newsletter, and the event was even advertised on the radio. Yet, despite these extensive efforts to advertise the event to a broad range of city residents before the event, a heavier focus of outreach was to pull passersby in to check out the event, particularly library patrons, on the day of the event.

One unique tactic used in the library to raise interest in data was to display chalkboards leading up to the event which asked people data-related question (for

instance, 'do you like fries on your salad?' alluding to the famous Pittsburgh culinary tradition, or 'what data question do you wish you could answer about the city or library?'). These were meant to catch strangers' attention as they walk through the library, and provided an easy way of creating collective, analog visualizations.

The organizers of the event based many of their values off of the success of Data DiscoTechs in Detroit. Primarily, the events encouraged awareness and accessibility of data for those in attendance. They were meant to humanize and personalize data, from the hands-on activity at every table at each Data Day to the presence of a 'data professional' to answer questions about their work and pathways into data analysis and data science for attendees who were interested in pursuing a career in the field. These events were meant to be interactive and therefore followed the DiscoTech philosophy of 'show-and-play.' They were meant to build confidence among attendees in understanding and working with data so that they would know how and when data could be used in their lives, and so that they would know where to find this data. Finally, diversity was a primary value of the event in terms of race, gender, age, and the topics covered by table leaders. To appeal to participants of all ages, the event was designed to be exciting and accessible to children—through playful stations, theme alignment with other activities being held in the library (like Halloween events), and providing candy at many tables.

The organizers were wary of only drawing the same audience who had attended the previous iteration of their User Conferences (although that group was not cast aside, they were still encouraged to participate in Data Day). Instead, they wanted to go

beyond their immediate network in attracting both participants and table leaders. This was meant to challenge ideas of what counted as 'data work' or who could be seen as a 'data user,' as well as to expand the diversity of the city's network of data users. To bring in a diverse range of table leaders, WPRDC and the Library had to connect with community partners and personal connections to see if they could help identify people working on projects that had a loose data component. They also encouraged activists to join as table leaders, following the broad definition of 'data practitioner' used by DDJC. Ultimately, table leaders included representatives from the city and county governments, some of their colleagues and students from the University of Pittsburgh and Carnegie Mellon University (which are both in close proximity to the main branch of the Carnegie Library), "civic hackers" or representatives from civic technology start-ups, and, crucially, community members. Organizers observed that the diversity of participants reflected the diversity of table leaders; their most diverse assembly of table leaders was for Data Day 2016, which was also their most diverse and well-attended event, drawing between 150 and 200 participants.

Each time the event was held, it featured around 10 to 20 tables. The organizing team instructed table leaders to make their presentations as interactive as possible. For some table leaders, including those who were community members, this meant that the organizing team took a step back and allowed them to show off the art project or civic technology project that they were working on (and did not require that table leaders had to explicitly discuss the open data portal or open data at all). For others, the organizers had to help them creatively brainstorm a way to make more procedural data issues

more interactive or relatable. For instance, when an environmental group that had conducted a census of fish in a nearby body of water led a table, they were encouraged to create a fishing game where the game pieces matched the fish in their census. Participants were able to simulate the process of discovering and collecting data on fish, and from there were able to ask questions about the process and discuss the importance of collecting that data.

After the event, the organizers took different follow-up strategies with table leaders and participants. Table leaders were asked to describe their experiences after the events so adjustments could be made. The organizers highlighted that a greater consequence of these events, however, was the growth of the data network, an impact that they had not originally expected. As far as they know, there have been a few one-off projects from partnerships made between table leaders (although it can be difficult to tell whether those bonds were necessarily forged at a Data Day event). The DDJC DiscoTech model encouraged special opportunities to be set aside for table leaders to see what each other were working on and network with one another. After the first years, separate events (like a pre-Data Day party) have been hosted for table leaders to get to speak to and get to know one another.

Among participants, these events served as an 'on-ramp' to become new members of civic data and technology space within the city. WPRDC and the library hosted a series of data literacy trainings, called Data 101, on basic data concepts, visualizing data, mapping data, and storytelling using data. These workshops were intended to be a next step for Data Day attendees who were new to using open data but

wanted to build basic data literacy skills. By separating awareness and literacy-raising efforts, WPRDC was able to create unique sessions that catered to more specific audiences in each case. In addition, WPRDC holds monthly open office hour in their offices. These sessions allow data users to ask deeper questions about using and applying data, troubleshoot software or tools, and generally receive support from Data Center staff.

Data Days was ultimately run three times, in October 2015, October 2016, and October 2017. A 2018 event was postponed as, by that point, Eleanor had left her position at the library. A new staff person has been hired to take over the position, and there are intentions to conduct further workshops. On the other hand, the Data 101 trainings have been placed on hold as the Data Center considers the next phase of its community outreach. Office hours continue to be offered.

The collaboration between WPRDC and the Carnegie Library also kickstarted a conversation about the potential for partnership between libraries and local data intermediaries to better serve data users, further democratize data, and support more equitable access to information. This model inspired the Civic Switchboard initiative, where stakeholders in Pittsburgh would help publish best practices in library-local data intermediary collaborations and seed pilots in other places around the country, funded by the Institute for Museum and Library Services. The first round of nine grant awards kicked off in 2019.

Chapter 5: Neighborhood Data for Social Change: Technical Literacy and Data Stories in Los Angeles

The Sol Price Center for Social Innovation was created in 2011 within the University of Southern California's (USC) Price School of Public Policy to conduct research and develop strategies to improve the quality of life of residents in low-income urban communities. The Center focuses broadly on initiatives which evaluate programs and policies, promote research, and enable social innovation.

A year after the Center for Social Innovation was founded, a push for an open data intermediary began in LA. A series of conversations, organized by Melody Winter Head at the Los Angeles Branch of the San Francisco Federal Reserve Bank and Caroline Bhalla at USC's Price School of Public Policy and encouraged by the national NNIP staff at the Urban Institute, launched about open data publication and access in LA. Soon that conversation moved to the Price School of Public Policy, where Professor Gary Painter, Director of the Center for Social Innovation, agreed to house the effort in the Center. In 2016, the Center for Social Innovation applied to be and was accepted as an NNIP Partner, and by November 2017, the Neighborhood Data for Social Change (NDSC) platform was launched to make community data accessible to Los Angeles and the surrounding area.

