

## Transcript | Network Book Forum: Democracy's Data by Dr. Dan Bouk

Panelists: Dr. Alex Hanna

Host: Ronteau Coppin

August 11, 2022

### Ronteau Coppin

12:00:01

Welcome to today's Network Book Forum, spotlighting Democracy's Data by Data & Society alumni, Dr. Dan Bouk. I'm Ronteau Coppin, People and Culture manager at Data & Society here with Dr. Alex Hanna Director of Research at the Distributed AI Research Institute. I will be your host alongside my colleagues behind the curtain. So let's get started.

Data & Society is an independent nonprofit research organization. We study the social implications of data-centric technologies and automation producing original research that can ground informed evidence-based public debate. We believe that empirical evidence should directly inform the development and governance of new technology. Data & Society began in New York City, an island node in a large network of Hills rivers and mountains in the Atlantic Northeast known as Lenapehoking, the ancestral land of the Lenni Lenape people. Today, we are connected online via a different infrastructure, a vast array of servers, humans, and computers. In the United States, much of this system sits on stolen land acquired under the extractive logic of white settler expansion. As an organization, we recognize this history and uplift the sovereignty of indigenous people, data, and territory around the world. We commit to dismantling all ongoing settler colonial practices and their material implications on our digital world.

The census isn't just a data collection process. It's a ritual and a tool of American democracy behind every neat grid of numbers is a collage of messy human stories. You just have to know how to read them. In Democracy's Data: the Hidden Stories in the US Census and How to Read Them. Dan Bouk examines the 1940 US census uncovering what those numbers, both condense and cleverly abstract, a universe of meaning and uncertainty of cultural negotiation and political struggle. I'm very honored to introduce Dr. Dan Bouk to say a few words about his book before I invite Dr. Alex Hanna to join our conversation. After that, we'll turn it over to you in the audience for a brief Q and A session. I encourage you to engage with us by adding your questions throughout the session, and upvoting the ones that resonate most with you. Thank you so much for being here today, Dan. I'm really excited to hear you give your author remarks and walk us through the backstory of Democracy's Data, which I know will help to set the foundation for our leader discussion. Please join me in welcoming Dan.

### Dan Bouk

12:03:07

Oh, thank you. Thank you. I'm feeling the welcome through the digital ether. And I really appreciate that. Ronteau, thank you so much for agreeing to host this conversation. I was really honored that you were willing to do this. Alex, it's such a pleasure to be able to talk about this book with you. I'm really honored by everyone else. Who's showing up here, either showing up

right now with us in real time, and those in the future who will show up with us in the digital replicated spaces and hang out with us, think with us. And I wanna thank Rigo for producing this, Nazelie for setting the groundwork, and I know there's a whole production team of Data & Society that makes these things go so smoothly. I wanna thank all of them too. Welcome to this pre book launch event. I'm happy to report that while, uh—I'll get my, the title of my book proper—while Democracy's Data does not ship for two more weeks, it exists and it is a really finely crafted object. I cannot wait for you all to get to hold it and smell it, and of course, read it.

So let's see if I can get the slides to work. There we go. One of the things that I think about a lot is the proper response to this slogan. Let the data decide, do what the numbers dictate and is the slogan that arises from the belief that the world would be better, fairer, if every choice was based on quantitative evidence and individual judgment was replaced as much as possible by explicit criteria. It is a slogan that accompanies the rise of predictive algorithms and machine learning. It's also a slogan with a much older impulse than one might think.

So, let me tell you a story. I'll sketch one out for you. In 1940, 120,000 people set out, marched out to count over 130 million of their neighbors. It was a staggering undertaking, akin to mobilizing an army. It was often spoken of as something like mobilizing for war, indeed, this army of enumerators generated. Then these figures that you're seeing on this table, these numbers show how many people lived in each state. In 1941, Congress used these numbers as acquired by the constitution to apportion seats in the US House of Representatives and votes in the electoral college to apportion them amongst the various states. Congress also made it so that each future apportionment, which each one of these splitting up of political power would happen automatically without its explicit action. So for nearly a century now, letting the data decide has been a fixed principle in American democracy.

Now the upshot of letting data decide was, and is that seats swing seamlessly from state to state political fights about apportionment have moved out of Congress and into the courts. And I think most importantly, the entire legislative system has entered a period of profound stasis in its early history. You can get a sense of this by looking at this graph here from my friends at [isapportionment.org](http://isapportionment.org) in its early history, when Congress and politics decided on the organization of the house, its membership grew steadily with population albeit a bit more slowly. Now the house is fixed at 435 seats for no particular reason. And as the country grows, the relationship of legislatures to those they represent rose evermore thin, distant. This is one consequence of letting the data decide.

