

All Data is Local featuring Yanni Alexander Loukissas

March 3, 2020

Good afternoon, everyone.

Welcome to Harvard Law School and the Berkman Klein Center Lunch Talk, a longstanding tradition, really an institution at BKC. A couple of things I want to say, first of all, I'm Matthew Battles. I'm Associate Director of metaLAB at Harvard, which is a project of the Berkman Klein Center. And you'll hear a little bit more about metaLAB because Yanni was my colleague there for a couple of years a few years ago.

Housekeeping matters, first of all, the event is webcast, and it's recorded. So you'll want to bear in mind your own telegenic personality when you contribute to the conversation. And if you want to contribute using Twitter, you're welcome to point tweets at the center @BKCHarvard. So that's the housekeeping.

And just to give a little bit of a sense of run of show, Yanni will be talking for 20, 30 minutes. I'm going to take that microphone privilege to ask him the first couple of questions. We'll have a little bit of a conversation, and then we'll open it up to the room. So with that, I want to welcome back to the Berkman Center, as it was when he was here, Yanni Loukissas. Yanni is on the faculty of the School of Literature, Media, and Communication-- I think I got the order of those right-- at Georgia Tech.

[INAUDIBLE].

And he runs there a group called the Local Data Design Lab. And Yanni's going to be talking to us about a book that I've been excited about, as I've been aware of its germination over the last few years, his new book, All Data Are Local I've been excited about it in part because a couple of the cases that he approaches in the book, he began to develop this research when he was with me at metaLAB.

And I say "with me" because for the first couple of years we were together, it was really Yanni and myself in an empty room trying to figure out what a metaLAB would be. It emerged out of our kind of blended practices of attention and focuses interest to blend design and art practice with a really critical attention to the way data, the way digitally, the way technologies and platforms are lived in situated contexts in our neighborhoods, in our cities, in our institutions, in our families as well, as well as how they penetrate, as you'll hear from Yanni, when he talks about the Arnold Arboretum, the natural world as well, and the city, and urban structures, and cosmopolitan contexts.

That interest, I think, in the city probably began for Yanni with his training as an architect at Cornell. Yanni went on to study Design and Computation and do his doctoral work in that program at MIT. Following on that, he spent time in postdoctoral study with Sherry Turkle and David Mindell in STS scholarship, always grounded in lively, improvisatory mix of design practice, technology making, engagement with communities and scholarship.

And so I'm really thrilled to have a chance to share with Yanni the new book *All Data Are Local*. I think it's a study, but it's also a kind of field guide to data in the wild and a provocation as well for us to think about data in contexts beyond mere openness, to think about the conversational, the human dimensions of this thing we call "data." And so with that, I want to turn it over to Yanni Loukissas.

Thank you. Can everybody hear me?

Yes.

In the back? Well, thanks for all being here, and thanks for that incredibly generous introduction. It's really a pleasure to be back. I spent two wonderful years here at metaLAB in the Berkman Center, and the book really got its start there. And being back here to talk about and share with you all is tremendously exciting, and I look forward to having a discussion with you. I'm going to give about 25 or 30 minutes of remarks, give some examples, gonna ground us. For those of you maybe haven't taken a look at the book, which I'm sure is most of you, and then we'll get to some discussion.

So I'd like to think of my work as bridging or bringing together two areas of study. On one hand, the work of data visualization, which means, how do we graphically present evidence about the world we live in? And data studies, which tries to look for, actually, the limits, the assumptions, the implicit values, and often the absences in those data, and kind of contend with them. And so when we bring those two together, I think there's all kinds of potential to rethink how we do visualization to make it more effective and more just.

My focus is really on public data, so data that's kind of out there and accessible in the world. But I think we could be-- it's not enough for data to just be open, and I'll explain in the following examples what I mean by that. I'm really interested in how we can look at public data in new ways. Oops, jumping ahead a little.

This is a book that I began writing-- so when I was at metaLAB between 2012 and 2014, now looking back on that era, particularly in terms of data, it seems like another time. Big data-- specifically, were on the verge of transforming how we do science, how we do government, how we do design. And a lot has changed, and that perspective seems a bit quaint now, right? A lot has changed for me personally, as well. I left metaLAB. I moved to Georgia. I had two kids. I'd like to say that probably shaped the book as much as anything else.

But societally, data and algorithms have taken center stage. And I'm sure you've all seen headlines like this. These are actually from quite a while ago, but they just kind of keep coming. I think societally, we've embraced the fact that data contain biases, unanticipated impacts, there's

even blatant misuse around data, right? And data can be racist, data can be sexist, or it can be simply fake.

And the examples run the gamut, from p-hacking to election hacking. But I don't think we're ready to give up on data as a way of knowing about the world. So my inquiry is really about, what does it mean to work with data, knowing that data have all kinds of limitations in ways that are both ethical, and effective, and just, actually?

