Episode 3: Open Data and Open Source Technology

I'm Kevin Chambers, Technologist with NCMM, The National Center for Mobility Management. And this is Conversations with Leaders, a podcast funded by NCMM. Check out all our resources, aimed at mobility managers, at nc4mm.org.

Today we have the first of three segments where I interview Carol Schweiger on three recent white papers she authored on open source software and open data, Mobility as a Service, and automated vehicles. Carol is the head of Schweiger Consulting based in Boston. Carol has been supporting transit agencies around the world for over 40 years and is nationally and internationally recognized in transportation technology consulting. Her wide-ranging and in-depth expertise is in several specialty areas, including technology strategies for public agencies, public transit technology, traveler information strategies and systems, and systems engineering. Now on to part one of the interview.

Kevin Chambers: Carol Schweiger, welcome to this interview for conversations with leaders. I'm really glad that you're here and able to talk with us about technology and transit. You have recently written three white papers for the National Center for Applied Transit Technology. You wrote one on open source software and open data, another on mobility as a service, and another on automated vehicles. Those are three big topics in transit right now around technology. And I would love to hear your theory on that. So welcome to the interview.

First, to start off, tell me a little bit about you.

Carol Schweiger: Okay, so one of one of the key things, first of all, thank you so much for having me, and be able to discuss a bit about those papers, as well as some of my thinking of where we are in the mobility space right at the moment.

But I guess to start off the story that, I'd like to tell, because it's true, why people don't exactly believe it. But when you ask a lot of our colleagues, so how did you end up in the public transportation field, in general, and many of my colleagues ended up there by accident. They, they did not go to school, to learn about transportation, their career paths, took them in a way where they ended up in, in that field. Yep. I, on the other hand, ended up in this field, because I always wanted to be in this field. And it started when I was very, very young, visiting my grandparents in Brooklyn, New York. And they lived not very far from New York City subway station. And very unique feature of that subway station is that a few blocks away

from it. There was a pedestrian bridge that went over the subway tracks that led to that station. And as a little girl, I would stand on that bridge, and I would watch all the subway cars go in and out of the station. And at that time, on that particular line, they would run express service and local service. And they had two different route designations. Anyway, I was about eight or nine years old, and I was just enamored with this. I grew up in the suburbs of Boston. And I certainly had seen the subway in Boston, but I spent much more time visiting my grandparents in Brooklyn. So I would just for hours on end, stand on that bridge and watch the subway car. It's a nice way if doesn't conductors. I just thought this was the coolest thing since sliced bread. Isn't there's something down the lines and I could do. So I wrote letters to the Transit Authority, asking them why they ran some service Express and some service on local. And I'm sure at that point, my parents wanted to disown me. But they didn't. They were very encouraging. And they said, if this is something you're interested in, and again, I'm only about eight or nine years old here.

So fast forward to college. I was a math major. And quite honestly, my goal was to be in some part of the transportation public transportation field. When I first went to college, I wasn't exactly sure what it was, but it ended up I started out as a transit planner. After I went to graduate school, I graduate school. That time I was in graduate school was all devoted to learning about transit planning. And then in one of my first transit planning jobs, we won a contract with at that time the Urban Mass Transportation Administration, which is now of course FTA. And that is a task order contract. You were to go out and provide technical assistance to transit agencies that were beginning to deal with sort of what we would consider today, technology, like, sort of things, but scheduling systems.

And my boss said to me, “Well, Carol, I want you to go out to the Rogue Valley Transit District, help them go buy some scheduling software.” So after I picked my jaw up from the floor, I thought, well, I guess I better bit better learn about how we look at these things. And I've never looked back. That was the beginning of me being involved in public transit technology, and the beginning of providing technical assistance to agencies. And coming full circle from being a little girl, and very interested in it was teaching some national transit Institute classes in how to procure technology at New York City transit time. And that was kind of my full circle.

So now I'm focusing on some things that are relatively new. As you mentioned, in the white papers that I wrote, some of those topics are still relatively new. And so it's exciting to kind of be at this point right now, where things are changing so rapidly. So that's kind of how I got into this.

Kevin Chambers: That's fantastic, that that interest was sparked early and that you've been able to do that early love, and carried on to today. How many agencies do you think you've worked with over the years?

