Sheryl: Good afternoon. Good morning, wherever you are. Welcome to the webinar of the National Center for Applied Transit Technology or N-CATT. This is our second webinar and this one will be on open source software and open data. If you missed our first webinar, Framework for Technology Decision Making, you can go to our website and you can have access to that one. So, just one little direction for the webinar. If you have any questions, please go to the Q&A. That's where we'll be paying attention to those. Let me tell you a little bit about N-CATT before we go to our speakers.

There we go. All right. N-CATT was launched several months ago in October 2019. We are the newest National Technical Assistance Center that is funded by the Federal Transit Administration, and this is done through a cooperative agreement with the Community Transportation Association of America. Here, you can see our mission. We are focused on rural, smaller urban, and tribal transit systems and helping them to learn about and adapt and implement new technologies. What do we mean by that? We mean technologies that are developing, emerging and starting to be used in other areas of the transportation universe.

You can find us at our new website n-catt, C-A-T-T, .org, where you can find lots of great information. I'm going to introduce our first speaker of our three. We have three speakers. Carol Schweiger is our first speaker. She's going to be talking about the sort of an overview of the entire topic of open source software and open data. Then we have Sarah Anderson, who will be speaking about open source software and then Sean Barbeau, who will be talking about open data. So, now, to talk a little bit about Carol before she begins, Carol has over 40 years of experience and she is nationally and internationally recognized in the field of transportation technology, consulting.

If you know anything about her, she does have, as she says wide-ranging and in-depth experience when it comes to a whole host of technologies and to how rural transit systems smaller been and tribal transit operates. She's provided over 65 transportation agencies with technical assistance including developing technology strategies based on needs assessments, developing technical specification, and structured processes to procure and implement technology systems. She's developed and delivered webinars, trainings for the National Transit Institute, the Transit Cooperative Research Program and she's authored and co-authored CCRP reports. And without further ado, I have to say we can't get a better speaker with more broad knowledge than Carol. So, Carol, I will let you take over and I'm going to cede powerful remote control to you. There you go.

Carol: Oh, boy. That's scary. Okay, so let me advance the slide. It may take just a minute.

Sheryl: It may take just a minute.

Carol: But thank you for the introduction, Sheryl, and thanks for having me. I think this is going to be really interesting to folks who have often heard the terms 'open source software' and 'open data'. They can be rather confusing. So, hopefully, we'll dispel all of the myths associated with that. So, I'm just going to stop my videos so I had a little more bandwidth here as I click through the slides. So, there's our first one. Here's the agenda for what I'm going to be speaking about. What's great about this webinar is I'm going to give you some definitions of open source software, kind of put it in the context of demand response transit or paratransit, and then move into open data and sort of put that in context. But then I have two fantastic speakers following me who are going to really pick up on.

Sarah's going to talk about the open source software and some really great examples for you. And Sean's going to do the same with open data. So, let's move to the slide. I got a little bit of a lag here. There we go. So, you've probably heard the term 'open source software' quite a bit and very simply, you can see in that first set of bullets. It's a software with source code that anybody can look at. They can modify, they can make it better, and so that's what the open part of it is. Now, a lot of the software portion of it, most of you that are perhaps looking at using open source software, you may never get into the actual programming of it but you may be communicating with someone that is actually doing the programming sort of behind the scenes. And so, programmers that have access to the source code can be improving that.

If they improve the software, they're improving it for everybody who's using it. So, that's another aspect that I think is really important to take a look at. So, when you think about open source software, there is sort of these six principles that go along with open source software that I think is important to recognize. And one is really overarching. It means that there's an open exchange among the people that are actually looking at the software, making changes to it, and improving it because they want to make sure that it's beneficial for everyone that's using the software. So, that's important. The collaborative participation is extremely important, and this is something that sets open source software apart from open data. We'll talk about that in a minute.

Open source software, it also allows there to be sort of a rapid prototyping of anything that you want to change or you want to demonstrate in a software and it also makes the software transparent so that you know the changes that are being made, who's making them. There's usually documentation associated with that so this way, it's transparent to everyone who's using it or looking at it. And then it's based on a meritocracy. What does that mean? It means that the control of what's being changed is based on someone's knowledge of what's going on in the software rather than politically what's going on. So, it's based on knowledge of the software. Whenever you see these packages, open source software packages, they're really being developed in a bigger community.

They're really has a vested interest in making sure that software is doing what they think it should do, how it should be doing it. So, moving on from there, I want to talk just for a minute about the benefits. There you go. So, there are several benefits. People often prefer open source software to proprietary software because of these specific benefits. One is that a lot of people think that they have, and I think it's true that they actually have more control over that kind of software than they do over proprietary software.

So, I think that's really important. People who weren't programmers are benefiting from that software because they can actually use the software in any way they want to use it on not necessarily the way that someone thinks they should use it which is where proprietary software, that's sort of a disadvantage.

There's a trading aspect. If there are programmers out there, it really helps them to become better programmers when they work on open source because they're working with a number of other people. There's the issue of it being potentially more secure. They consider it more secure because someone might actually spot a problem with the software and be able to correct errors associated with it or any omissions on that someone might have missed when they originally developed the software.