So the way that I see forward for this kind of work really actually builds on a long tradition of scholarship. And Matthew started to kind of take us into that, but there's been work in STS, in science and technology studies, in feminists and post-colonial critiques of the way science gets done and the notion of objectivity-- people like Sandra Harding, Donna Haraway, even more recently Anita Chan have helped to show us that data are situated like any other way of knowing.

I think one difference I take to a lot of that work is often when we talk about situated, we think about being embodied and who's making data, who's using data. But I'm also interested in where we're using data, and I think place is a crucial piece of what we're missing when we talk about and think about what data are, and what they're doing, and how they're used in the world. And I'm very interested in how data get made differently in different places and why that may matter.

So here is the cover of the book. And I think the title pretty much sums up my argument here. All data are local. Now what I mean by that is nothing more than that data are made by people, often with their dutiful machines in a time and a place, using the tools at hand within existing organizations and, crucially, often for audiences that are conditioned or disciplined to receive them, right? So data are not just facts. They are cultural artifacts, right?

And when I use the word "local," I mean that in the relative sense, right? So local can be local to a neighborhood. It can be local to a network. It can be local to a city, to a region, to a country. Local can be local to the earth. So local is just a way of suggesting that there is actually something beyond that domain. And I think with each data set, we have to kind of redraw that boundary and figure out what are the conditions and the context in which we have to understand that data and what they're doing. And so in a way, local is defined by difference. It's different from other localities, right?

So let's talk about an example. I like to use this example. This is actually something that I kind of encountered early on in my work with metaLAB. We were doing a lot of work with Digital Public Library of America, which-- how many people are familiar with this project? A lot of people here. Maybe not as many as I thought, but it was a Berkman project at one point. It was a project to bring together the digitized collections from libraries, museums, and archives across the country and create one point from which you can access all these collections.

And one of their biggest collections was from the New York Public Library. Here's an image of its facade along Fifth Avenue. Now this is a place-- it's a destination. It's one of the biggest research libraries in the country, if not the world. A lot of people like to hang out on the steps here on Fifth Avenue.

And I was doing a project-- after I actually left metaLAB, I was at Georgia Tech, and I was still kind of grappling with this collection. And what's interesting about the DPLA is it kind of brings together data sets that were created in different places, brings them together into one place and allows you to have this kind of comparative view of those data.

And we were digging into the collections of the New York Public Library, which has books, but also newspapers, maps, photos-- I think they even have Winnie the Pooh in their collection, and they have data that describe these objects. And we came to this question-- we were very interested in-- we started to discover through this collection that it's not just that data get made differently in different places, but even in a single institution and a wealthy institution like the New York Public Library, you have this incredible heterogeneity in data. And we tried to simplify that into one question.

You know, when you look at this, you might think of a date, right? And depending on where you're from, it might be the month, day, year or day, month, year. And we wrote a little algorithm to try to figure out how many different ways are there of writing the date in a collection like in New York Public Library. And this is their online collections that we're examining. But does anybody want to hazard a guess? How many different ways are there of writing the date that a institution like this has in use?

20.

20, that's a good guess.

Depends on if you're going beyond English, it's a lot more.

That's a good indication of what's to come. Anyone else?

100.

100. OK, so we found 1,719 date formats. And that just has to do with how we wrote the algorithm. If we wrote the algorithm differently, we might have gotten a different number. I'm going to start to zoom in to give you a sense of some of these. And the number in red is the number of times that format is used.

And this is a collection of 800,000 objects, and it's actually just a small subset of the larger undigitized collections of the New York Public Library, which are about 50 million objects. So there are a lot of objects here, but there's a lot more difference than we would expect. And when you look at this stuff, sometimes there are Roman numerals. There are other languages. Sometimes the name of the printer is put in there. There's a lot of approximation.

So there's the sense that for each of these objects, we might have different degrees of knowledge about when they were made. And they're very human, right? You can imagine people creating these with kind of limited knowledge, and time, and so forth.

And usually, when we encounter-- so in the work of data visualization, often you get these heterogeneous data sets, and this kind of standard procedure is that we've got to clean the data, which means, essentially, normalizing it, trying to find the kind of common elements, and then getting rid of everything else and cleaning-- this term "cleaning" suggests that all that other stuff is just dirt and needs to be gotten rid of. And the way that I would like to reframe this problem, I think, comes from an anthropologist named Mary Douglas. She wrote a book called *Purity and Danger*, in which she asserts that dirt is not an objective category.

It's relative. So all of you are eating lunch, and maybe there's some sauce on your sandwich or something like that. That's not dirt, right? But if you were to spill it on your [? hand, ?] then it becomes dirt. And Douglas says that dirt is nothing more than matter out of place.