Democracy's Data tells a story. Actually, I would say it tells many stories from many different vantage points over the 1940 census. Here's why. The 1940 census was a massive data set—still is—one that made fundamental decisions for American democracy. And most importantly, all of the records from the 1940 census were available for close study. When I set out to write this book, those 1940 records were the most recent census records that I could investigate in full. A few months ago, or last April, that changed. The 1950 census records have been released. And indeed I hope that many of my readers will one day take what they learn and go exploring for themselves in these new records. At any rate, the 1940 census offered me

a unique opportunity to get the note, the messy innards, of a data set—one that was expected to drive direct or even automate governance. Now, in my research, I could look at archival documents, showing people in charge, trying to figure out how to get good data here, a bureau official named Kale Dietrich wrote to Milton Friedman, yes, Milton Friedman, the famous conservative economist, complaining about people who didn't like the new census question about incomes. So Dietrich wrote, "We are under extremely heavy fire because of the income question." I could also look at original census sheets filled out by enumerators in April 1940 or thereabouts. Now these are a species of records that any of you who are genealogists or who have done any genealogical work are probably quite familiar with.

So here's a record for Ray and Florence Dowd. Now look to the far right hand side, look at your right hand side of the screen, and you'll see that there's two zeros next to a 52 and next to an eight. And so that indicated that Ray and Florence Dowd claimed to have earned \$0 in wages in the previous year while they worked for 52 weeks and eight weeks. Now that could be true. Luckily for the 1940 census, I could also frequently consult other records to read alongside the original data. Like this drawing, which was sent by Florence Dow to a Senator in which she writes "I had these brass toes made just for such an occasion as she imagines kicking an enumerator, in the rear out the door." So maybe those income figures are not to be taken as entirely trustworthy.

One consequence of examining the data closely is that we discover how messy it is. Here's an illustrated chapter outline of my book. Here's at the beginning of the book as well for when y'all go and grab your copies. Don't worry, I'm not gonna go take you through every single chapter here. Chapter one deals with people who designed the form itself and decided on questions and categories and the final chapter considers the data as published and used by others. So these are the stages of data production the census studies usually focus on. So

I mean, great books like these: Debra Thompson, Melissa Nobles, Ken Pruitt, Paul Shore, those sorts of books. And of course, Margo Anderson, the American census, which I have in a different bookshelf.

But most of Democracy's Data instead thinks about the conversations that took place at millions of doorsteps. As a nation of people from all over identifying with many races and ethnicities in a stubbornly white supremacist system as these whole complex individuals tried to make themselves known to their country. This process was filled with creativity and also conflict with silences and errors with pride and puzzlement. I think we need to spend more time in each data set's doorsteps. That is a project for which the usual suspects, the quantitatively skillful are welcome, but we need the humanists, we need the genealogists, we need those who bring to the process a wide range of backgrounds, perspectives, and questions.

There are stories in the data. We just have to learn how to read them. And when we do, we'll be in a better position to deal wisely with our data sets and to know how to trust them and how far to trust them. Thank you.

**Ronteau Coppin**12:11:14([12:14](#))

Thank you, Dan. That was very insightful. It was good to have that background and have you detail the process and decisions behind the chapters of Democracy's Data. And I want to invite Alex into our conversation about the census. I'll start us off with this quote that, as I'm reading the book a few times now, I'd say is quite fundamental to the census story and that is: "To know America, each American had to submit to being known." Something that's quite integral in your documenting the census story is this idea of silences. So you make it a point, Dan, to distinguish, retain stories and numbers as you just did. So then how should we read for silences, the hidden. That is, how should one approach reading this data, knowing that everything isn't actually captured?

**Dan Bouk**

12:13:10

All right. Reading for silences, I think, is interesting. We have to distinguish the kinds of silences we're talking about because different kinds of silences will tell us different things. There's one easy way to say that, to imagine silence isn't necessarily a bad thing because it's missing, but I think maybe the baseline thing we need to say is that silences are necessary for communication. This is one of the crucial concepts of a book called *Silencing the Past* by Michel-Rolph Trouillot, which I'm sure a lot of people here have read at some point, but it's most important to pay attention to silences, where they come from, but to recognize that we can't say things, we can't study things, we cannot create data without choosing to make something as visible and making other things invisible.