Also, there's a thought that the software is a little more unstable and that's based again, on having a community working with a community of people that are working on the software. And nobody wants that software sort of disappear or fall into disrepair, and because you have this great community of people working on it, that's not going to happen.

And then we've talked already about how strong the community is in helping to make this software perhaps the best software that it can be for the reasons that it's there. So, here's something that I wanted to bring up at the beginning, and this is that a lot of people equate the term open source to it being free. That's kind of a misconception because open source is not necessarily free. On people who are programmers, open source software programmers, it can charge for the open source software that they create.

But in some cases, because the open source license might require them to release the source code to everybody when they sell that software, they often will opt instead to charge for services related to the software and support of the software. You'll see that quite a bit in our industry. That's typically what happens.

You really want that skill in programming and troubleshooting, because that's really extremely valuable. There are situations where employers specifically hire programmers with that experience of working on open source software because it's so valuable and if it working again in a community of people that are doing the same thing. So, let me give you just a few examples. Actually, Sarah is going to cover a lot of this so I don't want to really go into a lot of detail. But I'm guessing that a number of you that are participating in the webinar know about either OpenStreetMap or OpenTripPlanner. They're both open source software.

They're built by a community of people. On OpenStreetMap, it's a community of mappers that contribute to data about features that you want to see on a map. In OpenTripPlanner, that's actually a family of open source software that provides information for passengers. So, it provides itinerary information that might combine transit information with other modes like pedestrian mode, bicycles, other mobility services like bike share or scooter share. And then a couple of others that Sarah's going to go into detail about is One-Call/One-Click software that I'm guessing a number of you out there use. One specific piece of software that I always think is a good example to use when talking about open source software that's called RidePilot and Sarah's got covered that as well. So, let's move into open data and open data, there's-- No. Some confusion about that as well as confusing it with some open source software and we'll talk in a minute about the comparison of those two.

But the bottom line is open data is data that can be freely used or reused or redistributed by anyone. Typically, it is subject only to the requirement that it gets attributed to the originator and it's shared in the same way by everyone. So, here are some details to jump into that definition a little bit more. So, the data must be available as a whole and it no more than a reasonable reproduction class. So, that's preferably by downloading it over the Internet. That's the process that's used most frequently. We already talked about reusing and redistributing it and can be used and reused and redistributed and mixed with perhaps other data which might be very, very useful if they're using it for some sort of analytic capability.

And then everyone must be able to use, reuse, redistribute it. There should be no discrimination against anybody using it for any specific reason. So, let's talk about what is considered, "good open data". There's open data and then there's good open data. You'll see why this is important in a minute. First of all, open data by its nature should be able to be shared very easily. So, that's a given based on our definition. It also should be available in some kind of standardized or structured format. That's going to come up again in what Sean's going to talk about and I'll mention it again as well little down the road.

It should have guaranteed availability and should be consistent. And then the other thing that's very important is it should be traceable to where it came from so that you understand how it was built. So, let's talk a little bit about conditions.

Sheryl: Carol, this Sheryl. I'm just going to interrupt to remind people to put their questions in through the Q&A if in case you have any. Thanks and back to you.

Carol: Sure. So, some conditions of open data, one, we need to make sure that we take privacy into consideration, and this is actually going to come up in the other presentations as well. But that's something that's very important. You want to make sure that the data is closest to the source as possible and you'll see why that's important when we go into some examples. You want to ensure that it's very easy to exchange this data with others who want to use it. It should be accessible and that means in some kind of digital format, so as many people as possible can utilize. So, that's where kind of the standardized formats come into play. It should be easily readable by any computer. It should be mostly accessible for free and with no restrictions of views.

Again, we want to emphasize the format and it being in a standard format. Now, I want to go into a little bit more about what is open transit data, and why really is there such a thing? And one overarching thing that you see here—I'm going to mention on the next slide—is a lot of transit agencies specifically the ones that are represented, I think, in this webinar particularly, really don't have the bandwidth to do a lot of analysis on their data or visualization of their data that describes their operation. If you open up your data, you will allow a bigger community to actually conduct that analysis and to do some of those visualizations that you want to be looking at. Perhaps you're looking at performance information. Perhaps you're looking at data that's going to be very helpful to someone who's going to utilize your service. There are a whole host of reasons for having open transit data. But a couple of things are on the slide.

And I'm not going to repeat from what I've said before. One thing that open data is not is it's not static. It's always being changed, it's being updated, and it may not be comprehensive. It depends on the agency that is opened up their data and what they have chosen to share with the public. So, it may not be a comprehensive database, and it's available for anyone that has an interest in it. And data quality is something that can be evolving. One thing that I will mention in a minute—I'm running a little bit over so I'll finish up quickly here—is that the data quality is very, very important for those agencies providing that. It means that you actually have to spend some time, perhaps cleaning up the data before you open it up.