And I'd contend that we often think of as data dirt is really data out of place, data from another context that we don't necessarily understand. So in order to make sense of those data, we have to dig into those context and try to understand data as part of a larger system of knowledge that it's part-- data don't stand alone, right?

So often, we think, when we use this term "data set," we're talking about something that is portable, it's complete, it's closed. But none of those things are really true about data. And I prefer the term "data set." I think it kind of draws our attention to places like the New York Public Library, and its associated histories, and collections that it's drawn from that can help us get closer to and start to understand what these data mean within a larger system of knowledge.

I'm going to talk about a couple more things, just to kind of ground this example or show the implications of this example. First, I want to talk a little bit about, what do I mean by data? I get that question a lot. And I want to talk about the stakes of the book in terms of what Anita Chan has called, "digital universalism" because I really see this book as trying to challenge the idea of digital universalism. And then, I'm going to talk about the principles of the book. The book is organized by six principles of local data. And then, I'm going to end with kind of an evocative example that's from more recent work and tries to propose, if we take up these principles, what might it look like to work with, to explore data sets in a local way?

OK. So Christine Borgman and Michael Buckland have this great way of talking about data. They say, similar to the idea of dirt, data is also a rhetorical category. And if you start to look at varied data sets, you'll notice that they kind of don't share a lot of essential qualities. They assert that actually asking what are data is the wrong question. You should ask, when are data? Because anything can become data if it's taken up as part of a claim.

And Borgman's book, *Big Data, Little Data, No Data*, she writes that, imagine you find a box of photographs in your parents' attic. That's not inherently data, right? But if you start to make claims about it based on those photographs, maybe about your family structure, maybe about the clothing trends of the day, maybe about the mechanics the cameras used to make those photographs, suddenly, they become data. And I'm sure we can think of similar examples from elsewhere.

And I think this is a very evocative notion, but I think it misses something, and something captured by another theorist of science and technology, Bruno Latour, who has this cute phrase to talk about data [INAUDIBLE] inscriptions. He says they're immutable mobiles. And that means just what it sounds like. They are things or representations that we expect to move anywhere, and they don't change when we move them.

And I think you could kind of relate to that when you download something from the internet. You don't expect it to change on its way to your computer. But the truth is that data don't actually move easily. Or when they do move, they're difficult to interpret. And there is a kind of friction that we need to become more aware of.

So I'd like to think of data as allegedly mobile evidence. It's evidence that we think should move without changing and we expect to, but there are all kinds of problems we run into, and my work is really to try to kind of address and contend with those issues.

I want to say something quickly about my methodology or my approach to this work. I am not an anthropologist. I don't think of myself as an ethnographer in the traditional sense, but I like to take on what Sherry Ortner calls an ethnographic stance. And that really means that I'm using-- the ethnographic stances attempt to use the self as the instrument of inquiry.

So the book is really about my kind of journey through various data settings and attempts to understand them through a combination of just reading the data, making visualizations of them, interviewing people involved in creating those data, observing them at work-- so through a variety of what I think of as kind of participant observation techniques. And I sincerely believe that we don't all need to become ethnographers to have this kind of more culturally grounded understanding of data, but it requires curiosity and an openness to connecting with people around data.

OK, so let me say something about the stakes here. I mentioned this term "digital universalism." Now the idea of digital universalism, as Chan outlines it, is that it's an expectation by digital systems that we're all just users. And it doesn't matter where we are or who we are, that we're all the same in the face of these interfaces. And in her book *Networking Periphery*, she really challenges that.

But what I see here is a kind of underlying, what I call, place agnosticism in digital systems. And this can be found in the work of a lot of theorists, going back to Marshall McLuhan, who talked about electronic media collapsing space and time to people like Nicholas Negroponte, who famously wrote being digital means less and less dependence upon being in a specific place at a specific time. He suggested that place itself was going to be transferable. And I think these are inherently problematic statements that we need to challenge. I'm going to read you, actually, a little from the book to give you a bit of a sense of the tone of the book and explain what I think of as the stakes.

So I write, "The diversity and prosperity of the world's varied and contingent digital practices depend on our acceptance of data's locality. In fact, the stakes for the future of the internet could not be higher. If left unchallenged, digital universalism could become a new kind of colonialism

in which practitioners at the periphery are made to conform to the expectations of a dominant technological culture. Learning to look at the local conditions of data can be a form of resistance to the ideology of digital universalism and the threat of erasure that opposes [INAUDIBLE] data cultures.

OK, so having said that, I think we also need to be really careful that we are not romanticizing the local, right? And I want to tell you another quick story that actually came out of a workshop that Matthew and I ran together called "Beautiful Data" together with metaLAB. We actually hosted it a couple of times.