And so in that sense, like one reason we look at the silences of data set is because they tell us something about the values of a particular time, particular society, the particular people who are behind making the choices of what gets counted and what doesn't be accounted. We could dig into a little more, but I don't wanna get too deep into the weeds. I will say the one different kind of silence though, are unintentional silences and for a census, the most obvious kind of unintentional silence would be what we call an undercount. And in this case what I talk about in the book is in the early 20th century, the scientific discovery and quantification of the undercount of individuals in the census was led by Black scholars first in the census bureau, and then particularly a scholar named Kelly Miller at Howard University who very aggressively in the 1920s put forward this claim that, and very insightfully put forward this set of claims, that the 20s census had undercounted African Americans and the census bureau at that point disputed quite vigorously, that set of claims but then by the middle of the 20th century, it came to be widely accepted that in fact, there had been and continue to be persistent under counts where some people just don't get counted by the census.

**Ronteau Coppin**

12:15:27

Alex, do you wanna jump in here and kind of talk about how you might see silences in your work, um, engaging with the census?

**Alex Hanna**

12:15:36

Yeah, absolutely. And just want to thank you all for inviting me and for the team to put this together, I'm super thankful and happy to be here. I love this question about silences because there's so many different silences that we can think of here, and in the book, something that I love love love about this book is the way in which the stories in which the interactions are given flesh, given sinew, given blood, you can see what kind of media these people may have interacted with or what they might have been thinking at the time, the ways in which people concerned about privacy were still sending drawings of kicking a census enumerator out onto the doorstep and the things in which we're talking about in thinking about silences, um, silences suffuse all areas of the quote unquote data pipeline, right. In the chapter on the people who create data sets the question men, titans of industry, who come here and are asking, are interested in, this contentious income question and are less concerned and thinking less about things that might be of importance to the republic in terms of nationality and immigration.

So silences are those of which are asked and which are of a mission. Uh, some of the silences are, um, embedded in the enumerator training process, some of the enumerators go throughout and they collect some information, but they emit other information either because the form which has a constrained set of columns only fit so much or by their own kind of predispositions of how they identify someone by race or ethnicity or by gender. The affordances given by the computers at the time, this new technology of the punch card, literally these kinds of first digitizations, can only encode so much.

So these silences occur at different places. And, I think in modern discourse on machine learning or data driven technologies, things that we may call silences in this view are sometimes all looped together in something called bias, which I think does a lot to obscure what is happening at each point and what is even the cause, whether that's intentional or technological or institutional. And so I love this kind of thinking about how to read for silences and a kind of parallel of this work, there's a paper that Morgan Klaus Scheuerman and Emily Denton, and I wrote on data sets and computer vision and we surveyed over a hundred different data sets, and we're looking at what the values of the curators were in writing up these data sets and found huge silences and talking about how they treat care compared to speed or how they valued some kind of ideal of universality, whereas they eschewed this kind of concept of contextual, which would give data sets possibly a better kind of understanding in the training and evaluation of these machine learning models. These silences are really interesting and I think an important way of doing this type of data reading in whatever the context.

**Dan Bouk**

12:20:08

If I can just piggyback on two little things. One is I love that point about bias. I think it's really so interesting for historians, and probably a lot of people who do different kinds of cultural theory, right? Bias is such a strange word. I often talk to my students and, and say, oh, I don't want you to talk about bias as a bad thing, because like bias is the only thing we have. Like every source you read, bias is the worldview, it's our subject, we don't wanna remove bias. And yet at the same time, clearly there's a good reason that bias has become so important to this conversation.

But the other thing I wanna say is just about how silences are, especially as what I call this doorstep data, the silences are hard to keep silent. And so for instance, there's the story about this enumerator and, Alex got me thinking about this. In Texas wandering around and because of a big politics like politics at the international level, it's been decided that there can be no racial category called Mexican for 1940, but this enumerator wanders around in her district, and she labels all kinds of people Mexican. Now that will be erased later on, but at this stage, this is one of the interesting things, **even though there might be silencing in a later part of the data, this is the data too, at that moment of that enumeration. And we can see how people are speaking and making judgments and making claims that might later be silenced if we only look at those numbers.**

**Ronteau Coppin**

12:21:40

I'm glad you brought that up because I think for me too, in hearing you all talk, the silence is not just the census part, right, it's also in how we choose to answer the questions. We being now, but also the people who actually filled it out in 1940 and onwards. And so I think something I wanted to raise, which I knew would come up would be how people chose to tell the story of who they were. And so in that regard, too, we could be silent and, you know how we say how we're counting kids or, you know, how we are, describing our race or our gender or any other identity markers. So I wanted to throw that in there as another thought for y'all to kind of riff off about how it is that the people themselves kind of have a play in the silence. And either of y'all can start. Maybe Alex, you can start.

