PUBLIC
PUBLIC

INTEREST
INTEREST

TECHNOLOGY
TECHNOLOGY

BY
MARIA FILIPPELLI

PUBLIC INTEREST TECHNOLOGY

By Maria Filippelli

There's a recurring conversation I have at every organization where I've ever worked. It begins with a colleague approaching me with a technical issue. It could be that a website is down or a product is glitchy — basically, something is not functioning as expected.

So I troubleshoot with them. I ask what was supposed to happen, like a website being available 24/7. Then I ask if it's happened before. If so, the next question is about the resolution, with the hope that we could repeat that solution. If not, the conversation shifts to me asking about the kind of maintenance in the vendor contract and the contact person listed. The response is often, "I don't know." When I ask to see the contract, service and maintenance terms are minimal if they're listed at all.

Sometimes I'm approached about something a little more personal, like an email regarding a data breach from one of a person's accounts (pick a retail store, hotel chain, or any other place you've had a digital interaction). Unfortunately, it's near impossible to do anything after the fact. The next best thing is to use it as an allegory for better personal and organizational cybersecurity. Especially in the past few years when personal and work devices have become so entwined.

In either situation, the conversation is happening after the fact. As a Public Interest Technology (PIT) practitioner, I often balance fixing problems after something breaks while designing products, policies, and practices that minimize the chances of them happening in the future. PIT practitioners often advise on technologies, reconcile new technologies with legacy technologies, and take the larger technical infrastructure into account while building solutions.

The "public interest" piece of PIT can be an elusive term to define, especially when it comes to technology. At its core, PIT reduces barriers to entry for public services — both for folks accessing that service as well as those operating the service. The datafied state[1] proposes helpful questions about facets of public interest, like procurement and accountability, which have long been parts of my work. However, it's important to note that technologies can also create barriers to entry, like requiring reliable internet access, so sometimes the right tech solution is not tech at all. Whatever the project, PIT involves understanding the problem you're trying to solve and extensive outreach, collaboration, and consensus building.

Throughout this paper, I illustrate some common scenarios at the intersection of the datafied state and PIT, like working with an outdated procurement process and interpreting highly technical concepts for a general audience while balancing various tensions such as timelines, stakeholder expectations, and policies. From a multimodal trip planner to the 2020 Census, I have worked with all levels of government and on every part of a technical project from design to evaluation. The issues are sometimes with the technology, sometimes with the process (outdated procurement language), and sometimes with perception (Are past failures indicative of future ones?). I include some tried and true lessons (asking questions is a

1 "The Datafied State," https://points.datasociety.net/the-datafied-state-a2a7101ba573.

good thing), as well as some topics that are still evolving and will continue to unfold in the near and far future (tensions between highly accurate data and a right to privacy).

An "Open Scope of Work" is Not in the Public Interest

The two guys across the table were not ready to present, fumbling with cords and making jokes the entire time. My fellow interviewers laughed along with faces full of anticipation. I did not feel the same. I couldn't understand why anyone would have such a cavalier attitude on a multi-million dollar contract. But how could they know how much work I put into getting to this point? Hundreds of hours with experts to update the procurement process, with my team drafting requirements, and with partner agencies envisioning "what could be" led to this day.

What got us sitting around the table started several years prior to that day. I worked for a metropolitan planning organization that focused on transportation planning and execution for the region. The organization had long managed commuter programs to reduce peak time roadway congestion. We managed a regional vanpool program with hundreds of vans, an employer-based carpool program, bike lockers, and a few other programs to support commuters. For nearly 20 years these programs were managed by paper and, eventually, a combination of paper and local databases. We were faced with a very common problem in the public sector — how do we digitize a paper-based program making it (in theory) easier for commuters (the public) to access and program leads to manage?

The advent of smartphones and big data had us wondering what was possible. Could we build an application that included all of our programs in one place? This was also before any of the big mapping services had multiple transportation modes available; they were in the vehicle traffic time and sometimes public transportation phase. There was an opportunity to build a new and needed solution.

The first step in the process is research and landscape analysis. We wanted to know which metro areas had something similar to what we wanted, who was thinking about making this happen, and what vendors were in this space. This early in the process, our project was not funded and had no open requests for proposals (RFPs), so we could converse with vendors without any conflict of interest.