And one of the participants, Maria McWhorter was a historian and a curator at the Smithsonian. And she told this incredible-- Maria is an African-American woman, and she told the story about during her time at the Smithsonian, doing a search of the collections for the word "black." And what was returned was a number-- and I worked with her to try to recreate this search a little-- a number of artifacts in the collection that were labeled as made by black artists, or coming from black culture, or, in some ways being associated with blackness as an identity.

And then, she searched for white and whiteness. And she came up with stuff like this. Plants that had white flowers, maybe the name white attributed to an artist, nothing to do with racial identity. And what she took away from this was-- which I thought was incredibly powerful-- was that in its absence whiteness is represented here as the default. And I think this goes beyond when we talk about bias in data. This is really data being shaped by invisible values, assumptions, a kind of ideology of white supremacy, in fact, that McWhorter is calling attention to.

And this kind of ideology, these kinds of absences, are only visible through comparative analysis. And I think when I talk about locality, that's what I mean. It allows you to kind of access this kind of comparative perspective on data. Clifford [INAUDIBLE] wrote a lot about local knowledge, liked to say that we can't understand the local in a relationship to some imagined universal, but only in relationship to some other locality.

OK, so these are the six principles. I'm not going to go into them in depth. I hope they will pique your curiosity, and these examples will kind of invite you to check out the book. I'll just mention briefly-- so, of course, the first principle I've already talked about, the second principle data have complex attachments to place-- came out of work that I started with Matthew at the Arnold Arboretum. The Arnold Arboretum is part of Harvard University, out in Jamaica Plain, incredible place I highly recommend visiting. It's a place where you can really understand data's relationship to this idea of place, how places shape data in certain ways, how data shape those places, in turn.

And so the book is really trying to come to terms with what do data look like from a particular place, and what are those sometimes unexpected relationships that we might find? And that's kind of grounded in data on plants-- trees, vines, and shrubs there that they've collected for over 140 years, and they have kind of an extensive data set.

The next principle data are collected from heterogeneous sources, turns to this case of the Digital Public Library of America and ask, what happens when you bring data from different places together in one place? And what does that reveal? And what kinds of place attachments are retained in the process? And how can you discover those through visualizations that highlight the classifications, the schemata, the constraints, the errors, even the absences and the rituals that might go into producing those data because those data very much hold all of those elements of their formation.

I have a chapter on how data and algorithms are entangled. These days, a lot of the focus is on algorithms, but algorithms aren't really-- can't be disentangled from data so easily. Data are made to be recognized by certain algorithms, and algorithms are made with the expectation that they're going to interact with certain kinds of data. So there's a kind of deeply historical relationship there, and I examine-- I do an examination of natural language processing algorithms, which I'm sure we've all heard about and how they're grounded in historical texts that have a lot of assumptions about what we mean by natural when we talk about natural language.

And then, the chapter 5 is about interfaces and how interfaces actually, first, decontextualize data, kind of take data out of the context in which they're made, and then create new contexts, new settings in which the data are meant to be fully understood. And I use the example of Zillow, which is an online housing marketplace maybe some of you are familiar with to show how interfaces do this work of contextualization. So context is not just about where the data were made. It's also about where they're used and how that usage is kind of shaped and set.

And then, the summit of principle has to do with recognizing that data are really no more than indexes to local knowledge. Now what do I mean by that? Well, if you were to look at the index of my book alone, it's kind of like a data set. You could probably learn a lot about what's in the book, right? But it's infinitely more useful in combination with the book itself. And that's kind of what I'm asking people to do, to look at data as really a starting point for further inquiry and understanding. And look, what are the deeper sources that data can point us to? The sense that data never tell the whole story.

OK, I also go into a variety of principles. This is a book that's meant for people who are interested in understanding data in more depth, but also for designers, people who want to make you do systems around data. And I'm not going to talk about these explicitly, but rather I'm going to show you one final example that I think tries to bring together [INAUDIBLE] a number of these principles called the Map Room Project.

So the Map Room is a collaboration with Jer Thorp, a digital artist. And together, we've been trying to build a network of local spaces for mapmaking and data exploration that allow people to explore data in relationship to their own lived experience by trying to make maps of the places they live that kind of combined different sources of knowledge together. It's a project that doesn't assume that data speak for themselves, but rather that data are understood in the context of other kinds of knowledge.

And we want to understand what context means in different places and different situations. And these are spaces really for collaborative data exploration, so there's inherently social spaces and

creative exploration. So we want to actively engage people in interpreting and thinking about what-- and sometimes challenging what data seem to be saying about the places they live.

