

## Transcript

Title: Webinar - Social Data in Action: Social Data for Public Empowerment in the USA

Creator: Prof Jane Farmer, A/Prof Anthony McCosker, A/Prof Sarah Williams

Year: 2021

Audio/video for this transcript available from: <http://commons.swinburne.edu.au>



JANE FARMER: Welcome, everybody. Conscious that people still keep coming in as we start here, but thanks very much for being here for our third webinar in the social data and action series. And today, we are super excited to have Sarah Williams from MIT with us today. And her talk is going to be about social data for public empowerment in the US, possibly other places since there's a number of sites in there.

So Sarah is an associate professor of technology and urban planning at the Massachusetts Institute of Technology, where she is also director of the Civic Data Design Lab at the Leventhal Centre for Advanced Urbanism. Sarah combines her training in computation and design to create communication strategies that expose urban policy issues to broad audiences and create civic change.

She calls this process data action, which is also the name of her recent book published by MIT Press, which is truly awesome. And she's going to draw on some of that today. So thank you for being here, Sarah. It's evening time in New York, where Sarah is. So obviously it's early morning for us, so people are dribbling in. Paul, next slide.

So I'd just like to acknowledge that I'm hosting this webinar from the lands of the Wurundjeri people of Kulin nation and acknowledge all of the traditional custodians of the various lands of which you're all working today and present today and the Aboriginal Torres Strait Islander people participating in this webinar. And I want to pay my respects to elders past, present, and emerging and celebrate the diversity of Aboriginal peoples and their ongoing cultures and connections to the lands and waters.

OK, so Sarah is going to speak for a while, maybe up to half an hour or so. We encourage you to put questions or comments into the chat as you go, but we will leave all of the questions and handling of these questions in chat till the end. At that point, we might ask you to ask your question, or you might want to raise your hand and ask a question spontaneously. We will be recording this session. If you have any challenges with being recorded, then please get in touch with Paul who's on here, and [paullavey@swinburne.edu.au](mailto:paullavey@swinburne.edu.au).

So thank you very much. I'm now going to hand over to Sarah.

SARAH WILLIAMS: Great. Thanks so much for that wonderful introduction. And also, thank you for inviting me to join you from way far away, I can imagine. So I feel very lucky to be able to join you all, or the pandemic has created this lucky situation so that I can speak with you today. So thanks so much for the invitation. Let me just share my screen.

Just going to organise it so that I can see my notes. Hi, everyone. I'm an associate professor of technology and urban planning at the Massachusetts Institute of Technology, where I also direct the Civic Data Design Lab and the Leventhal Centre for Advanced Urbanism.

The LCAU is a cross disciplinary Centre, which combines the research of architecture, urban planning, landscape architecture, and systems thinking, not about the problems of yesterday, but of tomorrow. We are motivated by radical changes in our environment and the role that design and research can play in addressing these.

I myself am very interdisciplinary. I combined training as a geographer, architect, urban planner, and data scientist. And I really combine these skills to take data into action. I believe multidisciplinary teams are essential for using data for action, as they help contextualise the work within the broader policy arena, communicate its results more broadly, and allow for new innovation. These teams bring their diverse training to the table.

Here is an example of a team I brought together from work I did in Nairobi, Kenya. We included computer scientists, architects, policy experts, data scientists. People from NGOs, such as the World Bank are also included. And I think it's really impactful what we can do together.

So what do I mean by data for action? I'm going to use an example of a project, which is now close to 15 years old, but had such a great impact, a project which I did with Laura Kurgan, Eric Cadora, David Frankfurt, and myself, who was a data scientist at the time.

In this project, we took data from the criminal justice system of where people lived before they went to prison. We added it up block by block, and those red blocks show where costs over a million dollars to incarcerate people. When you zoom in, you find whole communities such as this one where \$17 million is spent to incarcerate people.

How could this money be spent to alleviate the systematic reasons for mass incarceration was the question we all asked ourselves. The maps expose our response to poverty is incarceration, rather than prevention. We took these maps to the Centre for Architecture in New York, where we also showed maps of 10 other cities.