Though it was not a crowded space, a few tech vendors were in the multimodal commuter space. They, too, saw the potential in the increased use of smartphone applications and maps. The conversations, akin to an informational interview, followed a similar pattern. I explained what we were trying to accomplish and the vendor would let me know what pieces were developed, what parts were under development, and what parts they couldn't (or wouldn't) do. Through the vendor's limitations, I uncovered what tech concealed and why it is so challenging to protect public interest.

Tech companies want to own their source code, the unique documentation that undergirds their tech products. Their ownership makes their tech proprietary, competitive, and profitable. In addition, I discovered that tech companies prefer little customization or bespoke products, if any. They are thinking of scale — "If we do this in one metro area, how can we replicate it for others?" While bespoke products are an option, they are often too expensive for the public sector. The weeds of the customization part of the

conversation is usually when I heard, “If you just give us a general idea of what you want, we can figure out the details,” or “We like an open scope of work,” or “We don’t like too many specifications.” Vendors that fail to design what works best for your organization and those not willing to have a conversation and collaborate are, by default, creating your tech policy: a policy that likely benefits the tech company more than it benefits the public interest it is intended to serve.

Once we conclude the research, a procurement can be drafted. Procurements include heavy documentation, including scopes of work, budgets, and timelines. Government procurements, at this time and place, were designed where the government owned everything procured. However, this type of procurement would not work for tech and especially tech start-ups, as they would not give up ownership of their source code, and owning source code is not a position governments necessarily want or need. To own source code requires staff to manage it; often licensing a tech product is the correct fit. So contract language needs to be flexible, and in my case — updated.

The process of updating our contracts so that we could license tech products and create customized off-the-shelf products with vendors required … another contract. We needed an expert in intellectual property to help us identify needed changes in the contract and procurement language. The process proved successful as it yielded us terms of ownership over data but not code and set up the organization for all tech products moving forward.

With the research and updated contract language completed, I began drafting the specifications. To date, this is the only time in my career I ever drafted a full technical specification document with purpose, needs, requirements, functionality, and behaviors that we wanted from our tech product. It

is a highly collaborative and detailed process, requiring conversations with anyone who will interact with the product. It's one of those things so foundational to creating tech that it should be mandatory for anyone wanting to work in tech or PIT.

After the RFP is made public and bids are received, they are reviewed and scored. Then, there is an interview for finalists — that led to the conversation with the two joking guys representing their particular tech company. They asked about the usual things, such as flexibility in timelines and budgets, and then asked about the technical specifications. Did they really need to adhere to all of them? Yes. While I felt confident not owning source code, providing a sub-par product to the public is unacceptable.

Everyone is a Stakeholder in PIT

Fast forward nearly a decade from that first procurement process and I'm standing in the hallowed halls of Congress for a day of education and advocacy regarding the 2020 Census. A colleague, from a less technical background, also there to provide census education to advocates and congressional staff asked me, "Do you think you know more about the census technology than we do?" It struck me in two ways. One, I had been hired specifically because I work at the intersection of technology, data, and policy (all intrinsically intertwined, especially in the public sector). Two, I was viewed as an outsider, challenged on my tech, policy, or advocacy chops — I wasn't sure. But I knew there were long-standing, unresolved tensions between DC tech advocates and Silicon Valley … and that comment let me know I was somewhere in the middle.

In the years between my first tech procurement and that fall day in 2018, the world became more digital: data brokers had more information about individuals, the harms of social media were known, and lack of accountability for tech firms persisted.

The 2020 Census underwent a number of updates from the 2010 Census. It would be conducted largely online, so the design of the online form and cybersecurity were two of the biggest concerns. It was also administered in the age of mis- and disinformation (the Cambridge Analytica[2] scandal broke in the spring of 2018). I expected external challenges like limited information from the Census Bureau and apathy from social media companies. The internal challenges were unforeseen.

My work was cut out for me. Not only was there new technology to understand and inform folks about, but it was happening in a crowded space. A network of hundreds of organizations, thousands of stakeholders and volunteers, and many others supported the Census Bureau's work and ensured that the Census Bureau could count everyone "once, only once, and in the right place."[3] There were voices just trying to get the word out about the importance of being counted, advocates trying to ensure the census had proper funding, a presidential election, and a global pandemic vying to shut it all down.