When the NDSC data portal was launched, it was accompanied by a curriculum for training and technical assistance, as well as visualization, mapping, and narrative tools to make data more usable. Staff at the Center for Social Innovation began to offer monthly data trainings on-campus and community-based trainings, hosted by

community partners, on a semi-regular basis. This type of work is what sets NDSC apart from other efforts in the city, which has a growing shared data environment. While the city government has published data through Open Data LA, and other organizations like the Southern California Association of Governments and Healthy City, have published indicators, NDSC attempts to further democratize data in the NNIP mold by engaging with the community and focusing their outreach on low-income communities.

Neighborhood Data for Social Change: A portal, a storytelling platform, and a series of trainings

Caroline Bhalla, who is now the managing director for the Center for Social Innovation, was previously at New York University's Furman Center for Real Estate and Urban Policy. While she was there, the Furman Center developed an online data portal and ran trainings, both on-campus and around the city. Through this work, Caroline built a commitment to community engagement through data and gained experience running workshops in the field. The format of the trainings in LA were inspired by the trainings Caroline gave in New York.

NDSC trainings are run in two different formats. They are offered on USC's campus on the third Wednesday of every month as one-hour workshops beginning at noon. These trainings are generally held in a computer lab in the Price School of Public Policy. They have also been offered periodically off-campus at partner organization's offices or meeting spaces. Partners have ranged from community organizations and foundations to local branches of state and federal agencies to librarians and grant

writers. These are generally held in the morning for staff or members of the host organization.

The exception were a few off-site workshops run by EMPOWER LA (an initiative within the Mayor's Office for Neighborhood Empowerment) which NDSC participated in. EMPOWER LA hosted nine or ten community-based trainings, held from 5:30pm to 8:30pm in community-based sites around the city. These trainings would feature several 30-minute trainings offered by a rotating stable of city data employees, representatives from civic data organizations, and trainers from the geospatial software company Esri. NDSC presented at the end of a few of these trainings, and found this to be an especially effective way of connecting with community residents who were not as deeply entrenched in the civic data community.

For on-campus trainings, between 15-40 people usually sign up and up to 30 attend each month. Attendees are most often students and researchers, but some non-profit and foundation staff members, city employees, and, occasionally, members of neighborhood councils have attended. While some participants travel across the city to USC to attend these trainings, most of them are doing so on behalf of an employer and are compensated for the time they take to attend. On-campus trainings are primarily advertised on the NDSC website and through the NDSC mailing list of 2000 recipients, as of February 2019. On the other hand, the outreach for off-campus events is left up to the hosting partner organization, so the audience for those trainings are generally representative of the partner's staff or community. The decision to leave advertising to partners is, in part, is a strategic move by the Center for Social Innovation; community

organizations have stronger connection to community residents than the Center does, so collaborations help them reach new audiences. Local partners are also responsible for selecting an appropriate venue to hold the trainings and providing the minimum necessary technology to host the trainings: a projector, a screen, and internet access powerful enough to support multiple users rendering maps and visualizations. All participants are encouraged to bring their own laptop, although the training team have often brought a few spare laptops in case someone did not have one.

So far, NDSC largely has relied on community partners who have either been previously connected to the Center for Social Innovation (either through previous collaboration or referrals by someone close to NDSC) or who have attended an on-campus training. Yet, the NDSC staff are working more to broaden their outreach and impact. Currently, the Center for Social Innovation is working with a few community organizations to tailor their trainings to better match the skill level and skill needs of community members. They have also tried to reach a wide range of geographic locations in their off-campus workshops by hosting them in areas that have been historically underserved by data, including several trainings in South Los Angeles and one in the Coachella Valley.

The most successful trainings in attracting an average citizen were those hosted by EMPOWER LA. This was due to the success of organizers in the city's Neighborhood Empowerment department working closely with the Neighborhood Councils to mobilize and invite community members to the trainings. Those workshops were hosted in local spaces, particularly large communal spaces owned by non-

profits—libraries, learning rooms, conference rooms, etc. They attracted a diverse range of participants in terms of age, race, gender, and ability, which pushed trainers to deliver their trainings in a more inclusive way.

In all of the trainings they have run, NDSC staff have centered two primary goals. Trainings focus on building data literacy skills and confidence to independently access data among participants. To those ends, these one-off trainings focus on tool literacy. Trainers demonstrate different features of the website, and then allow participants to explore the various ways that data has been documented and explained on the platform. Trainings also try to highlight relevance of different data sets to participants' work. One trainer mentioned that a goal for the trainings was for participants to have a complete understanding of what is available on the portal, and for them to share what they learned with others. As a result, the trainings begin with personal introductions, and each participant is encouraged to share what they hope to get out of the training. Trainers generally modify the training to match interests in the audience (for instance, by using contextually relevant data in their mapping and visualization demonstrations).

The current curriculum of the trainings mirrors these objectives. Student workers at the Center for Social Innovation developed the training materials with the goal of showing a new user how to use the platform, and students help deliver or support most of the trainings. The trainings have iteratively grown to cover all three portions of NDSC (i.e., the mapping tool, data stories, and portal to explore and download data) and include a description of the features that set the NDSC portal apart. The workshops conclude with a feedback form, which asks for participants to provide feedback on ways

to make the platform more intuitive, whether there are data sets they would like to see on the portal, and provides them with an opportunity to submit a testimonial. More recently, the NDSC team created a worksheet for participants to take with them with reminders of the skills that covered in the training, and they are working to create a webinar version of the training. The online mapping tool already features a short training video and step-by-step description from Socrata, the software vendor tapped to create the platform, on how to access data at different geographic levels and how to create a user-defined "neighborhood" (a custom geography created from smaller standard units, such as census tracts) to uniquely aggregate data. Presumably, an additional webinar would cover the other aspects of the in-person training, such as introducing users to the site's data stories or exploring data indicators.

While the mapping tool is often the most popular one among new users because of its ease-of-use, the data indicators provide extensive information for data users. They are grouped by 'policy area,' framed with metadata explaining both the source and definition of the variable, and include a short description of why each indicator is important to measure, with links to related data stories and citations to external research.

NDSC's data stories combine data insights, charts, and maps with a narrative, providing greater context to a policy or planning issue of note in greater Los Angeles. These stories often focus in on a particular place in Los Angeles, Los Angeles County, and, as of recently, the Coachella Valley, whether that is a neighborhood or a small region, to both ground the narrative and leverage the platform's ability to easily create

unique aggregated geographies. The stories from Los Angeles cover a broad geography—from downtown and Central Los Angeles to South Los Angeles and Long Beach to the San Fernando and San Gabriel Valleys, Stories also connect local neighborhood issues with larger policy issues through the addition of background literature; for instance, research has been cited to frame the importance of a particular indicator in understanding or measuring a broader outcome. The stories also try to humanize the issue using photos, videos, or first-person quotes, if possible. While the first few stories published on the platform were based on data trends identified by the Center for Social Innovation's staff, the rest of the stories have been created alongside a community partner. As of May 2019, the platform features twenty-three stories about greater Los Angeles, and three stories about Coachella Valley.