So, I wanted to show you a chart that I've always liked from Tim Moore, who's at the Bay Area Rapid Transit System. When open data was becoming very, very commonly discussed back about probably seven or eight years ago was the beginning of it, to put together this chart that I think very aptly describes the fact that here, BART is the transit agency and they're generating a whole host of data publicly. What they're doing is they're opening it up to people that want to use it for a variety of reasons, and they're typically developers that want to take that open data, analyze it, perhaps put it in another format for consumption, whatever they want to do with it.

But eventually, it's going to customers, to BART customers. So, there's really this ecosystem and flow of open data that I think is very important to be thinking about when you're thinking about providing open data. So, quickly, basic characteristics are-- So, there are two basic kinds of open data.

One is static data which describes transit schedules and you have separate fixture service, and they may not change very frequently during the year. And then you might have real-time information which gives estimates as to when a vehicle is going to arrive at a stop or maybe you'll show where a vehicle was located. That can change very frequently. And then when we talk about the magnitude of data, there are two schools of thought. One is, you dump everything you've got and you just let people consume it however they want. That's called a fire hose for obvious reasons. The other is to have faucet where you were looking at a specific subset of data that you want to be analyzed. So, you can sort of imagine these benefits of providing your data in the marketplace. It actually provides some increased awareness of your services that you provide.

It sort of empowers customers. They have more information to work with, makes them often feel more comfortable about using your services. Open data encourages some innovation that you might not be able to accomplish inside your agency. It provides some great transparency to people who utilize your system and fund your operation. It gives some opportunities for private business as well. So, here are some additional benefits and I'll just mention one of them. When you make your data publicly available to developers and other users, they're going to be able to tell if there are issues associated with your data if there are data quality issues, or there are some anomalies they're seeing, and they can immediately tell you that when you might not be aware that that is the case. So, that's the one thing I wanted to mention there. And then some examples of some smaller agencies that are actually they've opened up their data. I just wanted you to get a sense that this is not just for major urban areas.

It's also for much smaller agencies. And then here is really my sort of wrap up. So, a lot of people get confused between what is open source software and what is open data. I think by the descriptions, you can probably you could have written this slide yourself. So, when an open source software, it's being improved, updated by a community of people or programmers and it's being used in an environment that supports transparency and talking about issues that need to be solved. It really allows the documentation of what it does. There are places where people have provided tools that utilize that as well. On the open data side, it's published by a single organization. That's one of the biggest differences.

So, a transit agency provides open data, it's by that specific agency. So, there isn't always a lot of insight into how the data was actually put together, but typically, there is some documentation about it. You can access it through some kind of portal or some kind of infrastructure. There really is not a lot of support for collaboration except for open systems like OpenStreetMap or OpenTripPlanner. Here are some resources. I think, Sheryl, the slides are going to be made available in the recording so I'm not going to spend any time. There are actually many, many more reasons resources that folks will have to sort of go into this a little bit further. One of them in particular is a report that I wrote several years ago for TCRP that talks about open transit data. So, that, I'll leave it back to you, Sheryl.

Sheryl: Okay. We have a question from Marcy Jaffe, she has worked with RTAP and she's asking, are there no-cost tools to generate open data transit GTFS? I think we're going to get to that a little later in the program, Marcy. So, we're just going to table that question for now, but we're not going to forget about it.

Carol: Yeah, that's a great question.

Sheryl: I am going to forward us to, there we go, to the next slide. I'm going to introduce you, Sarah, in one moment. First, I'm going to give control of the keyboard to you. Now, I'll introduce you so that we don't waste that time lag. Okay, now I've known Sarah for a long time now, I think 8 or 10 years. So, it's really a pleasure to introduce you as well, Sarah. Sarah Anderson is a principal and Vice President at Cambridge Systematics. She has over 14 years of experience in the transportation technology field. She oversees the software business line for Cambridge Systematics, and prior to joining that organization, she spent eight years in software development as a project manager.

You'll see exactly why she's right for this audience because she's implemented custom technology solutions for transit agencies. Her areas of expertise include human services transportation, transportation technology, trip planning, and mobility management. She's an advocate for expanded use of open data, open data standards, and open source technologies in the public transportation industry. She's passionate about the role of technology in driving better results for transportation providers and consumers. Okay, I'll let you take it away, Sarah.

Sarah: Thanks. I, as the middle presenter, I wanted to craft a presentation that really builds on what Carol presented earlier, and I'm trying to change the-

Sheryl: Oh, is it working yet?

Sarah: Not yet. Oh, there we go. So, my presentation is in two parts. First, I want to focus on the business case for open source. Why would someone choose to go in this direction as opposed to something traditional and proprietary? And then I'm going to give some examples of tools that I think I've seen successful projects, particularly in rural and small urban communities. So, it's covering two very different topics but they're related. And they build on Carol's presentation. So, I'm still having a bit of a lag with my-- Here we go. Okay. To recap Carol's definition, one of our attendees who is a participant, Liz Peake, once gave a presentation where she described open sources being similar to architects looking at blueprints.