We asked the community, policy experts, urban planners, criminal justice advocates, if they had just one million of that dollar, how might they reinvest it in the community. So the idea here is if we spent just a million dollars on better education or better health care, it might alleviate the systematic reasons that people enter the prison system to begin with.

The images were seen in the architecture league by the Museum of Modern Arts curator, who asked to include them in an exhibition. They were seen on the walls by a congressman, who asked if he could use them for evidence in the Criminal Justice Reinvestment Act of 2010 and 2020, which allocated \$25 million to reentry programme.

And while this is just a drop in the bucket, if we look at that \$17 million, it shows how by bringing data to broad groups of people, we can change minds and hopefully policy. This is a great example of the data action methodology because communicating with data in this way requires the ability to

ask the right questions, find or collect the appropriate data, analyse and interpret that data, and visualise the results in a way that can be understood by broad audiences.

Combining these methods transforms data from a simple point on a map to a narrative that has meaning. Data is not often processed in this way because data analysts are often not familiar with the techniques that can be used to tell stories with data ethically and responsibly. And my book seeks to do that.

I want to just note that one of the reasons I use this project is also because even though it's close to 15 years old, it's been coming up on my Instagram and news feeds quite often in the conversations in the United States around defunding the police, much so because the message in a way was about that, how do we reallocate fiscal funding to things that can help alleviate the systemic issues, rather than focus on policing.

So data action isn't just my methodology. It's also my recent book that was released in December. And data action was really started with the premise that in the last decade, we've created unprecedented amounts of data, which is only set to increase by tenfold, largely devoured by private companies, such as this data set from Google's autonomous vehicles, which creates points every second as you can see here.

And while not all the data is closed, like the previous example, all the excitement from civic hackers, two consulting groups, such as McKinsey, left me with a pain on my side. Did these groups know the ways in which data intentionally or unintentionally can cause harm. Do they know how their excitement to do good might be marginalising or reinforce or forget whole populations?

And while criticism began to mount from academics about the potential misuse of data, as you can see in these examples above, and rightfully so, I found we needed a way to show people how to use data ethically and responsibly. I found we needed to explain to data enthusiasts, my MIT students alike, the ways in which we could use data for empowerment rather than oppression.

So I decided to develop data action, which I believe presents a corrective standard to data practises by acknowledging that data represents the ideologies of those who control its use. Data action at its heart was developed for data enthusiasts and novices, civic hackers, urban planners, and policy experts to provide guidance for using data ethically and responsibly. Data action asks evidence of big data to rethink how they work with data to make the process more responsible to the people their work affects.

So in the first chapter of Data Action, "Big Data for Cities is Not New," I position the reader within this larger narrative by providing a historical account of the ways those in power have used data to shape our cities, how they have reinforced structural racism, how data has helped enforce the marginalisation of various populations. At the same time, I illustrate the many benefits that analysing data has provided societies, such as establishing social services, and stopping the spread of disease.

The juxtaposition of these two outcomes provides a reminder that the person analysing the data defines its use. And so in this chapter, one of the many things I talk about is gerrymandering. And I talk about a character in the United States called Thomas Hofeller, who has pushed around data to

help create Republican districts in the US and is often considered the reason that the Republican Party has been so dominant in the past years.

A self-professed data nerd, he started his work since the first digital census came out in 1970 and has systematically used it to push around congressional districts towards the Republican favour. So one of the things that I mentioned that he does in the book is he creates in the '80s all Black congressional districts under the auspices of civil rights, saying that African-Americans should have their own districts. But what this actually did is consolidate Blacks into districts, decreasing their vote and their ability to make more diverse districts, one of the main reasons that we see a heightened amount of Republican districts across the United States.

A recent discovery of Hofeller's personal notes from his computer confirmed how he used data for the Republican advantage, redistricting and creating districts that he hoped would subvert the margins. In this chapter, I talk about characters like Hofeller, but I also talk about characters from public health that really helped improve society using data as evidence.