Data stories provide a deeper and more structured opportunity for outside partners to use NDSC to talk about an issue that is important with them. It also helps build a relationship between the organization and the Center for Social Innovation. As one staff member described it, working on a data story gives community organizations a research partner to help pursue a short, month-long project. From the opposite perspective, collaborating with community organizations provides the Center for Social Innovation a way to ground their data in issues and initiatives relevant to local stakeholders. So far, the process of co-creating a data story has been initiated when a community partner has reached out to a member of NDSC's staff. In some of those cases, partners have made a connection to NDSC after attending a training. Partners have included several collective action networks and neighborhood organizations, a

business council, a food bank, other small- to mid-sized non-profits, an elementary school, Federally Qualified Health Centers, and several municipal initiatives and county agencies.

The process of creating a data stories is facilitated by NDSC staff. Partners are engaged at the beginning of the process to identify which data available on the portal are of most relevance to their work. Partners may suggest potential story lines or goals that should guide the narrative creation process, as well as the intended audience for the story. They are also encouraged to identify and share data or other media that they hold to provide more local information to the story. At that point, the partner's role has generally shifted to providing feedback. NDSC produces an outline of the story and then a full draft of the story, both of which the community partner is given the chance to review, but they are not actively involved in the drafting process. As these data stories provide a unique opportunity to community organizations, the Center for Social Innovation have already identified partner-driven data stories for the next six months.

Chapter 6: Grounding the Framework in Reality

The framework proposed in chapter three attempts to provide a broad basis to guide the values and decision-making processes when organizing a data-based community engagement event. But, by its construction, the framework is building from the theoretical combination of several different fields of thought. To better define its value in the context of what is potentially attainable within real-world constraints of organizing and achieving such an event, it is helpful to compare it to work that is currently being done.

This chapter compares the considerations suggested in the framework to the cases presented in chapters four and five. In doing so, the intent is neither to invalidate considerations in other fields, nor to critique the decisions made in the two case studies. Instead, it is to highlight the challenges in implementing different portions of the framework and suggest strategies for organizations interested in putting on their own data-based engagement events. The chapter reviews each pillar of the framework individually, although aspects of each case study often span several pillars. Finally, the chapter concludes with general proposal for how to use this framework in the future.

Reviewing the Pillars of the Framework

Defining Primary and Secondary Audiences

In Pittsburgh, organizers held equity and diversity as central goals, and determined that the best way to reflect that in their attendance was by selecting a wide range of table leaders. By many accounts, they succeeded in reaching members of

communities who have historically not been included in data work. The organizers at WPRDC and the Carnegie Library intentionally cast a wide definition of what could be considered "data work" and purposefully looked for non-data professionals to present. This was meant to challenge notions of what constitutes data work or a data "professional." Community members were recruited as table leaders through WPRDC's community partners and even, in a few cases, through serendipitous meetings. Often, those potential table leaders needed to be convinced of the relevance of their work to a data event or, at a deeper level, the legitimacy of their presence at the event as individuals conducting this work informally outside of the context of larger companies or institutions. The organizing team succeeded in recruiting hesitant community members by inspiring trust, excitement, and confidence in them. They also provided members of the data community impactful roles in shaping the events. By ceding much of the control over the design and management of tables over to a diverse range of partners, the events reflected a more organic representation of the range of uses for data.

Their efforts to engage an audience who would not otherwise attend a 'data festival,' particularly by persuading library patrons to check out the unexpected event, also seemed successful. Compared to the previous iteration of the user conference, WPRDC and the Carnegie Library were able to expose new audiences to many different forms of data collection and use, in ways that were relevant to their lives. At the same time, the presence of content-area experts or previous data users waned compared to the previous iteration of the conference. Some continued to attend based on excitement around data, but others may have been dissuaded by the informal nature of it. As many

of the attendees did not have a background in data, it was pointed out that the
conversations between attendees and table leaders seemed to remain surface-level.

In Los Angeles, the connection to a broad range of community partners is
growing hold now as the platform develops. Currently, many of those who have
attended the workshop on-campus were associated with the university in some way, or
had found out about it through the organization's newsletter. A few of those who found
out about it in this way were already working with data as a large part of their jobs. As
the portal and trainings are still relatively new, the Price Center is working to provide
different resources to each of its audiences. On one hand, they aspire to add to the data
available on their portal so that data-savvy users (including grant writers, nonprofit
leaders, and academics) have easy access to the indicators they need at various
geography levels. On the other hand, they are modifying their trainings so that they are
better able to engage residents to learn about and change their neighborhoods. This is
particularly important as their trainings expand into the Coachella Valley, which has a
very different landscape, geographically and in terms of data use. NDSC have
responded with a unique training that has been deployed as they enter this historically
underserved area.

The data story partnerships provide a unique model for deeper engagement,
particularly for those less conversant in data; however, as these partnerships are
currently constructed, they are predicated on an existing organization with existing data
capacity. If shifted slightly to formally give community members points in the process to

actively engage in the narrative creation process, the data storytelling process could

open up more of a dialogue than simply serve as an educational publication.

However, despite the fact that data story partners have been given a rather rigid

roles of providing feedback through the process of drafting data stories, they have taken

the lead on defining audience for the stories. This has happened both at the beginning

of the story writing process and after it has been published to determine how the story is

used. For instance, one of NDSC's partners conducted a community survey and held a

data dialogue with local residents to better contextualize the survey results. While this

was done prior to the story creation process, it was a critical step towards making the

data story that was relevant to the community. The community organization brought that

perspective to the story creation process, and wrote a story that directly responded to

the community's concerns about health. After stories were published, partners also

have some control over how they can be repurposed. A story originally written for

policymakers or senior staff members leading a non-profit, for instance, can be used to

showcase data driven insights to potential funders or grant makers. This ability to

repurpose the story to fit an organization's needs provides the community continued

autonomy to define whom they are speaking to.