Carol Schweiger: I actually did take a look at that fairly recently. It's about 77. Zero, now. And that includes some work on a technical assistance side, not only directly with all of these agencies, but also several of the federally funded agencies, right. And FTA, FHWA, as well as TRB. And some of those white papers written for the National Center for Applied Transit Technology evolved from work with the National Aging and Disability Transportation Center, that whole group. But that's the approximate number.

Kevin Chambers : Great. Well, you know, it's precisely because of that depth of background that I thought it would be really great to interview you for this conversation with leaders kickoff. So let's start off talking about open source software and open data. And we'll keep this conversation dialed in to what's really useful for that particular audience of the mobility manager, the general manager, the people who are figuring out how to design their services, what they need to be thinking about, you know, on a horizon, maybe 3 to 5 to 10 years. What are the things that they need to be aware of that maybe they think they can do now or things they just need to keep on their radar? So around open source software and open beta, maybe we can start off with a definition of what you mean and with a few examples?

Carol Schweiger: Okay. I think on the on the open source, and open data, I had conducted some research through the Transportation Research Board several years ago now, when agencies were just starting to think about it. And so the whole idea is that it is available to the public for free. Although I will explain why it is truly not free. And I remember conducting research, when I surveyed people. And I asked, "What are you most concerned about." At the beginning of open transit data, it was somebody who's going to do the wrong thing with this data. And they will misrepresent us as an agency. Frankly, that really never happened. And that's the good news that it didn't.

But it goes a lot deeper than just providing your operational data to the public. One of the things and again, relating specifically to the more rural and smaller urban marketplace, it does take effort for agencies to provide their operational real time data to the public, right, because they need to make sure that data is clean. And that there is nothing in that data, that would create some kind of misinterpretation. So it does mean that somebody's back in the office, before that data is released, has to really go through a validation process. And fortunately, in today's world, we have a number of validation tools that are available to us.

But back to the purpose of, of open data. One of them is for an agency to be transparent.

I've been very fortunate to work for an agency here in Massachusetts, that actually started transparency, more than 10 years ago, any checks that they wrote, the data was available for the public to know exactly what they were spending their money on, kind of an open checkbook, if you will. And they very freely embraced their operational and real time data. Because they wanted to be transparent, as well as obviously, helping their customers know where their vehicles were in, you know, when the vehicle was arriving at a stop. But there's also a number of other hidden benefits. But it's also where customers want to take a look at how things are operated. And all the information is available to them. But with a small agency, like the one that I work for, and I'm still working for in Massachusetts, they did not have the bandwidth to hire someone to develop the apps around that data.

Kevin Chambers: Now we think of an app, we don't even think about what does it take to develop it? Because now it's done in a very different way. But by producing this data, there were individuals who wanted to develop these apps, and basically did it for no cost to these are lightning people in the community. These are people and people in the service district who wanted to support the effort.

Carol Schweiger: Yes, they were supportive of transit. They typically were software developers, that was typically their roles, but they wanted to help and this was a challenge.

In Boston, when the Massachusetts Bay Transportation Authority decided to go with open data, there was a app developed within about three days. And that never, ever would have happened, had it been left to the transit authority to develop an app, it would have taken much longer, and the resources just born fair. So that's something very important.

And the other thing is, it elevates transit in the community, to a point where people who might not have had an interest in writing or knowing anything about public transit. It, it elevates transit to be much more visible in the community. And I also think that's something that we do not talk about, we assume everybody is walking around with a smart phone, and they have their collection of apps. But what does it mean, when you actually are using an app, that business or that transit agency is somewhat elevated in your perception? I think that was also an important part of open data.

Now on the open source software inside that is also, I don't think we've utilized that to its fullest extent, it always seems to be a bit hidden from view. And the reason is, many transit agencies develop RFPs. And I'm someone who helps them do that, as well as write specifications for software. But they'll tend to do that before they will look around the community and say, Well, is there some open software that would help us to do what we need to do, and perhaps that already exists. And I might actually be able to utilize that, particularly, again, with rural and smaller urban system. And when I was doing the research for that aspect of the white paper, I knew of several open source software programs that were already being used in the field. But again, just don't get a lot of press. And hopefully, with people reading that part of the white paper, they realize that there are more and more of these systems that are being developed, like the one that you're developing, for example, very, very important aspects of open software.