It's not surprising given all the hype swirling around massive amounts of data in today's world that some essential data is still missing. This missing data usually represents the interests and needs of those on the margins of society, but it also represents topics that government and companies seek to keep hidden. Chapter two, "Build It!" challenges those communities and data specialists alike to create the data needed to encourage policy action.

The examples in "Build It!" show that today anyone can go about collecting data with very little training. This is because innovations in digital technology make it easier than ever to collect personal data. From our mobile devices, to environmental sensors in our home, modern life is full of tools to measure aspects of our lives. And it's up to us to understand the data we need, how to use these technologies to develop narratives that can make an impact on policy.

In this chapter, I show a project in Beijing. And this is a project that I did during the Olympics with the Associated Press to expose the extremely poor air quality, which is now famous in Beijing. And it might be surprising to you to know that there were no air quality measurements being given out by the government during the Olympics. And we developed a sensor, which was the only tool that measured air quality during the Olympic games.

Here you can see the red represents Beijing, and the measurements of particulate matter are far above the World Health Organisation for a developing country and far above what you see in New York and London on the same day. The Associated Press translated these into data visualisations that they sold to their subscribers, and the information was also included in The New York Times. And really, the idea of this project was to show the impact of air quality and use the Olympics as a global stage to help bring this issue to broad groups of people that might not have been aware of it.

Although data collection is important, this chapter also explains how building data together strengthens communities around their shared interests. Building sensors, learning about data measurements, and collaborating with one another creates the bonds necessary to create change. "Build It!" inspires us to create our own community data collection projects, not only because they can literally change our community, but also because they can help create new communities, like

this one which was created in Louisiana during the oil spill to make sure the government did the proper clean up.

While "Build It!" describes how to collect missing data and use it to fill gaps in knowledge, "Hack It" shows that oftentimes data already exists openly, although we might not see it that way because it's stored, maintained by private companies. In some countries where data is tightly controlled by the government who produces it, privately owned data is the only data that exists openly for analysing the dynamics of our lives.

This third chapter argues that we ought to be creative in the ways that we obtain data to answer important questions about society, nonetheless acknowledging that data acquired for one purpose and applied towards another holds numerous ethical concerns that must be considered. Data action provides guidance for those who seek to use data openly.

So in this project, we look at the-- or in this chapter, one of the examples is looking at the issue of ghost cities. If you're not familiar with ghost cities, they are developments that lie vacant. And mapping vacant residential developments can identify risk in the Chinese real estate market.

But data about vacant developments are not available through the government, are not widely available at all, in fact, are really important to planning but also to the global marketplace to assess the risk. So we decided to create a model that tried to identify where these vacant development exists.

And our model is based on the idea that a thriving community needs amenities, schools, places to go grocery shopping, banks, entertainment. And if we are to have that, we might have a thriving community.

Given that data is hard to come by, we scrape data from the Chinese Yelp or DianPing And then we got residential locations from a AMAP and Baidu's API. We overlaid a grid of 300 by 300 cells to mark residential locations. And then we took the Centre of each cell and measured its distance to the various amenities.

We took into account whether an amenity had a review to determine whether that's something that people actually went to. And then once we had this, we applied the Hansen's gravitational model that measures urban accessibility, which takes into account that if we live in the suburbs, we would travel further to get to an amenity.

We find the mean of the fitted distribution, and then we remove the cells that are above the mean. And then we perform spatial autocorrelation on those highest amenities scores, thinking that those that cluster are likely to be ghost cities. And then one of my methods is I really think you need to ground truth your data so we went to Chengdu, Tianjin, and Shenyang to see how well we did.

And we found things like this. Developments that were built maybe 5 to 10 years ago in which the government would say people will eventually move there. The idea is that the government thinks it's better to have the economic growth that comes with the construction and allow them to lie vacant until people will move in.