Addressing Physical, Technological, and Personal Barriers

In Pittsburgh, physical, technological, and personal barriers were addressed so

that the event could be more inclusive. The library was selected by organizers both

because of the strategic partnership with staff at the Carnegie Library of Pittsburgh, and

because libraries seemed to be a natural fit to connect with community residents looking for information. The set-up of the event, with numerous tables lining several rooms near the central area of the library, encouraged multiple levels of engagement. However, one side-effect from selecting the main branch of the library as the location was that the audience that were drawn in largely reflected the area surrounding the library: namely, students and middle-class families.

Data Days in Pittsburgh also incorporated a fairly even balance of digital and analog presentations across tables. This was especially reflected in engagement tools. From the chalkboard which posed data questions to library patrons before and after the event to raise awareness of it, to public participatory art projects that were set up during the event, the organizers and table leaders were able to inspire a sense of data culture that was separated from the need for technology.

From a pedagogical approach, the organizing team in Pittsburgh met its intended audience of people who were new to open data at a beginner's level. They encouraged table leaders to present projects and data sets that the table leaders had unique insight into, which helped them all, even those who were coming from fields outside of data, speak from a position that they were comfortable with and quite knowledgeable about. At the same time, the organizers pushed the table leaders to structure their activities from a place of accessibility and interactivity, rather than one that centered teaching "expert" knowledge. This seemed to be successful, as the organizers found that the tables that were most well-attended were those that had no learning curve or required little explanation. In this way, the tables at Data Day were not structured in a way that

placed certain levels or types of expertise above others, but rather encouraged all participants to try something new in a low-stakes yet engaging setting. This also provided participants a range of low-commitment access points to different data-related topics to try out. They were not obligated to stay at any one station for a certain amount of time, and they were not obligated to see all of the stations, making it potentially easier for a passerby to join the event. At the same time, the organizers used strategies, most notably a Data Day "passport" system where participants could receive stamps from every table that they engaged with, to encourage attendees to see all that the event had to offer.

In Los Angeles, the choices made around physical, technological, and personal barriers were often made to encourage a data-literate audience to navigate and use the tool. The fact that the trainings were held at the Price School of Public Policy raised questions about who the intended audience was: was it nonprofit staff and residents, or was it policy researchers and academics? Instead, NDSC trainings found greater success reaching a broader cross section of the population, particularly attendees without a pre-existing connection to an organization, during city workshops. This speaks to the power of organizing through the Neighborhood Councils that they were able to attract residents without clear existing connections to open data, but also to the power of meeting people where they are.

The sessions were predicated on technology. On one hand, this encouraged users to actively try the tool; compared to a demonstration where participants were only expected to watch someone else use the tool, the NDSC trainings were more hands-on

and broke down an initial barrier by building tool literacy and confidence. By splitting

trainings into two halves where participants were encouraged to explore the tool on their

own with opportunities to ask questions and make requests of the trainers, NDSC staff

gave attendees a supportive chance to explore what is available and receive immediate

feedback in case they got stuck. During the latter half of the trainings, NDSC staff also

reported to have received a number of questions about the availability of certain data

sets; this is noted down and periodically taken into consideration whenever the Center's

data publication schedule is being determined.

On the other hand, this situated the things that participants learned specifically

within what was shown; participants reported having trouble extending their knowledge

to other applications of the tool after the workshop. This also posed a slight challenge

when conducting off-campus trainings. The need for additional technologies, such as a

projector, screen, and sufficiently-stable internet access that all participants could use to

render the mapping tool, could limit future partnerships, or prevent attendees from

recreating the training within their organizations (except in a one-on-one fashion).

The data and information published on the website provide information that could

be helpful to stakeholders with both a deep understanding of data as well as those who

are not quite as sure how to use data. The organization of data indicators under policy

issues and the inclusion of examples of how to use each indicator that is available on

the platform provides structure to those who are approaching data related to a given

topic for the first time. However, these contextual explanations, as well as the data

stories, tend to skew towards a more policy-driven audience in the language they use

and research that they cite. However, some of the data stories, particularly those written for community members, have been thoughtfully edited with that audience in mind. For instance, one organization included community leaders in the revision process. In these cases, organizations have also worked to have the stories translated into the other common languages spoken by their residents for greater accessibility.

Mitigating Risk

The third pillar of the framework seemed to be the least consciously addressed by either event, but particularly by those in Los Angeles. While presentations on the risks of divulging personal information may not make sense in the context of platform trainings, the process of receiving consent and having local stakeholders check data story narratives to ensure no identifying information about residents in the inclusion of local data and quotes could be incredibly useful.

On the other hand, Data Days more actively tried to at least bring up topics of risk mitigation in the face of great datafication of society through different tables. For instance, their partners from the Detroit Digital Justice Coalition were given a table at each of the three Data Days. Among the activities that DDJC conducted were ones published in the Digital Defense Playbook, which they helped contribute to. The Digital Defense Playbook,[123] published as a resource from the *Our Data Bodies* project, provides workshops, tools, and tips to help co-create and share knowledge and strategies to advance data justice. For future events that are interested in incorporating

---

[123] Lewis et al., "Digital Defense Playbook."

activities that raised awareness about privacy, anti-surveillance, and demystifying opaque algorithmic systems, this toolkit is an excellent resource to other organizations looking to conduct community-based data events.

Pittsburgh was also very conscious about asking attendees to hand over their personal information. The organizers decided not to collect attendance at the event, not only to avoid dissuading people who had not signed up in advance from attending, but also to avoid collecting more information than was needed. Participants were given the opportunity to sign up for WPRDC's newsletter at the event, but even this sign-up sheet was not heavily pushed upon attendees. Several table leaders seemed to be collecting survey information at their stations, and it is unclear whether the organizers were aware or tried to moderate that. At future events that are committed to the privacy of attendees (especially events that hope to provide a safe space to discuss and share anti-surveillance strategies), organizers may consider producing guidelines that favor privacy for table leaders for asking for informed consent before collecting information from attendees.

Providing Opportunities for Meaning-Making

The efforts profiled in the previous chapters provided the chance to connect data to its positional source and for community members to share their perspectives as they related to open data. The NDSC trainings provided participants an opportunity to speak to the NDSC staff who were accessing, processing, and uploading data from other sources to their platform. This gave participants an opportunity to suggest further data

sets to be published, and allowed them to speak directly to those who were drafting data stories. However, the city data trainings provided attendees even closer access to the authors of primary data. Oftentimes, those giving data presentations at city trainings were also the people directly responsible for managing that data within the city. If community members had questions for the data author, this forum served as an excellent opportunity to connect with them. Participants were encouraged to follow-up with presenters after the events, and a few did to gain greater insight or assistance in the way data was defined and published in the city.