And so, I'm thrilled that you're a participant today, Liz. I'm giving you a shout out and I used this concept in this image that I included on this side, and I think that's a really interesting way to visualize what open source is, and what Carol meant when she was saying lots of people can look at this thing. It can it's something that people can build on, modify, and improve upon. That's very much, it's sort of a similar to architects looking at open source plans or looking at blueprints before they go to build a house or something like that. A collection of folks can kind of look at the underlying source content before they make a decision to actually build and implement something. So, these bullets were ones that Carol already shared with us.

So, I want to quickly pivot to what are the advantages of this approach. In one of the advantages that's commonly stated for a business case as to why one might do this is that there are no license fees. So, there's no recurring costs for the privilege to use the software, which has pros and cons which I'll get to in a minute. That can lead to a reduction on an ongoing cost as sort of one advantage. Another is that the community shares the cost of improvements. So, everyone knows that software has to be maintained and updated over time. So, there is a cost, an ongoing cost to open source, but that cost is directly attributed to improving it over time.

It's a cost that can be shared across boundaries, across states and things like that. Across countries too in some cases. Some of these projects are not just in the US, they're multinational. You can also get the benefits of shared perspectives without any one particular agenda being forced which is another benefit, and then it also eliminates the possibility of someone being locked into one particular provider of technology for a period of time. So, the nature of it being open source means it's not owned by any one company. It's owned by the community. So, if you decide that you hired a college student, or you hired a company, a consulting firm, or you yourself implemented this, at any point, you could decide to change who is the entity or person who is maintaining it for you, which gives you a lot of flexibility in how you maintain and how much you pay for this software over time, which is another really important advantage.

So, why would someone make a decision to do this thing? What are some of the driving business motivations behind open source? So, these are four cases where I've seen the business reason creating the most successful outcome for an open source deployment. And this is based on experience, things I've seen in the field. Number one, I think open source makes the most sense when thinking about a regional group, statewide deployment because really, the cost then at that point is distributed among the broadest group of potential agencies. And the ongoing support then is also significantly reduced towards a proprietary alternative.

So, an example of this might be a state DOT transit division having an interest in supporting something that can then be consumed and used by a number of smaller rural transit agencies. I think that's a great use case for when open source can deliver the most value. I also think it brings a lot of value when there is an intent to customize. So, even if it's a single agency that wants to deploy, if this agency or regional group has a vision for something that really can't be matched with what's available off the shelf. So, there's a vision and an intent to grow and customize this thing to meet your regional nuances, which I have seen in rural communities who are often forced to adapt, modify, be creative, depending on constraints or other unique regional characteristics.

There's something about having sort of control over what the software can do for you in order for you to meet these unique regional constraints or funding constraints that open source really can provide some value there. Also, if you want to be a part of this broader community, if that drives your decision making, and then so it's not really a cost-based decision. It's a decision to be a part of a whole and contributor to this community. That could be another reason why you might want to engage in open source. And then finally, kind of related to something I said before, with that no vendor lock-in.

If you live in a community where you have access to local technical resources, either through a local university, or someone you have on staff, or a relationship with a local consultant who can help with support of this thing over time, that might be another reason why you would want to use open source versus something else.

So, now for the fun stuff, examples of open source tools. So, I said I wanted to expand on some of the examples Carol gave but this next section is structured by tools that our customers are facing. So, ones that would be things that writers would engage with. The second section focuses on agency oriented tools. So, tools that would be used by folks inside of the transit agency. Here is my general public. So, OpenShift planner, people are familiar with it. It's a very well recognized open source trip planning tool. But the example I want to give is the one that touches on a statewide deployment that was sort of started, initiated by the DOT, but ultimately helps all of the providers in Vermont, small, medium, and large.

I think this is a really fabulous example of the value that open source can bring to a regional or statewide deployment. So, Vermont VTrans started this project where they deployed open trip planner, and then layered into it, an open data standard called GTFS-flex, which Shawn's going to talk about. And ultimately created this statewide platform to discover different types of transit modes beyond what Google Maps was able to provide and share. So, I could incorporate all of the unique, specific different types of services that the rural providers have to serve in order to meet the needs of their community but a little bit different than what a traditional fixed route service might be.

So, what I mean by that, I have some slides that come with that, if it will go. Great. So, this is a beautiful example of the benefits to rural communities when you add open source software with an open source data standard. So, on the left is an image of what a flag stop would look like using this trip discovery tool. So, if someone wants to go from point A to point B in Vermont and on the right hand side, you're zooming in to see how that's communicated to the writer that you want it to start at the green flag, but if you just walk a couple of hundred feet, you can flag down that transit vehicle at any point along that sort of dotted blue line area. And as an example, the same trip that's discoverable because of the use of OpenTripPlanner and GTFS-flex would not be discoverable using Google Maps as an example.

So, the same trip that I tried to complete in Google Maps, it said there was no transportation option available to me, I have to walk, but that's not true. Obviously, we do have an option. It's just the flag stop, it's a little bit different. So, I hope this is a good case study that kind of shows the business case and value behind why bringing these things together can be so important and powerful, particularly enrolled communities. So, similar thing, the same project also looked at how do you discover a Devi in fixed route stops, and it even brought in Call-n-Ride as well? Another example, so I'm not going to go into all of this text. I figure this is recorded and it's available online, so I just want to kind of quickly go through some examples.