**Alex Hanna**

12:22:37

Yeah, absolutely. I mean, this is great. And thinking about the stories we tell about ourselves, and, the book, and this is just a plug for the book, there's such a, there's such a vividness in these individual stories of these people, of this Dowd character of, of this enumerator in Texas, of people and what they're doing at what time. And I mean, and you do that yourself in the epilogue in a way that's very sweet and I could even go about this thinking: if

I'm going to fill out the census, as you rightly point out, you know, Egyptians are to be counted as white, and there are still, you know, to be counted as white in the 2020 census and everybody from middle east and north Africa. And there's some great scholars like Neta Maghbouleh who's written about that as well, and the fight for that. I'm a transgender woman, and over the years, my gender will have changed. There's of course this body of research that focuses on how people's self-perceptions also change, depending on the stories they tell about themselves.

So the stories of self, this idea of where we are and how we can fit in it are stories of intentionality, but they are also stories of resistance, right? There's this great story in the book about this idea of Mexicans and Mexican American advocacy organizations encouraging people to classify themselves as white to gain the same advantages as of citizenship. And then on the other side, you have people fighting against that and saying, no, we're not white. And it reminded me a lot of Neta's work actually in, in this, in her book *The Limits to Whiteness*. In the first chapter, the limits to whiteness, she talks about these cases in which Iranians operate as what she calls as "racial hinges," you know, the way in which there are appeals made to Iranians as both a foil by other groups saying that they're not like those Iranians, they're not like these zoroastrians, these pagans, we are actually white, compared to others who are coming from south Asia who are making appeals and saying, oh, we're just like them, they're Arians and we are actually of this white group.

They operate in this kind of dualistic way. And so much of it is a story of self and a story of our perceptions. And as you rightly say, this major shift happens in 1977 with the office of management and budget directive that allows self classification and terms to race and gender. And it's a function, both of the stories of self, but also the kind of affordances that we have to squeeze into these forms and these boxes and whatnot. And there's that tension there in the epilogue, which is the tension of wanting to, you know, be a good citizen of trying to have political power, but also that there is this inherent violence of this recognition when we can't put ourselves into those boxes. So I love that. I love you calling up that tension.

**Dan Bouk**

12:27:10

This is the thing that more than anything else has always drawn me in all of the work I do, which is I at one level, understand that our identities are very powerfully constructed by bureaucratic systems. They literally often give us the language and the terms with which we can organize ourselves and work with one another and see one another. And at the same time, our own senses of self, both because Dallas points out, right. They change over time and because they're peculiar and situated so seldom actually work. I written about life insurance, I've written about the census now, and it's not because I like finance or bureaucracies. I think it's because they both frighten me a great deal. And they feel very foreign to me. And as a result, I spend a lot of time being fascinated by how it is that they actually work. I wanna spend some time trying to understand these things that I understand are really powerful in my life and to see, like, how

have people managed? Like how have people gotten by, in the midst of all of these really powerful forces?

**Alex Hanna**

12:28:26

I love that you study these things, but they terrify you a bit like, it's great. They also frighten me, insurance and the sort of these logics of data keeping and which technologies you're replicating from institution to institution. And there's something very brave about jumping into these systems instead of dealing with them in a way that's head on and trying to understand their internals.

**Ronteau Coppin**

12:29:06

I'd like to jump in here because I feel like we're kind of heading in this direction, so let's just do it. I feel like a good theme to talk about is data genealogies and governance, right? So there are all of these systems that we play into whether we want to or not. So I wanted to ask about how the census serves a kind of structural governance function. Dan, I'll let you kick off this one.

**Dan Bouk**

12:29:35

You're right Ronteau, we are totally going in this direction because I was even thinking as I gave those comments. I should say these systems frighten and fascinate me and I know that they're important and we probably can't have a mass society without them. I have pure ideas about insurance. Like a good social insurance program in particular, I think it is really, really important. Private insurance systems, I think are still very useful, even if they're prone to more difficulties. But as I say in the book, I believe in the census despite everything, but I say in the same way that I believe in democracy. I'm sure there is a way, but I don't know how to have a democracy without something like a census without this kind of data system. Because the premise, the ideal we'll say the ideal has always been, every person will be counted and then they will be represented in the government based on whether they're counted.