The trainings also provided opportunities for NDSC staff to check themselves and the work they were doing. In connecting with community members and content-area experts who attended the trainings, NDSC staff were able to "ground truth" some of the other work that the Center for Social Innovation does on researching issues facing low-income communities. This also serves as an opportunity to ensure data is defined in the most useful way. For instance, during one data story-creation process, a food bank alerted NDSC staff that the demographic cross-sections of food insecurity that were being published on the platform were obscuring the experiences of groups who were generally at the greatest risk of experiencing hunger (namely, senior citizens and children). NDSC republished the data with these vulnerable subpopulations separated out from other ages so that others would be able to access the most relevant information related to this issue.

In Pittsburgh, Data Day provided a similar opportunity to interact with those creating and publishing data. Some tables were run by the authors of various sets of

data, including city and county employees who were responsible for creating and

publishing data tools, as well as representatives from community organizations and

grassroots data collection efforts. While the informal nature of the event may have been

an excellent opportunity to inform the public of the availability of tools and data sets and

to field impromptu questions, the format of each table influenced how likely it was that

casual participants would pose critical questions. Other tables, including those run by

researchers, journalists, artists, advocates, and entrepreneurs, explicitly engaged

participants in a process of sharing their perspective. This came in the form of

crowdsourced data collection activities (such as a crowdsourced community asset

mapping), narrative creation activities, and prompts to provide feedback on tools and

research.

Creating Direct and Indirect Impact

Organizers in Pittsburgh and Los Angeles were able to achieve goals that they

set for themselves, and saw unexpected indirect benefits from their events. In Los

Angeles, the data stories have provided partners a way to speak across literacies about

their work. This has been helpful for community organizations to inform their processes

to be more data-driven, explain their impact to outside funders, and better advocate

their policy positions to policymakers. However, organizations have highlighted the

importance of creating a second step of this process that is run within their local

networks of organizations and residents to create a two-way, ongoing process. This

could ensure that service partners and the populations they served were on the same

page and speaking the same language, but also would lower the barrier for individuals

to speak to institutions which one interviewee identified as the most empowering result

this could bring about.

A useful model to open-up the process to greater community participation comes

from Rahul Bhargava and Catherine D'Ignazio's DataBasic platform.[124] They advocate

teaching data literacy not using a tool designed for users but one that is explicitly

designed for learners. Such a tool or activity should be focused, guided, inviting, and

expandable in its pedagogical approach. While the existing NDSC trainings are guided

and generally focus on one tool, a future training could be framed in a more inviting and

less policy-specific way to engage a broader audience of beginners. That could suggest

a less technical and more conceptual lesson that goes over some of the skills needed to

make an interesting and relevant inquiry of the data, and could be practiced a few times

so that individuals would get used to asking relevant questions that could be answered

with data from an open data portal. The current worksheet that NDSC staff have

developed as a take-home reminder of skills discussed could serve as inspiration for

future expandability of community-focused trainings.

Whereas Los Angeles set tool literacy as a central goal for their event, Pittsburgh

was able to narrow their scope to engaging citizens who were new to this topic without

overwhelming them by placing the events within the context of a broader ecosystem of

engagement. WPRDC and the Library saw success, in part, as whether or not these

events introduced local residents who held a broad range of technical and personal

---

[124] D'Ignazio and Bhargava, "DataBasic."

backgrounds to different applications of data. Going a step further, they hoped that these events would encourage participants to engage with the Data Center or the Library further on the topic of data, through Data 101 trainings, WPRDC office hours, or other data events (such as data exhibits or talks) organized by the Carnegie Library and Museum. Because attendance was not tracked at the events, quantifying their long-term impact was difficult to do, but these follow-up workshops were popular among attendees and organizations hoping to recreate them. As a result, WPRDC published these resources in a toolkit through the NNIP Network.[125]

Pittsburgh especially succeeded in providing an opportunity for the community participating in data-driven policymaking and planning to grow. One of the key takeaways from Data Days was the importance of creating the environment to foster social networking. The events succeeded in beginning conversations between participants and table leaders through hands-on activities. Some tables presented topics that were related to current political action or direct civic engagement, encouraging participants to continue to follow and support the issue outside of the context of the event. The table that was offered each year allowed participants to meet a professional data scientist allowed for professional networking.

The importance of allowing table leaders to engage as participants was unexpected in the first year, but encouraged in following years. Connections were made between table leaders that led to further collaborations; for instance, a local data journalist was invited to present at a hack night at Carnegie Mellon based on a

---

[125] Western Pennsylvania Regional Data Center and Carnegie Library of Pittsburgh, "Data 101: Data Visualization, Data Literacy, and Storytelling."

connection formed at Data Day. This was an opportunity for those in the civic data and technology community (particularly those on the periphery of those movements) to connect with other parts of the ecosystem, such as with those running the Code for Pittsburgh brigade. However, the connection between professional data participants and community members who led tables was not always easy to form. Some community members felt separate enough that they didn't try very hard to forge connections with other table leaders.

The Impact of Participation on Data Platforms

In addition to the pillars defined above, a key outcome of participation are changes and improvements to the data platform itself. Examples of this are discussed above, particularly in the analysis under pillar four of NDSC's experience of refining published data through collaborating with a food bank on a data story. Outside of this instance of deep engagement, however, the community-based data events show mixed results in changing the design and content of the data platforms themselves.

In Los Angeles, the NDSC trainings conclude with a feedback form, although response rates are inconsistent across trainings. As of March 2019, NDSC trainings have received a total of 127 responses. Some of the feedback and suggestions collected by the forms relate specifically to the trainings themselves. As trainings are malleable and determined largely by the individual trainer who is leading the session, these comments are relatively easy to act upon. This kind of feedback helped motivate

NDSC to create a take-home worksheet to encourage continued use of the platform after trainings were completed.

The feedback survey, along with conversations between training staff and participants, have also inspired considerations about larger changes to the platform's design or the data that is available. This information is shared with NDSC's steering committee when they meet three to four times each year. The steering committee includes representatives from the platform's sponsors and major local partners, such as employees from government agencies, researchers, and non-profit leaders, but does not include any community members. Based on this feedback, the steering committee is able to measure the feasibility and cost of adding new data sets and making modifications to the platform. For instance, one highly requested feature has been for the platform to be released in Spanish, which is currently under consideration by the committee and NDSC's staff. The feedback form provides a small, yet constrained way for participants to affect change upon the platform.