We also found a number of stalled construction sites next to partially occupied housing. In this particular case, the occupation was by those who previously lived in the village in which these housing were redeveloped. We found older housing blocks, communist housing blocks that were being readied for removal and redevelopment. And then we found whole cities that lie vacant, such as this one. And this was meant to be the Science Centre.

I should note that some of these vacant developments lie vacant but completely sold because Chinese citizens often buy four or five homes. This is encouraged by the government and was a tool for investment before the Chinese stock market existed. So in the first example, those homes might have been completely sold but lie vacant, a kind of thing that China has kind of replicated in many places.

So we decided to create a data visualisation that helped to explain the model. So when you click on each cell, it tells you how that cell's amenity score was created. So here, there's great distance from malls, from schools, and from, banks which is one of the reasons that this particular location had a high amenity score. And we showed this map to planners. And what was really important about this is they became much more forthcoming about the reasons for these developments. And we were able to create trust through them. Vacant developments are controversial to local politics and that the decisions were based mostly on theory and not open data says these planning directors, but the real estate developers also told us that the bubble of real estate in China might have irreversible impacts on residents and that the mismatch between supply and demand would be a big problem looming ahead.

And I think that this kind of verification really helped us to create what I would consider the next foreclosure map, showing us where there's risk in the market and the potential for houses to be foreclosed upon. And as I said, one of the things that we like to do in the Civic Data Design Lab is really bring our projects to new audiences. So we brought this project to the [INAUDIBLE] biannual, where we presented the data visualisations linked with drone footage that we took on site to allow more people to understand the phenomenon that has now been exported many places across the world.

While "Hack It" asks us to innovatively find, acquire, and analyse data for policy change by communicating the results of those analytics ingenuously, chapter 4, "Share It!" stresses the importance of sharing data both in its raw form and also through visualisations. Sharing data helps the public have access to information, acquire knowledge, and ultimately make better civic decisions. Sharing data through visualisations can communicate the insights of data without asking everyone to be data scientists.

So in this project, we use the example of Nairobi. And in Nairobi, matatus or buses you see here are the main form of transit for 3.5 million people. Yet when I started working in Nairobi, there was not information about where these buses went, not even a map. This data is essential for planning and creating transportation models but also can leverage to develop applications that can help decrease congestion that plagues the city as exhibited here. While matatus are essential for the operation of the city, the government had no data about where they were.

This is very familiar to me. I started working in Nairobi by creating the first GIS data set using machine learning from satellite data. And so for this project, we really thought, how could we create a data set for our model but a data set that everyone could use. And we thought could we leverage the ubiquitous nature cell phone use in Nairobi to try to capture that data about an informal transit system, which most citizens depend upon and open that data for anyone to use and build upon.

So our research team collaborated with the University Nairobi Computer Science Department to develop an application that collected data on the routes and stops. Our collaboration with the university was important because we wanted the local knowledge on how applications are used. But more importantly, we wanted the knowledge of how to build these systems to stay in Nairobi. The collaboration of skills allowed for transfer of knowledge between cultures.

We collected the data using the GTFS data standard, which is the same standard that Google Maps used to route you on public transport. It's a simple format that holds latitude and longitude data, which draws the network of matatu system as volunteers collect the data class. And you can see here it's building the road network in Nairobi.

Developing a visualisation that would allow the public to understand the complexity of the system became a challenge, and we began to play with the data to determine ways to visualise the information. We began by giving each quarter a colour and legend. Still, it was still hard to see the multiple overlapping routes. Ultimately, we decided to create a stylized map, much the same that you would see in New York, London, Paris, even Boston.

And the map was edited with matatu drivers, owners, the government, and collaborators. And the collaborative way that we created this data really helped to build trust in the data set. And here you can see them planning in the north where you see there's not a lot of routes, and they're discussing why that is and instantly using the map as a planning tool.

We held workshops to get a sense of users would be able to read the maps. And we performed community engagement with the matatu drivers, owners, and riders. Ultimately the maps went viral on the internet and social media. Newspapers published them as inserts so that people without phones could have access to them.