This a project that we use with students, with different kinds of neighborhood groups with planners. We're talking about doing some map rooms and libraries. Here, you can see some students using the map room. It's quite a simple technology. There's an overhead projector, which guides students in drawing maps by hand. And the projection is the students or participants select an area that they want a map, and the encouragement is that people are actually mapping places where they live or that they know in an intimate way.

And then, as they draw their kind of paths, and where they live, and issues that they're concerned about, then they can turn on different data layers and see their experiences in relationship to data sets. And those might be data from census. They might be city-made data. We've experimented with a lot of different things. This is a kind of deeply inherently accessible system. It's not one that you really need to be trained on. And as I said, we've worked with a variety of audiences. Over here on the left, this is a Dr. MacLaine, who's an environmental justice activist in Savannah, who we were working with on issues of coastal flooding and air pollution in neighborhoods in Savannah.

And then on the right, you can see a young kid at a pop-up version we did in the Atlanta planning office. And so people show up to these events, and they draw, and they draw, and they talk, and they reflect on what data-- how data seemed to be representing places they know well. And they create these shared representations.

So one version of the map room is in my lab at Georgia Tech. It's for making pretty big maps. So these are paper maps. This one's 16 feet long. And I won't go into kind of detail, and the apparatus, and how it works, but there's an overhead projector. It's on a rail that moves. You can see these are hand-drawn maps. Because it's paper, you can also incorporate all kinds of media, photographs, other kinds of images. Here's a technical diagram of how this version works.

And we hope that this is a new kind of technology that helps people to examine data in new ways, and challenge those data when thinking about what those data might be leaving out, and they know about places the places they live.

We have a more mobile version that we've used extensively. This was in New Orleans. Here's the kind of maps that people make. You can see that people are leaving with traces of their hands. Everybody has their own way of drawing. You kind of can't make a mistake.

And then, we can overlay different kinds of data layers. Here, we were doing a lot of experimenting with data on gentrification, changes in median income, and the percentage of college-educated residents, and race, and percentage of occupancy in different neighborhoods around New Orleans. Here, you have-- this is a class of high school students in Savannah working with climate scientist Kim Cobb on making a map of their neighborhood.

So I'm going to kind of leave it there. I think the takeaway, really, is thinking locally can be a way of thinking critically about data. And we want to continue to explore through vehicles like

the map room, what does that mean, and what do these local contexts-- how do they shape the production and use of data in different ways?

I've had a lot of students collaborate with me on these projects, on the book, and on the Map Room, and I just want to acknowledge them. And I want to leave you with one thought. So while I was writing this book, I would think a lot about, who's going to read this?

And often, my students would come to mind as the first audience. But as I said, I also had a couple of kids while I was writing this book, and it kind of weighed on me. If my kids read this book someday, what do I want them to take away from it? And I think it's this. And it's a message that I think should appeal to a broader audience.

As you seek to explore the world, data can be a wonderful starting point. They can be a bridge to get closer to the people and places beyond data. But let's not take the availability of data, the easy access to data as permission to remain at a distance from those people and those places. Thank you.

[APPLAUSE]

Thanks, Yanni. And can you hear me? Yes, I can hear myself. So as I threatened, I'm going to take the opportunity to ask Yanni a few questions and pick up a few points that I'm intrigued by and that have been part of our ongoing discussion for years. And then, we want to open the floor. I'll share my microphone down here, and we have another microphone that we'll be able to pass around up in the back. And so we'll try to get coverage and get all voices engaged. Thank you, Yanni. It's really, again, thrilling to see the book is out and beginning to build its audience, and connect, and find its uses.

I wanted to think about with you a little bit that ethnographic stance, the new instance at the top of the discussion, and, in particular, the reflexivity of the project that you referred to, that you're thinking of this question of the locality of data, which shaped not only by encounters with collections data in the context of some of the metaLAB projects you described, but also as you made a life transition, had children, moved to Atlanta. And I think when you have talked about the work-- the case study of Zillow in particular, you've really brought that to life, that question of both grounding an experience of a vital and anxiety-producing platform that many of us are familiar with when we're looking for a place to live or thinking about other people looking for the place where we live and both the sagacity you bring to studying the data and the way they express in the interface on Zillow, but also how that impinged on your life as you moved into Atlanta and became aware of the community.

Mm-hm. So, wait, just to summarize, so how does this ethnographic stance, how is it formed by some of these changes? Yeah, so certainly, I wrote this book in the first person. It's interesting because some early reader responses-- you know, the publisher sends out the book for review, and they said, you know, this book would sound a lot more authoritative if you did write in the first person. I said, well, that's kind of the whole point of the book is that I didn't want to take this authoritative stance and rather say what my experience was with these data sets and understand that others may have very different experiences.