During Pittsburgh's Data Days, on the other hand, WPRDC's open data portal was not a major focus of the events. In fact, staff from the Data Center viewed their roles primarily as organizers rather than presenters. While they described being ready to plug themselves into tables if necessary, they stayed away from doing that as much as possible as the portal wasn't a primary feature of the event. This stands in contrast to the previous iteration of UCSUR's data user conferences, especially the last one in 2015 where their new data portal was presented and feedback was solicited on its design and contents. Now, WPRDC maintains a separate process for user-submitted

feedback on the portal through its website. However, reintegrating the portal as one aspect of future Data Days is something that WPRDC can consider going forward. Doing so would actually bring WPRDC closer to one of the primary goals of DDJC's Data DiscoTechs; those events were run to solicit feedback and ideas in constructing community-based guidelines for the City of Detroit's open data portal.[126]

Applying this Framework to Future Events

In the analysis above, it is clear that any single event will likely be unable to fully address all points in the framework presented in this thesis, but with the proper amount of forethought and planning, can address portions of each pillar. This should be done, first and foremost, with the values and goals of the event in mind. For instance, if the goal is to encourage more data users to visit and utilize an online data portal, as it was in Los Angeles, questions of who should be involved and how to lower barriers of entry for them should enable the goal of increased utilization (and, therefore, may need to segment training sessions into beginners and power users to address their unique needs separately, despite the benefit both groups may have in interacting with one another).

Perhaps the most appropriate way to apply this framework, then, is as inquiries into the goals and values directing the event. Organizers may take each of the considerations nested within each pillar and ask themselves, "how does this relate to

---

[126] Detroit Digital Justice Coalition and Detroit Community Technology Project, "Recommendations for Equitable Open Data."

our goals? Based on that, how might we design our event to incorporate this consideration?" The answers from those questions may be conscious choices that would have been made without the framework; but it is possible that having a fully documented set of considerations could make implicit choices more explicit and thought out.

In order to promote data literacy and awareness more broadly, however, it is important that these types of events (in conjunction with broader in-person data ecosystems that promote community access and technical support) are able to scale. At one level, this can be made possible by publishing resources and materials created for one event online for others to repurpose and reuse, following the lead of the Detroit Community Technology Project and DataBasic. At a greater level, organizers of these events could encourage attendees to recreate these types of events, particularly trainings and opportunities for knowledge creation, for their own communities. This would be especially helpful if the trainings offered by local data intermediaries were already framed to engage a broad audience and could be easily adapted by community organizations. For instance, beyond running platform trainings for community organizations, trainers at NDSC could run data storytelling training workshops, which could both provide useful outputs for their participants (i.e., data stories) and teach them methods that they could share with the community members or service partner organizations that they work with so that NDSC's impact would not be constrained by their ability to run their own trainings.

There is also a critical role that funders play in the longevity and sustainability of these types of efforts. Inclusion of this type of framework in the instructions for community-based work that is being funded may help incentivize or expose grantees to the considerations suggested above. This type of relationship was in fact demonstrated by the Sunlight Foundation. In moving beyond open data policy, Sunlight created a *Guide to Tactical Data Engagement* in 2017, which suggests methods adapted from tactical urbanism and human-centered design to increase the social impact of open data.[127] The concepts advanced in this framework were field-tested thanks to support from Bloomberg Philanthropies with their What Works Cities grantees. The resulting lessons and methods were published this past year as Sunlight Foundation's *Roadmap to Informed Communities*.[128] By embedding this type of framework, which takes into account strategies and philosophies from PPGIS, Data Feminism, Data Activism, and Data and Digital Justice, into requests for proposals and funding for future community-based data events, foundations can encourage local data intermediaries to be more intentional in preparing citizens to understand and share their perspectives as data interpreters in data-driven cities.

---

[127] Sunlight Foundation, "A Guide to Tactical Data Engagement."
[128] Sunlight Foundation, "Roadmap to Informed Communities."

Bibliography

Bhargava, Rahul, Ricardo Kadouaki, Emily Bhargava, Guilherme Castro, and Catherine D'Ignazio. "Data Murals: Using the Arts to Build Data Literacy." *Journal of Community Informatics* 12, no. 3 (2016): 197–216.

boyd, danah, and Kate Crawford. "CRITICAL QUESTIONS FOR BIG DATA: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15, no. 5 (June 2012): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Brown, Greg, and Marketta Kyttä. "Key Issues and Research Priorities for Public Participation GIS (PPGIS): A Synthesis Based on Empirical Research." *Applied Geography* 46 (January 2014): 122–36. https://doi.org/10.1016/j.apgeog.2013.11.004.

Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," n.d., 15.

Cardullo, Paolo, and Rob Kitchin. "Being a 'Citizen' in the Smart City: Up and down the Scaffold of Smart Citizen Participation," May 15, 2017. https://doi.org/10.31235/osf.io/v24jn.

Center for Urban Pedagogy. "Soda Census," 2010. http://welcometocup.org/Projects/CityStudies/SodaCensus.

Certomà, Chiara, and Francesco Rizzi. "Crowdsourcing Processes for Citizen-Driven Governance." In *Citizen Empowerment and Innovation in the Data-Rich City*, edited by Chiara Certomà, Mark Dyer, Lorena Pocatilu, and Francesco Rizzi, 57–77. Springer Tracts in Civil Engineering. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-47904-0_4.

Costanza-Chock, Sasha, Maya Wagoner, Berhan Taye, Caroline Rivas, Chris Schweidler, Georgia Bullen, and the T4SJ Project. "#MoreThanCode: Practitioners Reimagine the Landscape of Technology for Justice and Equity." Research Action Design & Open Technology Institute, 2018. https://morethancode.cc/T4SJ_fullreport_082018_AY_web.pdf.

Currie, Morgan, Britt S Paris, Irene Pasquetto, and Jennifer Pierre. "The Conundrum of Police Officer-Involved Homicides: Counter-Data in Los Angeles County." *Big Data & Society* 3, no. 2 (December 1, 2016): 2053951716663566. https://doi.org/10.1177/2053951716663566.

Dencik, Lina, Arne Hintz, and Jonathan Cable. "Towards Data Justice? The Ambiguity of Anti-Surveillance Resistance in Political Activism." *Big Data & Society* 3, no. 2 (December 2016): 205395171667967. https://doi.org/10.1177/2053951716679678.

Detroit Digital Justice Coalition, and Detroit Community Technology Project. "Recommendations for Equitable Open Data," 2017. https://github.com/datajustice/report/blob/gh-pages/downloads/DataJusticeReport.pdf.