And one of the things that I like to talk about when I talk about that project is how do you measure success in an open data project? And I think that's when others leverage the data we created to generate their own policy change. And so we were really excited when the government invited us to a press conference to make the map the official map of this city.

And what I think is important about this is while they were largely disinterested all along the way, when they saw the benefit of the map, they were included in the data collection process, they felt like they could trust it, so much so it is now the official map of the city. It was the first informal transit navigable on Google Maps, and we have now have 40 other cities do the same.

Because the map became an iconic symbol of the city, the World Bank copied our visualisation to create their bus rapid transit map. They wanted to piggyback on the popularity of our visualisation. We have worked together with the local technology community, and there are now five apps that use our data as a base.

Semiformal transit provides mobility around the world, not just in Nairobi, but the majority of cities outside the US and Europe use these systems. And many of these cities have followed our lead. From Amman to Managua, we now have a network of many cities that we've helped develop this data.

Chapter 5, "Data as a Public Good." I believe it's important to think of data as a public good, a non rivalrous commodity that can be valued by all who consume it, a commodity similar to electricity, which needs regulation so it can be used equitably by the public. Society should work with private companies to find ways for them to share the data ethically and responsibly so that we can use data towards the betterment of society.

In this chapter, I discuss how we are actually missing a lot of data, and Africa is a great example of that, so much so that the World Bank has said it could lead to the denial of basic rights and became an important component for every part of the sustainable development goals, which emphasise the need for that particular goal to collect data. Yet with all this lack of data, it's surprising that data actually exists.

But it exists in the hands of people like Facebook and Google. So the fifth and final chapter reminds us that in a rapidly growing data landscape, there is a growing divide between the people who have access data and those who do not. And while data was once something that only landowners and governments controlled, now private companies are accumulating exponential amounts every day. Some believe this amounts to data colonialism, where private companies extract our data as a resource and use it as a tool for control.

Putting the idea of data colonialism aside for a moment, I believe it's important to think of data as a public good and that we can actually use this to improve society. So in this example, I took Facebook activity data to identify areas in Nairobi that lacked proper internet connections. And here we see many of the low income communities outside the CBD experience, very low ability to connect.

The value of using this Facebook data can help us develop infrastructure. But we know that infrastructure development also means social development. These maps led me to a project called the Living Data Hubs, which is a project that creates a community owned, community built data collection tool and wireless network in Kibera, which we installed wireless hubs during COVID to increase internet connections, using our map as a guide to the neediest locations.

So I'm going to leave us with the data action principles that I left everyone in the book with. While each story I tell in the book explains this, I want to leave you with some final thoughts. First, number one, do no harm. We must interrogate the reasons we want to use data and determine the potential for our work to do more harm than good.

Build teams. Building teams to create narratives around data for action is essential for communicating the results effectively. And I think the Nairobi example is a great example of that.

Change power dynamics. Building data helps change the power dynamics inherent in controlling and using data while also having numerous side benefits, such as teaching data literacy. Again, the Nairobi example is great. Collecting data where the government isn't and then having them use it as such an empowering tool for the matatu drivers and owners.



Expose hidden systems. Coming up with unique ways to acquire, quantify, and model data can expose messages previously hidden from the public eye. However, we must expose ideas ethically, going back to the first principle of doing no harm. So here, we expose the system of vacant properties, but we also showed examples of exposing hidden systems with the million dollar blocks project.

Five, ground truth. We must validate the work we did with data by literally observing the phenomenon on the ground, asking those it affects to interpret the results. We saw that in all my projects, including the ghost cities, but also matatu projects, really asking the people involved in the map whether what we interpret rings true.

Sharing data. Sharing data is essential for communicating the need for policy change and generating debates essential for that work. We can't all be data scientists, but we can read designs communicated with data. And I believe that's one way that we create open data sets.

Create your own ethical standards. Remember that data are people, and we must do them no harm. We must seek to develop our own ethical standards and call others to do so. And really, one of the things that I'm thinking about here is that technology develops much more rapidly than our ethics can be applied to. And it's up to us to create those ethical standards.