A combination of two open source projects, the OneBusAway and TheTransitClock, come together to bring a real-time next bus type information system for agencies. Certainly, some value there in both urban and rural contexts. Another example is the 1-Click mobility management tool. So, this is much more focused on trip discovery for persons with disabilities, persons with special needs. The next one is a variation of One Click, Link2Support, which really brings the resource dimension into the concept of mobility management and trip discovery.

So, the nuance here is really making sure that folks are able to find services that they need, particularly as we think about COVID recovery and making sure people have access to food and access to other basic human services. And then after sort of discovering that, how do we connect them to those services based on whatever transportation infrastructure is available in their region?

Sheryl: So, almost like a 211 type of model for that one?

Sarah: Yeah, it would have to fit things in the 211 database and layers transportation discovery along with that.

Sheryl: Okay, great. Thanks.

Sarah: Yeah, definitely bring that in our databases is critical. So, I think that covers customer-facing open source tech lists out there in use and have some really good real case studies. So, I'm going to pivot quickly to agency center tools. Hold on. You're going to see all of them. I can't even talk about them. Sean, I'm sorry that I'm eating into your time here.

Sean: No worries.

Sarah: Okay, we'll start with RidePilot. There was one before RidePilot but we're not going to get to it. Unfortunately, I can't seem to get the slides to go there. Okay, RidePilot is a really fascinating platform. It's a CAD/AVL system, and just to be clear, there's no automated scheduling here. So, it was really designed for small 10 vehicle providers, very rural, very small. And bringing it up as another example of how a regional deployment—in this case, sort of not statewide—is allowing these very, very small providers to give service and have software technology available for them to provide services in areas where the need is staggeringly high. But they are too small to necessarily have gone out and purchased something on their own because there's a regional sponsor for this software.

They have access to it and it's sort of controlled by this regional player, this regional entity. So, lots of really good information and value for that as well. Okay, national our tap of resource you guys probably know very well offers open source or no cost tools for folks. These are ones I haven't had personal experience with but I've heard good things about others. So, there are links embedded into this PowerPoint presentation. So, if you can get a hold of it, you could go and access and find more information about the website builder, the GTFS builder, which I think was one of the questions asked. I have not used this myself. Sean may have had some experience with it. I know he's done some research on these things. So, this is another resource to get low cost or sort of free tools to help agencies with a number of needs that they have.

TransAM is a platform I have had some experience with. It's an asset performance management system and yet again, another good example of something that deployed, when deployed at a statewide or regional level, can give a lot of assistance to agencies with NTD reporting. So, that's a tremendous benefit to a system like this that helps you get some handle over like how much you're paying for assets, when are you going to replace versus rehabilitate them, and really eliminate a lot of time needed for NTD reporting.

Finally, I think I just have one other thing to touch on, which is Carol mentioned OpenStreetMap. This is interesting because it's none of the other things that I talked about. It's not necessarily inherently client-facing, it's not really an agency tool. It's a tool that other software developers embed in the back end of systems that they use.

So, it's a really interesting example of something that developers use to then go on and create products and services that are consumed. But I just wanted to make sure you guys had information on OpenStreetMap because Carol, it is a fabulous example of a really healthy good open source community. So, worth checking out and kind of seeing how that works. That was the last example that I wanted to provide. So, I will hand it off to Sean, or Carol will hand it off to Sean.

Sean: Thank you very much, Sarah and Sheryl. Sheryl, you're muted.

Sheryl: Technology is only as good as the user. Okay, yes, we are recording this. The slides will all be available. So, all of that in-depth information that's in the slides that you're going to want to look at again, those will be up on our website within the week. Sarah just talked about those RTAP tools, and they were just covered in the N-CATT newsletter, so you can read a little bit more about those and there are links to all of those tools as well. And next month we are doing a webinar on GTFS-flex, and that will go even deeper into that topic.

So, that'll be there as well. We have some question but I think we're going to go right to Sean, and then ask them. I know I have a bunch but there's others up. But let's go to Sean now. And Sean, let me just introduce you briefly. Dr. Sean Barbeau is a research faculty member at CUTR, the Center for Urban Transportation Research at the University of South Florida. And he's been there since 2004. He's led many research projects, designing, implementing, and evaluating intelligent real-time traveler information which is something we get questions about for mobile phones.

He has over 50 peer-reviewed papers and conference presentations and patented inventions. And his research interests include multimodal mobile apps, open source software, standardized open data, and cybersecurity for mobile phones and transportation infrastructure—another important topic. So, Sean, I will mute myself and I'll allow you to begin.

Sean: Great. Thank you very much, Sheryl. So, today I'm going to talk about open data and three high level topics I'm going to cover are why open data, why should you be looking at sharing your data with other organizations, the anatomy of transit data sharing, just to give you an idea. If you're new to this or just kind of want an overview of what generally we're talking about when your people throw around different terms, and what those are, and then talk specifically about GTFS and some related data formats that you may find useful. First, why open data? So, before we actually even get to why, there's the what. So, I'm not going to spend a whole lot of time on this because Carol did a great job giving you the introduction.