And of course, I have to specify that it's an ideal because for much of American history, huge swabs of individuals were not counted or not counted fully by the census. Very explicitly the constitution did not count Native Americans. There's a complicated story, which in 1940 has its own interesting story about how and why this category of the Indian not taxed came to be removed from all census calculations. But then also very crucially enslaved people were, were counted as three fifths of a person, which has its own result of, of terrible compromises between, um, states that were dominated by slavery and other states that also had slaves, it's important to say, at the time the compromise was made.



So all of which is to say infrastructure you need for democracy, you, you do actually just need to count people. The census has, in recent years, since 1940, or really the beginning to happen in 1940, come to take on all these other really powerful structural powers as well. It was for a long time used to shape congressional debates, but now Congress increasingly gives it power, which is one of the reasons the census has become so powerful. So contentious over the last 50 to 60 years, is that now not only does it shape people fighting over how to like draw district lines, which have to do with representation, but many laws automatically allocate, especially social funding to localities and to communities based on census data. And once you let the data decide, it suddenly becomes the battleground itself. It becomes the place in which people have to fight because to get them more money to get their fair share, they have to know how they're being counted.

**Alex Hanna**

12:32:43

And I would say, you know, the census place is maybe one of the paradigmatic structural data sets since the longevity it has, that it's enshrined in the constitution, that it has carried that kind of heft. It's very entrenched in how we think about the social fabric. And it's kind of interesting because there's data sets that have proliferated in the past, half a century, even in the past 20 years, that also take on this really infrastructural function, but don't have the longevity of the institution and the different iterations of the data set. A lot of my work, we discuss a bit about the infrastructural functions of some of the benchmark data that's used in machine learning, Rigo posted the paper that we wrote for logic magazine. There's another paper we wrote from Bringing the People Back In that talks a little bit about the infrastructure function of benchmark data sets.

But these data sets are pretty new things like ImageNet have only been around for something like 12 years, maybe a little older. I think a lot of what you do in this book and, you know, and there's a part of this book that I just need to read because it I have all kinds of arrows and things around it, but it's about unsettling that infrastructure. It's taking a bit of cues from infrastructure studies, but it's this line we say, "we need more investigations that the data history is not less, more people willing and able to read the stories behind the numbers."

And that just really drives the point home that whatever this way, however authoritative these numbers or these data seem to be, we need more stories, we need more methods for unsettling that. And as you said, at the top of the top, leaving that merely to the, the quants and the mathematicians is not sufficient, actually, you know, the kind of intervention, and I'll say this from my own work, the kind of intervention has had to been archival, ethnographic, textual, and nature, because there's just a whole host of action and people and doorsteps, you know, whatever that doorstep can be conceptualized as, that that's really where the shoe leather hits the concrete.

**Dan Bouk**Speaker 2: ([35:48](#))

The way you're putting it Alex, one of my hopes with the book is that even if somebody in the end is like, you know what I really thought about it, and I don't care about the 1940 census, they can wander away from this book, still having thoughts, like all right, this is a training ground where I can hone these skills that I can then take into the world to go and do these investigations in data sets that I think matter more to me at this particular moment. I'm glad that you're seeing that too, that is certainly one of my hopes

I will throw into the mix, just because I feel like Rigo needs more work, Joanna Raven's digital natives, as I know a piece that Alex and I both love, it is the one piece I always tell people is the most important history of data piece that everyone needs to read, it's exemplary in this project of why and how we need to tell the story the genealogy of a data set, and what it means when we then just use it afterwards,

**Ronteau Coppin**

12:37:01

As we're talking about how data drives governance, over time, do either of you find that the process has become more productive or less productive based on how the census has kind of changed? Or has it been pretty much on the same level? Dan, we can kick this off.

**Dan Bouk**

12:37:23

Alex earlier alluded to some of the really important things that have changed over the years, and I'll answer with a story. This is a story that a friend of mine gave me permission to share in the book. I'm grateful to her for allowing me to share this. It gives a sense of what happens when somebody encounters a census form. The first clear thing to say in is that folks today, anyone here who remembers taking the 2020 census and most people don't remember it because it turns out it's not a very memorable experience, except for that moment where, in which you're frustrated, because you're being asked to do something that you, that doesn't fit you, but then that even usually just goes away and you forget that you ever did it. But when people are filling out this 2020 census form, they're only answering like seven questions. Whereas in 1940, they would've been facing this big old sheet, like the size of the newspaper worth of questions.