The confluence of all this technology and its implications — census data is used for apportionment, redistricting, and federal funding programs (and that's just the beginning) — required that I learn fast and adapt my approach for different stakeholders. Functioning as a bridge is the best way to describe my work. I bridged technical documentation with everyday language, concerns with facts, and the Census Bureau with stakeholders. The comments about who I spoke with and how I approached my work continued, but the importance of the census outweighed all of that noise.

2 Facebook-Cambridge Analytica Data Scandal, Wikipedia, https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

3 "Counting Everyone Once, Only Once and in the Right Place," Census Bureau Blog Post, November 5, 2018, https://points.datasociety.net/the-datafied-state-a2a7101ba573.

Digitizing Government Forms

Prior to the 2020 Census, the last time the federal government had tried to launch a nationwide website was healthcare.gov in 2013 — it was considered an epic fail.[4] The questions I heard about the census was not *if* it would fail, but what we should do when "the inevitable" happened and folks couldn't get online to fill out their form. The fear was that people would try once, and if they couldn't get online, they wouldn't come back, thus leading to an undercount of the population and communities missing out on critical funding, among other things.

The information the Census Bureau provided to debunk the concerns of failure was slim. I was able to find load capacity reports[5] and not much else. All I could tell people was that the system was performing as expected with some flags, and it would be ready on time. The other piece of information I tried to scrape together was what browsers and devices were used for the testing. The response, a fairly standard "the most up-to-date browsers and devices," was actually insufficient. The assumption that everyone in the US has the most updated phone, laptop, or tablet is incorrect. In fact, I found out many lower-income households used devices outside of the tests.

Beyond that were questions of what the form would look like and how the experience would be. For example, "Can you start the form and then come back and finish it later?" I understood and felt the stress — how could I explain something unseen? It turned out the Census Bureau was keeping the UI/UX under wraps as a matter of security. I argued that not showing people what the form looked like was in itself a security risk. If you don't know what to expect you're more susceptible to fakes. Shortly before the form launched, the Census Bureau agreed to a demo of the online form. I took screenshots and quickly made resource guides for multiple audiences.

4 "The Failed Launch of www.HealthCare.gov," November 18, 2016, https://d3.harvard.edu/platform-rctom/submission/the-failed-launch-of-www-healthcare-gov/.

5 Load capacity tests how many people can access a system at one time before it fails.

Back to the load capacity tests. The system under development started failing their capacity tests near the launch date. In perhaps a lesser-known, but rather extraordinary, story about the 2020 Census is that a backup system[6] was built internally. Originally designed as a backup only, it became the main system that collected the majority of census responses. An excellent PIT lesson: redundancy is resiliency.

Reconciling Data, Technology, and Policy

In December 2019, at a National Academies workshop on the 2020 Census Disclosure Avoidance System (DAS)[7], I watched a presenter grip the lectern and adamantly argue that highly accurate data at low levels of geography should be released publicly. The presenter was concerned that the new DAS would negatively affect census data quality. But highly accurate data at low levels of geography makes for easy database reconstruction, leading to the re-identification of individuals. And the Census Bureau knowingly releasing our individual information is prohibited by Title 13, US Code.[8] How that law is interpreted and executed was cause for much debate leading up to 2020, and it remains a hot button issue today.

The Census Bureau does not release data in a digital or data vacuum. The proliferation of digital tools and low cost of computing power available globally requires consideration of the larger data and technology digital space. The Census Bureau understood this challenge and developed a comprehensive mathematical DAS, unlike any previous decade. Differential privacy, the new DAS framework, became the focus rather than discussing what individual privacy means in this current environment and how we should balance different stakeholder needs.

6 "2020 Census: Initial Enumeration Underway but Readiness for Upcoming Operations Is Mixed," GAO, February 12, 2020, https://www.gao.gov/products/gao-20-368r.

7 "2020 Census Data Products," National Academy of Sciences, accessed December 12, 2023, https://www.nationalacademies.org/our-work/2020-census-data-products-a-workshop.

8 "Title 13 - Protection of Confidential Information," US Census Bureau, https://www.census.gov/about/policies/privacy/data_stewardship/title_13_-_protection_of_confidential_information.html.