But this is really transit data you're sharing with the public and you typically do that by putting it up on your website or an FTP site or some type of web service as part of your IT infrastructure. And as Carol emphasized, I think it is really important, this is not just something you do once and forget about. This is something that should be updated regularly. You want to make sure that as your transit system changes, everyone else that has access to this data is able to reflect those changes accurately.

We, I think, did a pretty good job of differentiating between open data and open source software here. So, I think this is how I try to explain the differences and also differentiate with a third term open architecture, which sometimes you'll hear tossed around. First, open architecture is most mostly focus on standards within your own agency system.

So, this is if you buy a transit vehicle or you buy an AVL server, check the vehicles and you buy a schedule system and you want them to all embrace open standards so they can talk to each other. That's an open architecture, but you're not necessarily going to share all that literally low-level data with the general public. So, that's where you get to open data is that some data that may be exported from these systems and it's the data that's shared with the general public. So, that's the red arrow here and that's what I'm going to be talking about. And then open source software, as we've already covered of course is, if any of these components in the system that are implemented in software, if the source code for that software is open for others to download and use. So, getting to the why is open data important.

So, a very smart woman by the name of Carol Schweiger wrote a TCRP report in 2015, which did a really good job, talked to a lot of transit agencies out there at the time and came to some really important conclusions. And this is a report that I frequently cite. So, the conclusions were that the benefits to the agency strongly support open transit data. The availability of the open transit data encourages innovation that cannot be accomplished only by the staff. Some of the benefits saw we're increasing awareness of the transit services, empowering customers by enabling them to use new tools, and encouraging innovation which creates some of those tools and improve the perceptions of our agency and also provide an opportunities for private businesses to get involved and again, fill some of those gaps.

The transit agency just may not have time refunding the tackle on a day to day basis. Another important conclusion was that the legal fears that sometimes people freak out like, "Oh, we can't do that because we're going to get sued." Those legal fears really have not been realized based on any agency's experiences. And when we've gotten to quite a number of agencies sharing open data, which I'll talk about more here shortly. And also, I think something really important from this report is looking at the factors that lead to successful open data programs and these are something that can take a number of different shapes.

But at a high level, the first is just maintaining management level of support and buy-in for the program. This is something getting into the second point here. It does require resources to make sure that you're sharing data in the correct formats and maintain the quality of that data.

So, if you're going to spend time doing that, you really need management level buy-in to make sure that the management's okay with that. You should also establish ways to monitor data accuracy, timeliness, reliability, quality usage, and maintenance. A good way to do that is by having, getting to the fifth point and open dialogue with the developers who we're actually using that data. So, that's a huge advantage to have a number of these developers that are building products based on this open information, really taking a fine-tooth comb and looking at the data and understanding what is working and what's understandable and what's not, what may not be right about that. That can provide some valuable feedback to you as an agency as you start actually using the same data, which I'll talk about here in a second too.

And then one of the final things is creating and maintaining a licensing or registration program. These should not be overly restrictive but these should lay some of the groundwork such as you're not allowed to use our agency's logo. For example, if you have your own app, you shouldn't be representing it as the agency's app. And just a registration so you know who's using the data and get an idea, which also helps with the communication developers that are using that application or building that application. So, I'll go through the anatomy of transit data sharing quickly. Carol already covered some of these, so static data as your schedules and where stops are located, things that generally change three or four times a year. Real-time data can change very quickly, user arrival predictions, vehicle positions, and service alerts which are human-readable information about your transit system. And similarly, she already covered this magnitude of firehose which is just a dump of like this is everything that's happening in your system.

It can be static and real time. And then the faucet is just a slice of that's carved out very specifically to say the bus 124 will be at your stop that you're standing out in about 15 minutes. And this is what a transit data flow looks like. So, you as a transit agency sit on the very far left-hand side of this diagram as the producer of the data. And then you have consumers which can consume these firehose formats are like the dump of the entire state of the transit system. And then some consumers then implement their own apps. So, Google Maps is one example. Google is the consumer of this data, they take in the entire state of the transit system, and process it and serve it up then in their own Google Maps app, which is on the right-hand side, which is the app that the transit agency or sorry, the transit writer actually uses.

So, popular formats for this firehose, the dump of the entire transit system are GTFS and GTFS Realtime. These are the data formats as an agency should really focus on producing first. If you really want to get into kind of being this middle tier yourself, which some agencies have, there are standards which define the second sharing of information between the consumer and the actual transit app. Siri is one of those. But again, if you really want to focus on just getting started GTFS and GTFS Realtime are the best place to start.

Now, I'm going to actually talk about GTFS and some related formats here shortly. So, a successful open data formats are organic. They're created by the people that are actually producing and consuming the data. They're open for evolution with an open forum where people who are using the data can come in and suggest improvements. And as part of that forum provide open documentation that people can review and better understand the format.