The other important distinction is that generally 60% or so of households fill it out directly online, on paper, something like that. And then lots of other folks do in fact have an enumerator come to their door and ask questions outta them and fill in their answers. But those enumerators are now supposed to put the answers down that a person tells them. Whereas in 1940 quite frequently there are many of the questions enumerators were encouraged to make judgments and decide for themselves what was happening.

And so I'm gonna take this now in two directions, and I'm gonna wrap up my comment. Because one of them is a little bit of an earlier conversation that we were having. I think it's important to note sometimes we think about the state producing big categories and things that then cross identities and then individuals making choices about who they are. And I think one of the things that is most important to me in the book is to recognize how many of our identities and ways in which we come to know one another are much more locally social.

For instance, I've got a chapter about this category of "partner" and I look at how in different places in New York, in Hawaii, in San Francisco, people get identified as partners. And we often don't know what it means for a person to be a partner. What we know is this kind of relational encounter in which a person and an enumerator showed up and an enumerator wrote down "partner" saying this person is a partner to the other person in their household, and for me it's not that important. Or it's really important of course what each individual said, but I can't figure out what that is. All I know is that these people in this particular setting settled on this as a viable category. And one of the findings is that in queer communities, people are more likely to identify as partners, which doesn't mean that they were in what we would today call queer relationships. They were queer in a sense that they broke heteronormative norms at the time, but they constructed these social identities together. And then much of what identity work is and why we need to look at the doorstep data is because from place to place, time to time, social norms at local levels are shaping who we are to ourselves and to one another.

And so that's like one part of the story. The other part of the story, how things have changed then is now, so your census forms are supposed to be your own self-identification, but it's not like everything else has gone away. So, we're still identifying ourselves in our local identification and in like these larger political structures. So my friend who's a very savvy census operator, when she fills out the race categories, she does this careful calculus. Her mother is Black Haitian and her father would be classified according to the census as Asian Indian. And so she has the right to check both of those boxes, but again, being a very savvy operator, she recognizes that if she does that, in many tabulations she'll show up then under a "two or more races" category, just all bunched together. And so she doesn't want both of these groups to essentially lose their representation on all of those tabulations later on. So instead she strategically chooses one or the other. And so she, in the 2020 census, identified solely as African American or Black, precisely so that one group will definitely get her and will have her count in their count, she'll count for them instead of being lost.

So things have changed, but we still have these very complicated calculuses that we have to think through. And of course, most people don't think through them and instead there's, so the, the data we get the result of impromptu choices.

**Ronteau Coppin**

12:42:35

There are some questions coming in the Q&A. So I do wanna make time for that, but I didn't want to not allow you to respond if you wanted to, Alex.

**Alex Hanna**

12:42:43

No, I think that's beautifully put, and I love thinking about this as a strategic calculus, if one knows about this, I mean, for my own purposes, I refuse to be classified as white, I typically fill out the other. That's the story that I have told myself and the sort of strategic resistance, but know that there's consequences to that. But I'd love to get Q&A.

**Ronteau Coppin**

12:43:14

Let's do it. I encourage everyone to continue to drop questions in the Q&A, but I'll start off with a question that comes in from Katherine Morse and it's directed to you, Dan. What, 1940 census question did you find most rich and fascinating and what stories did it reveal? I know there's one about the presence of radios in US households, which I've been puzzling about.

**Dan Bouk**

12:43:40

The radio question I think introduced was in 1930 to begin with and was at that point quite contentious. So one of the surprising things is that there's a lot of census poetry, like much more poetry about the census than you could possibly have anticipated otherwise, often because people just write poems, satirical poems, and send them to the census bureau usually again in protest or in annoyance. So in this case for the 1930 census, there was a Manhattan woman who was annoyed about the fact that they were asking about the radio in her household. And so she wrote a poem about how annoyed she was about the radio question and sent it in there. But the radio question in 1940 would show up in the context of the housing census. So the 1940 census is the first time at which a special census of housing was first introduced and that's another part of this larger story, which is I see the 1940 census as crucial to seeing the US government shift towards suddenly being much more concerned about economic management and about thinking about the economic lives of all of its citizens.

So the housing census is there in part because the people wanna have a sense of whether FDR is right when he says that a third of all people live in substandard homes. And because they're also particularly interested in trying to stimulate home building as a way of getting outta the depression, right? Because it's important to remember that in 1940, as people

were going door to door, there have been ups and downs, it's still a very depressed moment in American economic lives. And people are really struggling.