So I leave you here. Data Action sets out to remind us that big data in its raw form cannot perform on its own, rather how data is transformed and operationalised that can change the way we see the world. More specifically, data can be used for civic action and policy change. Thank you very much. I'll open it up for questions.

PAUL LAVEY: Thanks so much, Sarah. That's such an enlightening, fascinating talk and run through of a huge body of work that has had clearly a massive impact in a lot of parts of the world. I'm just going to kind of help coordinate questions, and I'd really like us to have some discussion in this. And some people that I can see in the crowd gathered today to-- who are working in this space or around this space.

But I have a lot of questions myself, so I think I'll just jump in with one. And I'm sure you've preempted, and you have an answer for this question. But it just picks up on that last point around creating ethical standards and the pace of change with technology and in particular with data and data sharing practises in Australia at the moment.

The federal government has had a couple of years of going through some work around regular or new forms of regulating data sharing, but particularly amongst the public sector and how that public sector data might move outwards into the private sector is essentially how it's positioned. But one of the-- and alongside that is a lot of work obviously in creating ethical standards and imposing those sometimes as regulations, sometimes just as guidelines or principles. And I'm just wondering about your thoughts around-- you mentioned at one point repurposing data, so using open data for example for alternative purposes to what that data was collected for.

The question that I have and grapple with all the time is how much do the protections of data and data rights and personal sovereignty and care for people, how much do those protections imposed challenges on the way that we might use data for social good or do all of the things that you've done

with data for example? And if you have even any way through that conundrum or any ideas around that, it would be great to hear your thoughts.

SARAH WILLIAMS: I think this is a really good question, something that we're thinking about all the time. I think in some of the examples that I showed where we're using data sets that are openly available, maybe produced for one purpose, and then I use them for another, the things that I think about is have I aggregated enough-- the data to remove personal identification? So I think a lot of times when people are using like Twitter data in that way, they're actually still exposing the individual.

And I think that that is problematic, rather than let's say, taking the data in a more aggregated format. I think also, in each one of those projects, like where I do, let's say, scrape Twitter data, I also go through a ground truthing process with the community. And I guess when I talk about ground truthing, I mean really asking those inside the data or who made the data if what I have observed rings true to them, but also if it has any potential to harm those communities.

And I think that part of ground truthing is really important and that process that I went through in ghost cities was important part of that conversation with a Chinese community who could have been very defensive, but also actually used it as a conversation point. I also talk to-- in that case, we actually told DianPing that we were taking our data. So I always inform-- so in Twitter projects, in DianPing or like the AMAP, we informed them that we would scraping the data in that way just to-- I mean, even though it's, let's say, like typically against their terms, like they usually don't have a problem with it because it is publicly available.

And I think oftentimes, they wanted to see it be used for good. I imagine others might contact them, and they wouldn't allow for that kind of scraping. I'm not sure. But I do feel like it's important in my own ethical practise to make sure I reach out and make those contacts with the data providers to let them know. And that can be useful in helping highlighting bias in the data sets as well.

PAUL LAVEY: Great. Thanks. We've got a question from Anne. Anne, did you want to jump in and ask the question yourself? With mute off. Sorry, Anne. It sounds like you're on mute still.

ANNE: Apologies. I had the initial question which was around the methodology slide that you put up very early on, which included things like ground truthing, opening up data, so different types of data and so on. So I was interested in hearing a bit more about them, and you've explained more about what you meant by the ground truthing data. I thought it was a very interesting sort of series or way of differentiating different types of data and how to intervene and engage with it. I was wondering if you could say a little bit more about that from a methodology point of view.

SARAH WILLIAMS: Yeah, absolutely. So whenever I work on a project, I really start by building the team that can answer the right questions. In fact, I'm just starting a project right now with the World Food Programme to look at issues around migration from the central triangle in Central America to the US.