And I brought certain tools to the encounters, and others bring their own tools. And, certainly, being at metaLAB and being at Harvard gave me access to incredible data sets or data settings, I should say, that really kind of helped to spark the inquiry. I happen to live very close to the Arnold Arboretum.

I tried to take mine proximity to these settings as a reason to kind of understand them and my relationship with them. And I didn't want to study settings that were very remote from me. I think the whole point was not to try to use data as this place-- as this way to know about something at a distance in another-- a place that I had no other experiences.

And so moving to Atlanta and having to find a place to live, I mean, of course introduced me to Zillow. It's, I think, one of my favorite chapters in the book. I didn't talk about it a lot today because it's quite involved, and it takes-- explaining it thoroughly is not easy so I encourage you to look at it in the book.

But you know, Zillow is this interface, this platform that proposes to let you learn about real estate markets across the country without being there. And I was interested in this question of, what are the implications for Zillow in Atlanta where I was and not thinking about this as a kind of generic interface that meant the same thing for every place but realizing that places had a specificity to them.

And Atlanta, which is a minority majority city, that is facing enormous gentrification now, I landed in the city, and somehow I had to contend with what role I was going to play in that. And that's how I really started to grapple with Zillow. And I think it really helped frame my first couple of years in Atlanta and really interesting-- and I think use powerful ways. And I got involved with housing and justice activists in Atlanta because of that. And so I think this has been-- those kinds of encounters have been really informative to me.

And I can see how the Map Room project in particular helps to make that connection happen and make it lively. And I wonder-- I mean, it seems like it's a really kind of thick context for engaging in these questions of data and place with folks and an opportunity to learn how to contend with data, whether those are civic data, or environmental sensing data, or census data as you describe, and to grapple with how those data are commensurable or incommensurable with your own experience of neighborhood of city, of community. I wonder if you'd talk a little bit about what you see the impact of that work being.

Of the Map Room work?

Yeah, yeah, what can people take away.

Yeah, yeah, I think it's a good question. You know, I was asked by a colleague once, so is the Map Room an alternative to Zillow? And I say definitely not. But I see it as potentially an antidote. It's a place where we can learn to think critically about data by encountering those data in a setting that we know well in relationship to a place we know.

Zillow is based largely on tax assessment data. I interviewed the local Fulton County tax assessor in Atlanta, and he was quick to tell me about how error filled that data is and, in part, because that data-- those tax assessments are expected to be part of a social process where you're sent the assessment on your property, and then you can respond and say, I think this is wrong. And it's meant to be a kind of iterative exchange with the county. And data really kind of takes that context away.

And for Zillow, they don't have to be right. Their business model is not based on being right and having accurate evaluations. And actually, they say that their estimates-- they call them "the Zestimates--" are only within 5% of the sale value 50% of the time, which harsh critics have said is close to a coin toss. But they're just interested that you keep coming back to the site to check on, oh, how is my property value potentially changing?

And that, I think, feeds a lot of anxiety about the market and, obviously, we're in a situation today where the housing market is out of control. And platforms like this are only exacerbating that. But I think the Map Room, I really see it as, in part, an educational tool. And so we're setting up Map Rooms in high schools and, hopefully, a library nearby, and museums and so forth, but also a tool that communities can use to advocate for themselves and try to represent their neighborhoods in more encompassing ways that maybe all the data available still don't deliver.

Yeah, so one more question about the local, and then I'll turn it over to the local audience. You have. Your list of student contributors and collaborators. And I wonder if you'd talk about how students and students in Georgia Tech in particular, maybe, take this paradigm up. Because I'm thinking, these are students who are probably thinking very instrumentally or are encouraged to think very instrumentally about what data are and can do in the world. How do they take it up? How do they respond?

Yeah, I have to say, it can be very challenging because what I'm proposing makes their work more difficult. And I'm asking them to-- a lot of students who work with technology do so because they don't necessarily think they're that working with people is their forte. And it may be an escape for them. But what I insist is that in order to work with data, you have to work with people.

And so, as an example, I often give an assignment where I ask students to go out and interview somebody who is involved in-- well, they have to pick a data collection they're going to work with, and they have to go out and interview someone who is involved in making those data, or works with them on a regular basis, or maybe is a subject of the data-- but someone who can kind of add nuance to their understanding of the data set. And they're always resistant. Like, oh, it's going to be so hard to find someone. How do I find someone?

But they always do, and they always come back with their understanding that it totally transformed, which is wonderful. And I hope they will continue to do that, but I think it takes a lot of work. And I'm sympathetic to students not wanting to do that. And, yeah, it's hard for all of us, yeah.

Well, now, I feel inspired to find someone.

OK.

So what are your questions, your responses, your thoughts? When I hand you to mike or a microphone makes it to you, it's on, just to let you know. Who'd like to jump in? Vince.