D'Ignazio, Catherine. "Creative Data Literacy: Bridging the Gap between the Data-Haves and Data-Have Nots." *Information Design Journal* 23, no. 1 (July 20, 2017): 6–18. https://doi.org/10.1075/idj.23.1.03dig.

D'Ignazio, Catherine, and Rahul Bhargava. "DataBasic: Design Principles, Tools and Activities for Data Literacy Learners," n.d., 25.

D'Ignazio, Catherine, and Lauren Klein. *Data Feminism*. MIT Press Open, 2018. https://bookbook.pubpub.org/data-feminism.

D'Ignazio, Catherine, and Lauren F Klein. "Feminist Data Visualization." *IEEE Vis Conference,* 2016.

D'Ignazio, Catherine, and Ethan Zuckerman. "Are We Citizen Scientists, Citizen Sensors, or Something Else Entirely? Popular Sensing and Citizenship for the Internet of Things." In *International Handbook of Media Literacy Education*, edited by Belinha S. De Abreu, Paul Mihailidis, Alice Y.L. Lee, Jad Melki, and Julian McDougall. New York, NY: Routledge, 2017.

Doctoroff, Daniel. "Sidewalk Labs | Reimagining Cities from the Internet Up." Sidewalk Labs. Accessed December 11, 2018. https://www.sidewalklabs.com/blog/reimagining-cities-from-the-internet-up/.

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.

Evans, Dean. "How Cloud Computing Is Evolving Street Lighting." Intel. Accessed April 29, 2019. https://www.intel.com/content/www/uk/en/it-management/cloud-analytic-hub/smarter-street-lighting.html.

Flowers, Mike. "Looking for Smoke Alarms in New Orleans." Enigma, April 6, 2015. https://www.enigma.com/blog/results?search=new+orleans&CRAFT_CSRF_TOKEN=nFq8-G-x_F-slx97FJFCI03-5_N1GukvDrQQxvgA4xZUG-1NBLcxffg43oA6wJoP9cFTN3LfEFU_vNCCRGi_aUzCZKm1VYRTHH29GEj-Qg8%3D.

Freire, Paulo. *Pedagogy of the Oppressed: 30th Anniversary Edition*. Bloomsbury Publishing USA, 2014.

Gabrys, Jennifer. "Programming Environments: Environmentality and Citizen Sensing in the Smart City." *Environment and Planning D: Society and Space* 32, no. 1 (February 2014): 30–48. https://doi.org/10.1068/d16812.

Gleeson, Dick, and Mark Dyer. "Manifesto for Collaborative Urbanism." In *Citizen Empowerment and Innovation in the Data-Rich City*, edited by Chiara Certomà, Mark Dyer, Lorena Pocatilu, and Francesco Rizzi, 3–18. Springer Tracts in Civil Engineering. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-47904-0_1.

Goldsmith, Stephen. "Predictive Tools for Public Safety." Data-Smart City Solutions, August 18, 2014. https://datasmart.ash.harvard.edu/news/article/predictive-tools-for-public-safety-506.

Goldsmith, Stephen, and Susan Crawford. *The Responsive City: Engaging Communities Through Data-Smart Governance*. John Wiley & Sons, 2014.

Green, Ben. *The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future*. MIT Press, 2019.

Gurstein, Michael B. "Open Data: Empowering the Empowered or Effective Data Use for Everyone?" *First Monday* 16, no. 2 (January 23, 2011). https://doi.org/10.5210/fm.v16i2.3316.

Haraway, Donna. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14, no. 3 (1988): 575. https://doi.org/10.2307/3178066.

Healy, Patsy. *Urban Complexity and Spatial Strategies: Towards a Relational Planning for Our Times*. London: Routledge, 2006. http://www.tandfonline.com/doi/abs/10.1080/01944360701755584.

Heeks, Richard, and Jaco Renken. "Data Justice for Development: What Would It Mean?" *Information Development* 34, no. 1 (January 2018): 90–102. https://doi.org/10.1177/0266666916678282.

Hendey, Leah, Jake Cowan, G. Thomas Kingsley, and Kathryn L. S. Pettit. "NNIP's Guide to Starting a Local Data Intermediary." Urban Institute, May 27, 2016.

Horgan, Leah, and Paul Dourish. "Ambiguity, Ambivalence, and Activism: Data Organizing Inside the Institution." *Krisis: Journal for Contemporary Philosophy*, Data Activism, no. 1 (2018). https://krisis.eu/ambiguity-ambivalence-and-activism-data-organizing-inside-the-institution/.

Jay, Jonathan. "Can Algorithms Predict House Fires?" Data-Smart City Solutions, March 6, 2017. https://datasmart.ash.harvard.edu/news/article/can-algorithms-predict-house-fires-990.

Johns, Eileen, Jessica Raithel, Maryanne Schretzman, Robert Collinson, Ingrid Gould Ellen, and Davin Reed. "Predicting Homelessness with Individual, Building, and Neighborhood Characteristics." Accessed April 29, 2019. https://www.neighborhoodindicators.org/sites/default/files/publications/local_policy_brief_march.pdf.

Johnson, Jeffrey Alan. "From Open Data to Information Justice." *Ethics and Information Technology* 16, no. 4 (December 2014): 263–74. https://doi.org/10.1007/s10676-014-9351-8.

Kingsley, G. Thomas, Claudia J. Coulton, and Kathryn L. S. Pettit, eds. "Institutional Context." In *Strengthening Communities with Neighborhood Data*, 13–38. Rowman & Littlefield Publishers, Incorporated, 2018.

———, eds. "Introduction to the Field." In *Strengthening Communities with Neighborhood Data*, 1–11. Rowman & Littlefield Publishers, Incorporated, 2018.

Kontokosta, Constantine, Boyeong Hong, and Kristi Korsberg. "Equity in 311 Reporting: Understanding Socio-Spatial Differentials in the Propensity to Complain." *ArXiv:1710.02452 [Cs]*, October 6, 2017. http://arxiv.org/abs/1710.02452.

Laird, Frank N. "Participatory Analysis, Democracy, and Technological Decision Making." *Science, Technology, & Human Values*, no. 3 (1993): 341.

Lewis, Tamika, Seeta Peña Gangadharan, Mariella Saba, and Tawana Petty. "Digital Defense Playbook: Community Power Tools for Reclaiming Data." Detroit: Our Data Bodies, March 11, 2019. https://www.odbproject.org/wp-content/uploads/2019/03/ODB_DDP_HighRes_Single.pdf.