And so we've built a team that includes the Migration Centre, the other kinds of policy experts, but also like data scientists, architects, and designers as we begin to ask the questions, and then actually

acquire the data. So next step is acquiring the data and collecting it or analysing it or gathering it. So like in the Twitter cases of gathering it or scraping it.

In this case, we're doing a survey of 5,000 migrants from the central triangle. Then analysing that data is the process that we're going through right now, interpreting the results, and then ultimately visualising it is what we really want to do for the UN Assembly Council and finding out ways that we can communicate new insights about what's going on in this migration pattern ultimately.

And as we do that work, we're going to ground truth our analytics with people, the migrants within the data visualisations and ask them both to interpret or tell us about our results, but also we'll be doing video interviews with them, asking them to analyse, which we hope those videos will be part of the larger communication strategy of what's going on with the migrants coming from the central triangle, which is Guatemala, Honduras, and El Salvador.

And always as we do projects, we've gained new insights from these ground truthing processes that really often change the graphics or change the kind of data we want. And our teams kind of rethink it, and the iteration begins again. So that's kind of why I think of it as a cyclical process as analysing data really never ends. But it's a constant process of evaluation.

ANNE: Yeah. Thank you. I just wondered too whether have you used these approaches in, I guess, small settings like organisational settings for example, or do you tend to work at this very large bigger picture level?

SARAH WILLIAMS: We work with small organisations and big organisations. Yeah, depends on the project.

ANNE: Essentially decide the process, but scale up or down as needed.

SARAH WILLIAMS: And typically like in an Organisation setting, those teams are already there in a way. You have your policy person. You have probably somebody who knows a little bit. You might need to bring in a designer to help visualise it. But I think it scales down and scales up as it were.

ANNE: OK, thanks.

PAUL LAVEY: Speaking of scaling up and scaling down, Jane, did you want to jump in with a question as well around that?

JANE FARMER: Yeah, sure. Obviously, I've got heaps of questions. But I my question is really around like the projects you showed us are all great. And I think we are quite passionate about engaging the public with the idea that there are data out there. There are data that they have generated or are significantly represented.

I was just wondering if you can sort of give us any tips or big picture thoughts, maybe even about how to roll out that kind of engagement of the public with data. In some of our projects, it seems that there's two approaches to data. One is like super fear, that their data are going to be out there, and on the other hand, a bit of an apathetic kind of perspective. Yeah, so I'm kind of wondering how

do we sort of scale up and roll out this kind of idea of being able to engage the public more with data?

SARAH WILLIAMS: So I think that like in all the projects that I did, I tried to create let's say like this public interface with the data. And it like often just depends on the project. So oftentimes, that's like-- I like to do interactive websites, which allows the public to actually be using it, playing around with it. Doesn't take expertise, but also allows them to find their own knowledge in the work that you've presented. So that's one way that I think is really effective for engaging the public within your data set.

But I think perhaps you're also talking about like the public's fear of sharing the data in the first place.

JANE FARMER: Yeah, I'm just-- I guess I'm kind of thinking that you showed us lots of projects. And obviously, projects are great. But how do we kind of scale up that whole idea of engaging the public? I'm wondering, is there a role for NGOs or councils or community health organisations that collect-- they collect lots and lots of data. I mean, they maybe haven't grappled with what to do with themselves, but ultimately, hopefully they will.

And then they really, I guess one argument would go need to engage the people who produced the data in those conversations or how they can get benefit from data. So how do we go about that? Do organisations in the US kind of do that kind of thing, engage their end users with the data that they generate at all?

SARAH WILLIAMS: Absolutely. I think that organisations absolutely engage the public, users with data. I think that they go about it in lots of different ways. I think most often with shorter tidbits, things that maybe can be posted on social media. And even the million dollar blocks project I showed you, which is still used like often gets passed around on Instagram and so forth.

And so I think that's one way. I think that it really depends on what your organisation is. I know that I work with a lot of organisations that do housing issues, and they often show interactive maps or use interactive maps when they're meeting with their communities to better explain some of the trade offs towards like development projects. And using data in that way to explain the trade offs can be really impactful community engagement in that it really helps to educate the community and break down knowledge barriers for them to be able to make important decisions about the places they live in.