Yanni, thank you for that talk. I think this is fantastic, and I very much appreciate your emphasis on place and on thinking of data settings rather than data sets. As a historian, I work as a historian, I'm equally interested in time. And I wonder if you'd talk a little bit about time and the accretion and accumulation of data sets. You mentioned that we should also be asking when the data sets are created and what settings. Would you talk a little bit about how you think about the accumulation of meaning over time when information gets turned into data?

Yeah, thank you for that question. And I should say Vince Brown's work is in the book, featured in the book also. And I hope you'll check it out and that it will inspire you to go look at his larger corpus. But time, of course-- and I'd like to say, no place exists out of time. So it's not as if we can ever look at a place independently of time. And similarly, we can never think about a time independently of a place in which we're kind of examining. So I think, in a way, those are-- we can't disentangle those things.

And often what I'll ask students to do is when they're working with new data is I'll ask them to investigate or ask someone what recent changes have there been to the way that these data are made? And I think that can be incredibly informative. You know, if you look at what's happening with the census now, this notion of whether we're going to have the citizen question or not, data are always changing. We're always rethinking how we make data. And I think that systems that assume that we are going to have this fix or we're just going to find the best format as if it's kind of independent of everything else that might be going on is really wrong.

You know, there's a great book called *Sorting Things Out* by Bowker and Star, and they write about data as kind of-- that data are lively. They're kind of transforming all the time. And we really have to have this kind of facile and light approach to thinking about data.

And furthermore, data are always historical. We don't necessarily think about data as historical artifacts, but all data were made in the past. And a lot of the work we're doing now with machine learning requires data to train systems. And then, we're surprised when the data have all these artifacts from the past. And in this country, we have a long history of sexism and racism that is embedded in those data. So in a way, we can never really be forward looking if we're data driven.

Thank you.

So my question is very practical. Having been recently user of Zillow, these platforms exist, and we use it.

Yeah.

And so how can we then design them in a better way or make something different available? Because we can't [INAUDIBLE] people using that. They will more and more need those kind of tools to find a house, to bring our kids in a certain school. So what's next? How can we bring this approach into the design process and redesign platforms and tools in a way that really changed the behavior of people?

Yeah, that's such an important question. I think the short answer is, I don't know. I don't necessarily see it as my job to completely redefine the way we design interfaces, but rather I want to present challenges and new models of working with data. You know, this isn't something that I'm going to do on my own. It really relies on people taking risks and experimenting with new ways of relating to data.

And instead of having a platform like Zillow, which is explicitly non-local and tries to give us this view from a distance, I think we can start experimenting with models that are meant to just deal with maybe even just one city. What would that look like? How would a Cambridge-based housing application look?

But also, I think it's a kind of cultural shift that I'm trying to-- I want to be part of, that maybe that platforms like Zillow don't go away, but people approach them with more skepticism and more critically-- so that they understand that they see a data point, and they don't expect that to be a fact. And they actively look for ways to kind of get closer, get on the ground, understand who's living in these neighborhoods where I'm looking at houses, how are they going to be affected by my move?

I think a lot of people don't realize if they enter a bidding war and they pay a lot more than the asking price for a house, they actually raise the taxes of their neighbors. And if your neighbors are fixed or low income, that may mean they're displaced. So I think getting closure means understanding things like that and understanding the implications of your work with data and that a platform like Zillow affects not just users, but also non-users.

Yeah.

Couple of questions over here. Anastasia.

Thank you so much. I'm wondering, so yeah, I think you started talking about, essentially, my question. Even when data are created historically or these platforms are kind of-- create their own data sets without any understanding of place or time, they do end up affecting very local issues. I'm specifically thinking about Google Maps examples when real estate developers would decide on their own four-letter acronym for a neighborhood entirely change like the look of it. Anyway, so I was just wondering if you do any work on that side of things, take data sets that are, essentially, fabricated by those in power or the people behind the platform and then examine the effects on the local populations.

So that's a good question. So I think there's lots of-- if we take a media frame for thinking of data, and we say that we can study how data are produced, we can do kind of a critical reading of data, and we can do a study of how those data are used, maybe folks in media states might call

that reception studies. My book focuses on the first two primarily, trying to understand why do data look a certain way? And how did they register the ways of knowing, the values, the assumptions of the places and the systems in which they're made. And how does reading those data closely help to reveal some of that?

And the Map Room is more of an effort to focus on the third. Because I do think by the time I get to the end of the book-- you know, you get to the end of these kinds of large projects and you realize what you haven't dealt with, and then that usually becomes your next project. And so the Map Room is really about understanding what it means for people to make use of data and in an intimate way, in a social way, in conversation with one another, and about places that they really care about. And I'm just really getting started on that work to see, how do people put data into context?