I think my question for 1940 and I'll leave some of this for the book you can read more in the book, but it would be the relationship category, which is really interesting to me. And I talked a little bit about how then that the partner was particularly a category in a way in which I could kinda think about some of the reasons that the relationship category was interpreted. It's really strange. So people wanted to get rid of the relationship category, and one of the primary reasons it still was left in the census in the early twentieth century was to prevent fraud because they believed folks in the census bureau believed that enumerators, if they had to make up plausible relationships, would be less good at that than at other things, and so it would be easier to catch people just making up random people, a process called padding if they had to also provide viable family relationships.

Of course, then the trick thinking about this in terms of living in a queer household is that queer households don't look like normal households and therefore are much more likely to look like fraudulent households or to look like people that don't otherwise quite fit. And so the relationship category is one, the other one, is that this income question was incredibly contentious. And so for me, I loved the chance to investigate the granular census records and try to see how did people show their resistance? How do people who I know didn't like the income question register that resistance, how do they avoid answering it or find ways to avoid answering it, or did they, as Alex alluded to earlier sometimes say, I can't believe you, that I'm supposed to tell this to the census bureau. I don't want anyone to know that I make \$262 a year, which of course now everyone knows because I've put in my book because they wrote it to their Senator. So it turns out the best way possible to broadcast your income to the world 80 years later is to write your Senator and tell them.

## **Ronteau Coppin**

12:47:32

Thank you for that. Your answer just now is starting to answer one of the questions that came up in the Q&A. It's a long question, but you're gonna hear all of it and know exactly how to answer, I trust you to do that! There's growing awareness around data privacy and the ways in which our personal data can be exploited by corporations, the government, and/or bad actors in the same way that the Dowds refuse to share their income in 1940, do you see more people refusing to answer questions on the census for fear that it might be used against them, or simply because it's safer not to share—similar to what you were just talking about—Over time, would that result in an increasingly less trustworthy census? And what impact would that have on resource allocations, investments, et cetera. If you've come across changes that the census has made to address these concerns, that would also be super interesting to hear about as well.

**Dan Bouk**

12:48:32

So there's an empirical question there about whether or not the census is in fact kind of has over time become less trustworthy in terms of what people are revealing. And I, my answer to that is I have no idea, and I definitely shouldn't answer that because if I were to do so, it would be entirely inappropriate if I don't have any empirical data to give. I can give some context though, which is to say that the undercount is significantly less than it was in 1940. So people are being counted better. And the other shift that's happened is that most people aren't answering a question about their income anymore. So the 1940 is the introduction of probabilistic sampling. I talk about this as like letting the casino into the census. And what that allows is, is that now the census bureau uses the decennial census asking just these seven questions to create a full map of everybody in the probabilistic terms, it would be like "the earn" of the entire nation and then like they pick balls out of that urn to decide every year through something called the American community survey, who to ask these questions of.

And so many fewer people are asked these questions to probabilistic sampling that are much more invasive questions. Part of that, being the theory that if people are fewer people are asked these questions less frequently, they're maybe more likely to be willing to participate than asking everybody of this also. Congress got very annoyed about this in the 70s and started to make a real thought about the census being a deep, deep invasion of privacy, not everybody in Congress, but enough people in Congress that it became, um, useful for the census bureau to kinda devote more resources to sampling and finding ways to ask your questions of most individuals over time.

The other tiny thing I'll add to that is it's interesting to think about why it is that the census bureau keeps individual data confidential over time, and one of them certainly is to encourage people to be able to tell things honestly. But even there I've been researching this over the last few months and one of the initial reasons that the census bureau was insisting on confidentiality in the middle of the 20th century was precisely because they knew that their data was not perfect. They knew people and businesses faced for the census question often kinda had to shoehorn themselves in a way that they didn't wouldn't want to be represented publicly or in any like high stakes thing in a court, for instance. So one reason exists I think is so that we can compromise. So we can make ourselves known in a timely fashion to the nation in a way that's good enough for the purpose of statistics, but that we wouldn't actually want to be understood of as ourselves individually, which again, **I think is a useful thing for those of who wanna find individual stories in the census in the past to know that the individual stories we're finding are an enumerators recording of a person as they allow themselves to be seen through that enumerator at a particular moment. It's not who they are. There's a much deeper sense of who they are. That's, that's lost to us, but it's something that's there.**

**Alex Hanna**

12:51:48

And I just wanna interject here. I think something the book highlights well and provides context and is still very relevant today, is that the refusals to answer came, I think, both from very privileged people and also very marginalized people. And so the kind of resistance of many people not wanting to answer who actually made more than \$5,000 a year in wages. I have a friend who's a census enum who was a census enumerator in, in 2020 and was saying actually a lot of resistance in her anecdotal experience was coming from people living in very conservative areas of small government, not wanting to give information maybe with upper middle class incomes.