Kevin Chambers: Right. Yeah. You know, this seems like there's these opportunities out here for both these open data that you talked about earlier, and open source software that opened up doors to transit agencies. And, you know, for the data, you know, you bring up the transparency, the community engagement that the Massachusetts agency got when it said it wanted to move towards transparency through open data. And suddenly they got this community of developers that sort of became like a brigade of people wanting to help out in that effort. And then, you know, these outcomes in terms of elevating transit's visibility, and it becomes sort of a marketing tool in that sense.

But I think a lot of agencies say that, well, that's all great. I just don't know where to begin. And I don't know where to start. It's a big world. And, for example, open data, what are some of the things that need to happen within an agency, so that they can think about data as part of how they structure themselves, how they prioritize their time, how they, how they conceptualize how they do things? Because if you start from a world where all the data is just internal, and it's meant to be internal, it's not really, even if you shared it, it really wouldn't be easy to understand. And then you want to move towards a world where you could be transparent. It seems like there's a number of steps or skills that need to be built up. And I'm

wondering if you'd be able to talk about those a bit. Like what does it take to start to move in this direction?

Carol Schweiger: Yeah, I think you're highlighting something very, very important. Always, where do you start? And I think with a lot of the rural and smaller urban agencies that are still going through deploying technology, the Open Data question, Where do I begin? What? Why would I even want to consider this? Am I not making more work for myself? I think the path is you have data that's very useful, not only to the public, but do you in turn awake, and making that transformation to be a data driven agency really requires embracing open data for a number of different reasons. And I think where to start is what, what data do you look at internally? What data do you bring to your board of directors? What questions are asked of you as an agency that you need to go digging into your operational data? To answer those questions.

And what I found with many rural and small urban agencies that I work with, a lot of that data is still either on paper, or it's in a spreadsheet that gets duplicated numerous times. And that makes it a lot less accurate, the more times it gets copied. What kinds of questions are you being asked by your local government riders by non riders, again, if it's a public service, you want to be open as to how you make decisions, those decisions should be data driven.

So in transforming to that sort of data driven thinking, you do need the resources internally, for someone to be able to take the information that's generated by your internal systems, and make it useful in order to be able to make decisions. And I think that's going back. And frankly, looking at the kinds of questions people ask you, as an agency. And if you're not able to answer that question, based on data, it's going to be very hard to support your answer. If you are a data driven organization, which I really believe this is where everything is going, eventually, even though you may need some more resources, I think it backs up your being able to manage the services that you provide. It also allows you to explain situations that may arise.

Another example that I'll give you, is that another agency happens to be again, it's started to write a data block. And a lot of us transit nerds for in seven, seven. They were writing about operational real time data. What did it say? But what it really did, and this was more important than telling the public about the data, is they started to use it internally to make decisions. And they started to do analyses that they never did before. They didn't even recognize how valuable that data was, until they started explaining it to people who don't work in this field. And they started to recognize, okay, so why is it exactly that our on time performance with this particular bus route? is never above 85%? You know, we know that this has been going on. But we've never lifted up the cover on the data and see what is it telling us here? Are we scheduling correctly. So the bus driver can never be on time. Because, you know, we have other issues out it out there in the field as far as traffic or whatever, that's probably a little more of an urban situation.

But even in a rural situation, you can have issues with scheduling, says you're traveling great distances. And you may be transporting people to medical appointments like dialysis. And you don't want them to be on the vehicle more than a certain period of time. All of these things that you would be analyzing, have to come from your real time operations, and all the data that is being generated, as you go, you know, moment to moment with your operation.

Kevin Chambers: You know, one thing I heard you say early on, and in your statements just now is that you find that people have been taking, they take documents, and then make copies of those documents, and then make copies of those. And then with each successive copy, the quality goes down. And it sounds like you're talking about a reconceptualization of how we manage data. Because I think for a lot of us, the way we think about data is sort of linked to documents linked to Excel spreadsheets, most commonly right, and that you pull those spreadsheets, and then you do some work on them, you distill down the things you need. And then you make maybe make another spreadsheet based on that data set. And it sounds like you're proposing a different way of thinking about things that you are. And I'm gonna give you a shot at maybe talking about what that is.

But one thing I hear is that, that there's different workflows around data, where you have, perhaps you have a core source of data, and then you draw from that, you know, you've sort of instead of taking a single picture of it and having that be a copy, maybe you have that same core data, and then you just extract what you need, always from that same source data, is that something close to what you're thinking about that it's a different approach, or just talk about, like what you see at a disease as they start to improve their skills around data management?