I really take Borgman's notion that data are rhetorical to heart and studying what people do in The Map Room. Anything can become data in The Map Room so we have a lot of data layers that people can turn on that are what we might think of as official or institutional data sets, but sometimes people reject those or neglect them. And instead, they try to put their lived experiences on the map, and those become data that they or others use to make claims about the places they live.

So I think you point to a really important issue and one that I'm really just getting started with examining. And I think it's incredibly difficult and hard to say in, I think-- it's hard to say in a generalized way how people use data. I'm expecting that's going to differ from one place to another, and I'm looking for what those differences are kind of attributed to. Yeah.

My question is sort of related to that. The city of St. Louis has had a big population exodus in the last half century. So abandoned properties are a large issue for them. And the city produced a bunch of open data about what they thought was an abandoned property. And the civic data group in the area produced a different data set. And the city's definition of what was an abandoned property was roughly who hadn't paid taxes on a property in X years.

And the civic data groups was roughly what buildings don't have roofs, because that's what most people think of, I think, on the ground. So they kind of engaged in this negotiation process for how to actually classify things. I'm wondering if you have any other stories that you think illuminate interesting aspects of that negotiation process, where there's kind of a imposed meaning for data that doesn't match experience.

[INAUDIBLE].

Yeah, that's such an interesting example. And I think it points to, again, the frame through which these different groups are seeing those properties, you know? If you're standing in front of the property, what do you see? If you're looking at tax records, what do you see? And the data are shaped by those perspectives, those standpoints.

I think one thing that came out-- this is a little different from the example you gave, but one of the context we've been using in The Map Room is in Savannah, I mentioned. And we did a

mapping exercise with residents in a neighborhood called Hudson Hill. And The Map Room, one of the partners in the Savannah project is this climate scientist Kim Cobb and her team, which includes data scientists and some local city officials who are trying to build a new network of high-resolution sea level sensors that are going to help us understand how flooding happens in heterogeneous ways across that region.

And so you know we brought that data-- we had that data layer, we asked them to make this map, and their response was, yes, we know that sea level rise is a problem, but we actually have more pressing-- what we think of as more pressing problems that we want data on, and they were particularly interested in-- this one resident said that she'll come out into her front yard some days and they'll be white, particulate matter all over her car and that she thought that was coming from a paper mill nearby. And so I said, how would you draw that on the map? And then that led to a discussion about kind of strange smells that people encounter in this neighborhood.

And all this led to now this-- the sensor project actually adding air quality sensors to that system. So I think the Map Room kind of opened up this conversation. I think what was great about that use of the Map Room is it wasn't narrowly limited to these data, and interpreting these particular data, and the place in relationship to them, but rather kind of cast open this web of possible sources and evidence that could help us understand and think about what's really important for this neighborhood and led to other data collection practices. So I think that makes me hopeful for the map room, as a much more open and grounded way of interacting with data.

Some of these comments have made me think about the humanity of data and how easy it is to divorce stuff on a screen or stuff on a piece of paper from the humans in the real place and made me think about something that's at least contentious in my life, and that's a study of a parking garage near my house. And what they said was, well, first of all, even though the real people who use it say it's full, they said, no, no, no, not at all, tons of room.

And not only that, they actually made up parking lots in the area, claimed they were real and said, besides, they could go to these places that don't exist. And not only that, we can simply manage it. We can just tell people they can't go there, and then there will be room. And it's these people sitting there, just massaging numbers.

Yeah.

And the other thing that someone else brought up about it was, well, did anyone go and see who actually uses the garage now? Well, you look and right now, it's a bunch of beat-up cars. We're planning to put, like, beamers and stuff like that in there.

Yeah.

So it really made me think about the real dangers of forgetting the humanity of the things that you're looking at.

Yeah, yeah. I think that's so important that we think of data as either something that just exists out there in the world-- I mean it comes from the Latin root means "given," the givens in the

world. But data are made by people or the machines that they program to collect those data. And they make certain decisions about how they want to-- what counts as data, and what doesn't, and what the shape of those data are.

And another exercise I give my students often is to go out and collect their own data. And it's incredibly frustrating for them because they realize all these decisions they have to make about, oh, well, what's going to be in the data set? What's not going to be-- a lot of students want to use photography as part of their data-making process, and then they have to, well, how far away from the subject am I going to be? And what angle am I going to be at? And all these things. And I think they come away with this sense of the humanness of data, as you say. And it's an important lesson. So thanks for that.

Thank you all for those questions. And Yanni, thanks so much for being here with us and talking about your book and your practice. And I want to remind everyone that the book is available for sale. I couldn't recommend it highly enough. Take it, use it, share it. Thank you, Yanni.

Let me know what you think. Yeah, thanks.

[APPLAUSE]