But I think that can have them both through these kind of interactive visualisations, but also static ones as well. I think the interactive ones are better because it allows the communities to find their own voice within the data set. But I would say we have organisations that do this more than others. Like we have like this group organisation called Policy Link that's-- in fact, the head of their data group just wrote a book about data visualisation. And so they do that quite often.

I'm trying to think. I feel like each city in the US has their own kind of data visualisation group now that either works with GIS or other kind of ways of making community engagement. We have a version of AmeriCorps, which is focused on technology and community engagement as well.

PAUL LAVEY: Can I jump in with a related question? I guess they're all related. I'm really interested in the parts of your work where you're talking about building data literacy and building skill sets among people who are-- communities who are engaging with the data. And clearly, some of the work is very complex. And I'm just thinking about the ghost cities example, some of that analysis of the analytic side of things. Analysis can be quite complex using complex mathematics or modelling or measuring.

What are some of the tools or I guess techniques that you use or know of that help to build skills? What are the kind of range of things that can be done to improve data literacy and in that sense across the community?

SARAH WILLIAMS: Yeah, I mean, I think there's so many great tools out there to get started with data, everything from like Tableau, which is a great. I think beginners if you want to start working with data or even advance in some ways. There's Plotly, which is also a great tool for like data visualisations and like putting data together.

I think that there's also like other mapping tools, like CartoDB and Kepler that make it really easy to make maps so easy, that almost anyone could take their census data and put it on a map. So I think there's a lot of great mapping tools.

I think, though, in terms of data literacy, it's about-- I would say data literacy is about the ability to make arguments with data. And I feel like while we have a lot of these tools available to make us say like a data map or a bar chart, we don't always have the skills to make the arguments that we need. And so I think that's something that really building the teams-- excuse me-- building the teams help with, in that, OK, I might be able to make this map and chart my data. What does it mean in the world?

And through like editing with your collaborators, it helps kind of build those skills and arguments as people test you and push you to try to move those arguments forward with the data sets that you have. But I'm just thinking like if I was a complete novice, like I wanted to use data I would totally go to Kepler or [INAUDIBLE] or Tableau and start experimenting in that way.

I do think that if you've never used data in your life, how do you get help? There's so many great meetup groups that help with thinking about data and data literacy that help-- there's even a Kepler meetup group to make maps. And so for those that aren't within an academic community, I do think there's a lot. We're lucky that we live in the world we live in right now where we have lots of ability to reach out for help through these different groups and organisations.

Even policy link that I mentioned often helps organisations and advises them like either on a nonprofit or collaborative way. And we have lots of organisations in the US that are specifically focused on helping organisations build their data skills and do consulting and consultancy work along those lines in the hopes that they can build the skills. They don't want them to become dependent on them, but if they are, really to transfer that knowledge.

PAUL LAVEY: Yeah. Look, I'm just conscious of the time. We've just hit 10:30 all of a sudden. It just crept up on me. So I'd like to thank you, Sarah, deeply for your insights and sharing them with us. And thank you to everyone who came along and the questions.

Obviously, we'll have more as we go, and hopefully we'll be able to keep these kind of links open in terms of our collaboration around working with data and data in action and the organisations that we work with here in Australia. It's been great to get your insights.

I just remind everyone that we have another of this series next week with Julia Stanovich from New York University, looking at AI and engagement, public engagement in AI and machine learning more specifically. And I think that follows on nicely from what you've spoken about, Sarah. So thank you again from us and thank you everyone for coming.

JANE FARMER: Thanks a lot. Thanks, Sarah. Thanks, everyone.

SARAH WILLIAMS: It was good chatting with you.

JANE FARMER: Have a nice evening.

SARAH WILLIAMS: You have a nice day, everyone.

JANE FARMER: Cheers. Thanks a lot.

PAUL LAVEY: Thanks, everyone. Bye

[END OF TRANSCRIPT]