And finally, there are different things that you can do to make sure data formats are easier for app developers to use, and if their support open source tools or that type of thing, it makes data formats very successful. And a good example of that is the General Transit Feed Specification or GTFS, which was initially created by Transit Google in 2005. And it's since become a de facto standard worldwide for sharing transit data. And so, just one count as of a few months ago was over 1500 agencies sharing open data in GTFS format. As Carol said, these are both large and small and medium-size. It really spans the gamut here.

So, if you're a transit agency, there are ways for you to share your data and peers out there that have already done that you can talk to and get some suggestions. And then more relatively recently in the last five or so years, GTFS Realtime, which is a real-time counterpart to GTFS, has become increasingly popular. This is for arrival predictions which are called trip updates, vehicle position, the position of the vehicle, and then also service alerts. Service alerts, that is the human readable information which may describe what's going on. These are increasingly important, I think, especially today when service is so dynamic and you may have drivers calling in sick and have to cancel routes on the fly and you can't get out of new GTFS scheduled data that quickly to represent changes. So, I would highly suggest looking into service alerts if there are things you want to share with your riders. If you're on Google Maps, as of today, Google has an internal tool that you can go in and enter service alerts manually. Keep in mind that those will only be shown then in Google Maps.

If you want to share alerts in other systems or sorry, with other apps such as Transit app or any other apps out there Apple Maps, there are tools out there that you can create service alerts in GTFS Realtime format without requiring a full-blown AVL system. These are really simple web-based tools where you can go in and enter an alert and get that data out to riders. So, I'd highly suggest checking into that if you have an information you want to share with riders.

As Carol said, quality is really important. In one study, 9% of riders, they took the bus less often due to errors in real-time information. So, there's been general research showing that sharing data with riders, sharing real-time information, in particular, can lead to increases in ridership but you need to make sure that you're doing that well.

And data errors and inconsistencies can also make analysis hard. So, internally, there's increasingly a number of tools that you will use as an agency that use these data formats as input and so, you want to make sure that as you're making decisions internally, they're based on high-quality patterns. There are some existing validation tools out there and I included a link here that you can go and use to test and both you GTFS and GTFS Realtime data and research is high quality. There are also best practices for GTFS and there's a link here which I will suggest checking out and include some recommendations and examples which is categorized also by the types of consumer applications like trip planning, arrival time, destination and time travel table generation, and that's a great resource.

An important best practices that you'll find here but I wanted to call out specifically is when you share GTFS data and I guess any open data, make sure you keep the URL the same. So, don't change it to add like GTFS April 2020 after the file name. Just keep the file name the same and then apps will always be able to find and search your latest data, that URL. You want to try to keep your IDs the same across GTFS data sets. So, if you have a route or a stop that's a particular ID now, you want to make sure the next time you produce your data, the ID of that or route have been and stops stays the same. GTFS data sets, you want to make sure that there's no interruptions to consumers consuming that data.

So that typically, what producers will do is merge data sets when they produce a new one. In that way, there's a constant, there isn't like a hard cut off when you share a new file where there might be a gap in service depending on how long the consumers take to load up that new file. So, there are some tools and best practices there that recommend that and then lastly, you shouldn't hide your data behind a login. If you want to track usage, you can use things like TKIPs but using logins really make it all developers hard and we don't want to do that. So, quickly, a few different GTFS related formats here. So, GTFS-flex was mentioned earlier and I think it is really important in the original GTFS data that entered was purely for scheduled service. So, if you have a fixed schedule, a fixed time, you can represent your data in GTFS.

If not, like Sarah shared earlier, you just don't even know if that service existed. So, with the graph teaching us about flex format and the experimentation that VTrans and some of the others did to showcase the new GTFS-flex capabilities. We're now seeing some of those features being adopted into the official GTFS format. So, this is a brand new, right off the press, right around a week old. One of the main features that Sarah actually showed in the Trip Planner having this continuous pickup, where you can flag down the bus and board anywhere along the trip path. This is now officially part of the GTFS standard and Google has started supporting this as well.

So, I think this is also a good showcase for the benefits of open data and open source that there was a pilot a few years ago where VTrans showcased this. Everybody realized how valuable it was and now we're starting to see it and now we're starting to see it percolate into other data producers like Trillium and Trimmed and then consuming apps like Google. And if you want to dig in and see how you can add this to your own data sets, I've included some links here. So, I'm not going to go into depth in GTFS-flex version 1 because I know as Sheryl said there's going to be an upcoming webinar that's going to do this. I'm going to touch briefly here on the newest version which is GTFS-flex 2 which is an open proposal. So, this expands in some of the places that GTFS-flex v1 didn't really touch.

So, things where you have service that doesn't normally operate unless the rider actually calls and books a trip. So, paratransit type services and things are flexible in terms of times. This is something that if it sounds like your service, the picture on the right is an example of like flex service in Tampa Florida where the vehicle can deviate from a fixed route based on call-ins, where you have to call a certain amount of time in advance and in that case the vehicle will deviate from the route, pick you up and bring you and drop you off and all of that. So, this is something I've included a link to the draft proposal at the top here. If this describes your service and you'd like to share your service in this data format, please reach out to me. And then mobility data is a nonprofit, helping to shepherd this in some of the other GTFS extensions along reached out to me or them as well.