And then the other side of that is of course, you know, we'd be remiss if we weren't talking about the census information and its use in Japanese internment and the way and the way that had been overtly a tool and those fears have been reiterated with the citizenship question under Trump and secretary Ross, etc.

These things have still persisted on and there have been guarantees, but at the same time it's still the state having a history of doing such things proven with Japanese Americans and not proven for the upper class of overwhelmingly white conservative people.

**Rontaeu Coppin**

12:53:57

I want to try to squeeze in at least one more question. Both of you will be able to address this. Can we talk a little more about doing data history? What does it look like as a historical geographical method? What does it allow us to stay as historians of data and what might it obscure? Dan, you can jump in first.

**Dan Bouk**

12:54:19

It looks like very old fashioned historiographical methods. I spent a lot of time wandering into the national archives and getting my nice boxes and opening them and taking pictures and reading through and creating large PDFs and this sort of thing. One of the things that was different and interesting about this kinda work was I needed to spend a lot of time reading a lot of census sheets and just absorbing and generating questions. And sometimes that would happen, because for this book, I experimented running a blog census of stories while I was writing it. And from that, publishing a little bits of work as I was working, and got new questions.

People asked me questions about things that they had seen in the census, and then I could investigate that. And that helped me to figure out what kind of questions to ask about it. I also have a history lab at Colgate also supported by Data & Society with stone foundation funding, which is a handful of people. I think Kevin's in the talk. Hi, Kevin! I needed a team and I

liked having a team too. It was fun to have people to work with who would, who could wander through just hundreds and hundreds of sheets, sometimes counting and tallying things. Sometimes just looking for anomalies, sometimes like taking a data set. So I went to Dartmouth and got all of the letters written to Senator Charles Toby who was against the income question. And then we went into the census records and tried to find as many of these letter writers as we could to see what they said and what they did in the census. And that was enormously time consuming work. I relied a great deal on the insights and work of my history lab,

**Alex Hanna**

12:56:15

To conclude, I'm not a trans historian, I'd love to be one but I'm not, I take inspiration from historians. I do think there are folks who do great work, and thinking about the work of what reading a data set looks like. I'm going to call out Lindsay Forrier who has done some amazing work, and one article I think that she's written on teaching in getting students involved is called ethnographies of data sets and it is about how to do critical analysis when you have the data. I dropped it in the chat, I don't think it went to everybody, I think Rigo can share, sorry. But it was a way of interrogating this both in terms of doing this from a quant perspective, but also how to take a step back and read it more as a text. I just want to give a shout and get some people and students and folks involved.

**Ronteau Coppin**

12:57:36

Thank you for that. Dan, I wanna turn it back over to you to give us final words on Democracy's Data.

**Dan Bouk**

12:57:45

Thank you everybody. Thank you, Alex, for showing up. Thank you, Ronteau for guiding us through asking these questions. Thank you everyone else who's been here with us. Thank you, Rigo for taking all of the things we say, and then translating them into links that people can use and actually find these wonderful resources. I wanna close with an inappropriately upbeat note. I'm just gonna do this because I like to sing. So I'm gonna sing you a song. The tune will be familiar to some of you, especially if you grew up in upstate New York, the tune is the Erie Canal song, and here we go.

I've got a book printed on mashed up tree it's Democracy's Data published by MCD. I've got a book printed on pumped up wood. It's a book, it's a look, at bureaucratic personhood. A book about the census might sound boring, filled with numbers far from stirring, but about this story, well, I must say, of data democracy and dignity. Census, everybody counts, census, so the nation sees itself, and you'll maybe find your neighbor or an ancestral pal in the lingering traces of a datafied tale.



**Ronteau Coppin**

12:59:21

I love it. Thank you. I love it. Uh, thank you for joining us for today's book talk and thank you, Dan and Alex, for a wonderful discussion. Democracy's Data launches on August 23rd. I encourage you to pre-order the book.

Data & Society's Book Forum Series provides a platform for scholars and researchers to present their work, frame key debates in the field and gather feedback from a peer network of thinkers and practitioners. For future programs and opportunities, make sure to sign up for our newsletter, have a wonderful day.

**Alex Hanna**

12:59:58

Thank you so much.