Carol Schweiger: So you described it very aptly. And, and this started for me, many, many years ago, when I started to do a lot of this technical assistance work. When you first learned about a computer, what do you have, you have three things, you have input, you have some processing of that input, and then you output something. So when I would walk into an agency, and I still do this today, what I'm doing technical assistance. I want to know, the data flows throughout that agency. I want to sit down with a dispatcher. And I want to find out, where do you get the information that you use to make decisions? And once you make decisions, who do you give that information to? So again, I'm oversimplifying here, but it's input processing, which includes decision making output. What that revealed to me over and over again, and it still reveals to me is that transit agencies don't always think about data flow or workflows in that way. And the reason is very simple. They're concerned about getting vehicles on the street on a minute to minute basis. And that typically is the focus. Now, we know that data supports that.

But sometimes that operational focus, which needs to be there sometimes overshadows the data. If you understand the flow of data through your organization, you can stop the duplication. You using the word spreadsheet business. You know, one of my pet peeves is over a very long period of time. I will go into an agency and I will interviewed numerous people in different groups, and some people. They were a lot of different hats in a rural agency, but you know, tell me what you do. With this information, well, you know, I put it in a spreadsheet, and I do these things to it. And then I give it to my colleague in this other area, you know, to look at for planning, and then they take that spreadsheet, and they do something else with it. And pretty soon, you've got this very disjointed picture of what's going on. And something you said before, is, is critical to the way that I think about it, the source data should always be there and not be manipulated to change, and you not be able to go back to it.

This is one of the reasons why using a data warehouse having a central location where anybody can pull information and doing analysis. But the source information remains. It doesn't get modified, but you have

the ability to access it, and to do any kind of analysis you want with it. I think that is the way I think about it. This is all about data flows. And what it will highlight is where there is a problem in that data flow, that maybe the person in planning has received data that has been changed 10 times before they see. And by the time they get it, it's pretty useless. It bears No, none of the original characteristics that plan or might have needed to know.

Kevin Chambers: Right? And it seems like these are the underlying issues behind open data that are crucial for transit agencies to understand is, it's not just sort of copying another copy of a spreadsheet to a websites where everybody can see it, it's actually changing how you manage your data and thinking in terms of these flows, rather than thinking of sort of static documents, or snapshots. And that sounds like an that could be a big piece, right? I mean, that's not that that shift in how to think about these issues might be a significant level of effort, right? And so when people are trying to think about, well, what are the costs to becoming this, that you talked about the benefits, there's the transparency, there's the community engagement, there's the marketing value, there's the increases in efficiency, lots of wins. But then there's this organizational change that needs to happen. And when you talk to agencies about how to sort of plan for that, or try to think about the effort that is going to require to get to that place, do you have any, like rules of thumb for helping them get to that place?

Carol Schweiger: I do absolutely in part of it is internal to a transit agency, you often times do not have to account for the various tasks that you spend time on. as a consultant, I've always had to account for all of my time. I you know, even though I've worked for myself now, I still have to account for my time. But I used to be filling out timesheets every day or every week. However, it works in the consulting world agencies, when I did the initial research about how much it cost. Most agencies really were not aware of how much how many resources they were using to clean up their data before they release it to the public. And they certainly didn't know how much time they were spending on the analysis side. So what I often say to my clients are that in order to get a handle on the cost, most of it really is a labor cost. It's really not a cost, the cost of storing and managing data is so much lower than it's ever been. Right? I'm not saying it's zero. But I'm also not saying it's really big, a big cost, the primary cost is going to be labor, and that agencies should really be looking at how much time is their staff taking to actually manage the data, to perform analyses to actually go through a validation process. Like I say, now we have some validation tools that we never had in the early days of open data. And that certainly streamlines the amount of time that someone might need to spend. I've always been very involved in agencies providing real time information to the public. And in doing that, I'm always suggesting that that real time information is monitored. Because if you tell an individual that the bus is coming in five minutes, but it doesn't come in five minutes, it either comes in two minutes, or it comes in 15 minutes, that person is going to lose their faith. Right?

Kevin Chambers: And then you've lost trust.