Light, Jennifer S. *From Warfare to Welfare: Defense Intellectuals and Urban Problems in Cold War America*. Baltimore: Johns Hopkins University Press, 2003.

Little, Robert Matthew, and Lynn Froggett. "Making Meaning in Muddy Waters: Representing Complexity through Community Based Storytelling." *Community Development Journal* 45, no. 4 (October 2010): 458–73.

Mattern, Shannon. "Mission Control: A History of the Urban Dashboard." *Places Journal*, March 9, 2015. https://doi.org/10.22269/150309.

McDermott, Patrice. "Building Open Government." *Government Information Quarterly* 27, no. 4 (October 2010): 401–13. https://doi.org/10.1016/j.giq.2010.07.002.

Meng, Amanda, and Carl DiSalvo. "Grassroots Resource Mobilization through Counter-Data Action." *Big Data & Society* 5, no. 2 (July 2018): 205395171879686. https://doi.org/10.1177/2053951718796862.

Meng, Amanda H. "The Social Impact of Open Government Data." Georgia Institute of Technology, 2016. https://smartech.gatech.edu/handle/1853/56292.

Milan, Stefania, and Lonneke van der Velden. "The Alternative Epistemologies of Data Activism." *Digital Culture & Society* 2, no. 2 (January 1, 2016). https://doi.org/10.14361/dcs-2016-0205.

Minkler, Meredith. "Enhancing Data Quality, Relevance, and Use through Community-Based Participatory Research." In *What Counts: Harnessing Data for America's Communities*, edited by Naomi Cytron, Kathryn L. S. Pettit, and G. Thomas Kingsley, 244–58. Federal Reserve Bank Of San Francisco, 2014.

Newcombe, Ted. "The (Hidden) Cost of Open Data," August 2015. https://www.governing.com/columns/tech-talk/gov-open-data-cost-problems.html.

Nucera, Diana J, ed. "Opening Data." *Detroit Digital Justice Coalition Zine*, no. 5 (2015). http://detroitcommunitytech.org/sites/default/files/librarypdfs/opening_data.pdf.

Nucera, Diana J, and Nina Bianchi, eds. "How to DiscoTech." *Detroit Digital Justice Coalition Zine*, no. 4 (2012): 64.

Obermeyer, Nancy J. "The Evolution of Public Participation GIS." *Cartography and Geographic Information Systems* 25, no. 2 (January 1998): 65–66. https://doi.org/10.1559/152304098782594599.

"Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge." Accessed April 27, 2019. https://opendefinition.org/od/2.1/en/.

Petty, Tawana, Mariella Saba, Tamika Lewis, Seeta Peña Gangadharan, and Virginia Eubanks. "Out Data Bodies: Reclaiming Our Data," June 15, 2018, 37.

Poon, Linda. "Giving Cities Room to Experiment and Innovate." CityLab, January 12, 2017. http://www.citylab.com/cityfixer/2017/01/giving-cities-room-to-experiment-and-innovate/512909/.

Raji, Inioluwa Deborah, and Joy Buolamwini. "Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products," n.d., 7.

Rawson, Katie, and Trevor Muñoz. "Against Cleaning," July 6, 2016. http://www.curatingmenus.org/articles/against-cleaning/.

Schlossberg, Marc, and Elliot Shuford. "Delineating 'Public' and 'Participation' in PPGIS" 16, no. 2 (2005): 12.

Schrock, Andrew R. "Civic Hacking as Data Activism and Advocacy: A History from Publicity to Open Government Data." *New Media & Society* 18, no. 4 (April 2016): 581–99. https://doi.org/10.1177/1461444816629469.

Schuurman, Nadine. "Trouble in the Heartland: GIS and Its Critics in the 1990s." *Progress in Human Geography* 24, no. 4 (2000): 569–590.

Scott, James C. *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press, 1998.

Segel, E, and J Heer. "Narrative Visualization: Telling Stories with Data." *IEEE Transactions on Visualization and Computer Graphics* 16, no. 6 (November 2010): 1139–48. https://doi.org/10.1109/TVCG.2010.179.

Shaw, Emily. "Making the Most of Open Data." In *What Counts: Harnessing Data for America's Communities*, edited by Naomi Cytron, Kathryn L. S. Pettit, and G. Thomas Kingsley, 108–22. Federal Reserve Bank Of San Francisco, 2014.

Sieber, Renee. "Public Participation Geographic Information Systems: A Literature Review and Framework." *Annals of the Association of American Geographers* 96, no. 3 (September 2006): 491–507. https://doi.org/10.1111/j.1467-8306.2006.00702.x.

Sunlight Foundation. "A Guide to Tactical Data Engagement," September 2017. https://sunlight-foundation.gitbooks.io/tactical-data-engagement/content/.

———. "Open Data Policy Guidelines." Sunlight Foundation, March 2014. https://sunlightfoundation.com/opendataguidelines/.

———. "Open Data Policy Hub | Why Open Data?," August 2018. https://opendatapolicyhub.sunlightfoundation.com/why-open-data/.

———. "Roadmap to Informed Communities," December 2018. https://communities.sunlightfoundation.com/.

Taylor, Linnet. "What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally." *Big Data & Society* 4, no. 2 (December 1, 2017): 2053951717736335. https://doi.org/10.1177/2053951717736335.

Townsend, Anthony M. *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. W. W. Norton & Company, 2013.

Wasacz, Walter. "How Motor City Mapping Became a Template for Citywide Collaboration." Model D, February 13, 2015. http://www.modeldmedia.com/features/DRF4MotorCityMapping021315.aspx.

Western Pennsylvania Regional Data Center, and Carnegie Library of Pittsburgh. "Data 101: Data Visualization, Data Literacy, and Storytelling." Data & Tech Training. National Neighborhood Indicators Partnerhip, 2018. https://www.neighborhoodindicators.org/data-tech/course-catalog/data-101-data-visualization-data-literacy-and-storytelling.

White House. "Memorandum on Transparency and Open Government." whitehouse.gov, January 21, 2009. https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government.

Williams, Sarah. *Data Action*. MIT Press, 2020.

———. "Data Visualizations Break Down Knowledge Barriers in Public Engagement." In *The Civic Media Reader*, edited by Eric Gordon and Paul Mihailidis. The MIT Press, 2016. https://doi.org/10.7551/mitpress/9970.003.0013.

Wilson, Matthew. "Towards a Genealogy of Qualitative GIS." In *Qualitative GIS: A Mixed Methods Approach*. Sage, 2009.