Quick-changing technologies rarely align well with slower-moving policymaking, especially in cases where we must protect an entire nation's data. However, anytime we release data it enters a global digital space. Layering mathematical and cybersecurity techniques, in addition to developing policies, are tools in the hands of all PIT practitioners that can ensure our right to privacy.

The Digital Ecosystem

The technology used to collect census responses went beyond what the Census Bureau built. While the Census Bureau was working on their internal systems, advocates and other stakeholders were gearing up to help with outreach, engagement, and to fill gaps in the digital divide. Local governments and advocates employed technologies to help create awareness about the importance of participation. They provided tablets, delivered hot spots to areas with poor internet service, and used text messaging services to send informational links and reminders to respond.

In one meeting about the use of these technologies, I confronted the idea of being tech agnostic. The conversation started by probing if additional technologies might be helpful. Then it flowed to what type of technologies should be put into use. Then someone said, "I don't care which technologies we use, I'm tech agnostic." At this point, I raised my hand and pointed out that perhaps we should not be tech agnostic, as that was risky if we're not considering how a tech company is collecting and sharing an individual's data or if we know their cybersecurity setup (the things we were demanding of the Census Bureau).

The response was a very frustrated "We can't afford to be picky about the technology we use right now." And I could not disagree more. We cannot afford *not* to be picky about the technologies we use, basically ever. Tech products collect data, and once our data is handed over to a product, whether through profile setups, cookies, or other social engineering, we have very little recourse when that data is mishandled or sold to data brokers. Sometimes PIT practitioners have to dole out the tough love and ask the hard questions, and we're not always given thanks for it. It is a fine line to walk, can feel awful, but we can't afford not to do it.

The Future is Even More Complex

As I write this paper, there are two big conversations happening at the intersection of data, technology, and policy that make the understanding of the Datafied State and PIT more complex and urgent.

One is that the federal government is making significant changes to the demographic data it collects across all agencies and departments, not just through the census. Specifically, the US Office of Management and Budget, which coordinates the federal statistical system, is reviewing and developing recommendations to expand the federal race and ethnicity categories currently utilized by the government. In addition, the National Science and Technology Council published the Federal Agenda on LGBTQI+ Equity[9] to assist federal agencies in creating evidence-based policy for the LGBTQI+ communities. Combined, these changes will better reflect the diversity of our nation in our nation's data.

There are requests from census stakeholders to release this data, disaggregated and at the lowest levels of geography possible, to better understand

9 "Federal Agenda on LGBTQI+ Equity," January 2023, https://www.whitehouse.gov/wp-content/uploads/2023/01/Federal-Evidence-Agenda-on-LGBTQI-Equity.pdf

economic and health disparities, among other issues. Simultaneously there is a scramble to understand the impact of generative AI, and regulate AI in general.

These conversations are not happening together.

The increased detail we provide on federal forms, like the 2030 Census, is necessary to ensure equitable distribution of federal funds and enforce civil rights laws, among other issues vital to our health and democracy. While the details of these changes are finalized (expected in summer 2024[10]), the proliferation of generative AI tools continues. AI pulls from a wide range of data sources[11] with little, if any, discrepancy. We need to be careful with detailed data publications, or we run the risk of re-identifying individuals, leaving us susceptible to known and unknown harms.

Remember census DAS? The Census Bureau is still working with stakeholders and conducting a participatory algorithmic design process to balance between quality data and privacy protections. And the procurement process changes? Those can happen anywhere at any time. These processes, policies, and practices reduce and limit harm when implemented in the public interest. Being more intentional with our technologies will limit the harm to ourselves and others.

When I run into former colleagues, they often tell me about something they learned from our time together. It may be that they didn't give their personal data to an app, or they included maintenance in a tech contract — something that lets me know the tensions and hard conversations about PIT and the datafied state are worth it.

10 "Frequently Asked Questions," OMB Interagency Technical Working Group on Race and Ethnicity Standards, https://spd15revision.gov/content/spd15revision/en/faqs.html.

11 "Higher Education and Generative AI: Evolving Lessons from the Field," New America, April 20, 2023, https://www.newamerica.org/pit/events/online-what-chatgpt-tells-us-about-the-future-of-ai/.