Carol Schweiger: If you as an agency, never monitor what you're putting out there. So what does that mean? It's labor, it's somebody's time, doing a validation of the real time information that's going out to the public, whether it's at the bus stop, or it's in an app, or it's on their website, did the buzzer actually arrived in five minutes or not? And someone's got to take some amount of time to do that. So I always

suggest that agencies need to have a much better handle on how much time their staff is actually doing that work. And that translates into a direct cost, even though they may not be filling out a timesheet. So and when I did the original research there, the information about cost was practically non existent. So now it's, you know, how much time does your staff take obtaining the data? So if they're using some of these legacy, CAD AVL technologies, they've got to go into some reporting system and generate the information may take them a fair amount of time to do that. And that's a cost right there. of labor? And then how long is it taking you to actually use that data for something? While so. So it's being more mindful? I think of what you're doing internally, to generate the data. How much time are you spending in the sense of open data? How much time is your staff spending, cleaning the data, to make sure that it, it shouldn't be going out to the public?

Kevin Chambers: Well, that's very helpful.

So one thing I hear about with both the open source and the open data and the mobility as a service and the automated vehicles, is that you really want to think about what you want to get out of the technology, and what are the non technical things you need to have in place. It sounds like you know, and in the case of the most recent piece on AVs, you really want to take a look at where your services are more difficult to provide where it doesn't pencil out, you need to have that clear. And with mobility as a service, you really need your partners in place, you really need to be thinking about what the larger market and outcomes you want, as well as what's your method for evaluating success. And then going back to open data, we hadn't talked so much about open source, maybe we'll talk about that another day. But going back to open data, that really how data flows through an organization really matters. And those incremental processes of how you track and manage your data matters. So all those things, to me say that technology itself is only one small part of the picture. You want to say more about that and how you've found that to be true in your work.

Carol Schweiger: Yeah, that that's really a very, very key thing that you just said, because, frankly, the technology works on Yes, there are new technologies being developed all the time. But it's more about what does the user need. And, and by the way, there are multiple users here. It's not just the traveling public, it's not just the internal agency staff. It's not just the funding agencies, there are a number of different users. And this always brings me back to a very basic systems engineering concept, where you don't build a system with blinders on in the back rooms somewhere, you start by determining what the user needs, what are those outcomes you're looking for, which helps you determine what the needs are. And then you build whatever you are going to build. So that when you get to the end of that, you can go back to the beginning and say, I'd satisfied that need, unable to make sure that my system covers those needs. And frankly, very early on in technology and transit, we didn't do any of that planning, people just jumped forward them. I'm going to buy this scheduling software, because that's what I think I was supposed to do. Well, what did you want the software to do? Nobody asked that question. They just went to the marketplace and said, I needed the scheduling product. Tell me what you got. And they bought one. But they never really knew other than third basic things like on the fixed route side, like actually scheduling service. But what were your goals and objectives? I don't know how to evaluate something

that doesn't have goals and objectives. And when you want to set the goals and objectives, what do people need to have to fulfill those. So I do totally agree with your sort of vision of kind of putting this all together. The technology is important. And of course, we also have to respect the fact that not everybody has some of the technology that may be on the user side, more on the travelers side, we have to respect that. But the technology is not typically the challenge. The challenge is more all of these other things. Have we collected the user needs and are we building something to satisfy those needs?

Kevin Chambers: That's a great note to end on. Those are wise words. Thank you so much, Carol Schweiger for talking with me today. Now, to close if people want to learn more about your work, where can they find you?

Carol Schweiger: So I have a very simple email address done very much on purpose. So nobody has to spell my last name. It's just Carol@tech4transit.com. Very easy to get, get in touch with me. There are some other resources, where you find those white papers that could be useful to folks where other people are talking about these same topics. We conducted some webinars, there are some fact sheets out there. And there are some other places to look for my work, but probably emailing me is the most direct way to find out about it.

Kevin Chambers: Right and your white papers are at N-CATT.org, as part of their Tech University portion of their website. You can do a search for white papers and find the white papers that Carol has written as well as several others are some on electrical vehicles, and a range of other topics. So N-CATT is a great resource for getting a sense for what's happening around technology in the transit world. Carol Schweiger, thank you very much for joining me today for this time conversations with leaders. And we'll close there.

Carol Schweiger: Thank you very much for having me.

Kevin Chambers: In part two, we will talk with Carol about Mobility as a Service. Thanks for listening!

This podcast was funded by NCMM. Check out all our resources, aimed at mobility managers at nc4mm.org.