I wanted to mention, as we start talking about mobility as a service and which I think is also going to be a topic of a future webinar here, transit I know is branching out beyond just running buses and in some cases performing partnerships with transit network companies like Uber and Lyft and also Bike Share. So, for the bike share world specifically there is a part to GTFS called the General Bikeshare Feed Specification or GBFS, which includes basically the locations and availability of where bikes and scooters are if you want to go rent those, which could include biking to and from transit and many trip planners that are out there today. So, it may be a good first last-mile solution if your transit doesn't serve certain areas.

So, this exists, I'm not going to take a lot of time talking about it but it's something you should know about if you start getting into the bike share world. I also wanted to mention, and this was TCRP project G-16, the development of transactional data specifications for demand-responsive transportation. So, this is another brand new, it's now TCP Report 210. This was just released in 2020 and what this standard is trying to do is standardize the process of actually booking a trip. So, if you have two systems interacting with each other where one, you're booking a trip in one system and then dispatching that trip and talking to say a different company or piece of software what that interaction back and forth looks like.

So, something to be aware of. This is still very new. I'm personally not aware of any producers or consumers. But again, if you want to standardize that type of software, you're looking at preparing the system, I would suggest checking this out. So, what's next? What are kind of the big takeaways that you should take away from this part of the presentation? So, before you're going to start putting out an RFP, if you're looking for help to produce open data, you really want to talk to your peers and understand that, like I said, there's a lot of people that have shared open data. And most likely, you can find a peer that's already done it and learn from them and what worked and what didn't. Mobility data, as I mentioned before, is a nonprofit where some of these organizations are starting to talk to each other.

And it's become a good hub that I would suggest checking out. RFPs and contracts require a few things when you're working with vendors that are implementing systems like automatic vehicle location and scheduling software. Again, you want to make sure that when these systems are exporting GTFS, GTFS Realtime data, that they are doing this regularly when service changes. This isn't something you just do once and then kind of forget about. You want to make sure that everyone's following GTFS best practices. You want to use existing validation tools to make sure that your data is of good quality and catch any potential problems before they become published out to the public. And then if you get different vendors, if you have one vendor doing your scheduling software, you're the one doing your real time system, make sure that they're talking to each other.

And you want to make sure that they can integrate so that things like IDs match between the two. You don't want to have a GTFS data set that calls a route 1 and then a different GTFS Realtime data set that calls at route 22. I want to make sure those can talk to each other. Important to test that equality prior to finalizing any procurements. So, especially try to work with some of the existing big consumers out there, Google, Apple, Transit app, or some of the big ones in the US to make sure that your data looks good before you sign off and say thanks, this looks great. Follow some of those practices that Carol wrote about in the TCRP 115 report for a successful open data program. Then again, if you're interested in GTFS-flex version 2, feel free to reach out to me or comment on the proposal link that I shared earlier. And that's all that I have for the open data presentation. Thank you.

Sheryl: Okay, thank you. We have a bunch of questions. Let me get to that. Let's see if there's any of them still unanswered. No, I think we've asked-- People have asked and answered a bunch of questions, but I have some questions. Sarah, for you, for the open source software, is this often sort of amenable to having a buyers group or regional group, take invest in this type of software together? And then when they update, they can update it together. Is that something that's happening anywhere? Unmute.

Sarah: Sorry. Yeah, it has happened that way. I think having a single contract holder in many cases has been an easier way to implement, even if there's a group of advisors or a stakeholder group of users who are providing input and commentary to that contract holder, even if they're providing funding in support of that but maybe that's sent through a different mechanism. I've seen cases where they've been sort of a regional deployment where each entity has their own contract as well. That can get a little bit messier only in that the renewals may come at different times.

And if they're using lots of different types of funding sources, sometimes that can get a little bit messy in terms of renewals and things like that. So, to streamline things, I do think I've seen a lot of really interesting different types of entities take that leadership role depending on what makes sense in the region. It doesn't have to be a state DOT or it can be sort of any entity, but it has worked out a little bit more smoothly when you've simplified the number of contracts that need to be held in order to create a regional deployment. Any question?

Sheryl: Okay. Thank you. I have a few questions for Sean and Carol. Since you're sort of our generalists, you can feel free to pop in at any time. You briefly talked about licensing and registration, are those kind of forms, whatever you have to do to arrange those, is that pretty easily available?

Carol: I'll give that one to Sean. I do cover-- So, in the open data side of things, there's actually licensing. We talked about it in both of them. There are some examples on the open data side in that TCRP report that I wrote several years ago. Sean, you want to take the report?

Sean: Yeah, sure. I'll actually share one more resource too that was is very hot off the press as of like yesterday or the day before. So, there is a new TCRP Report 213, Data Sharing Guidance for Public Transit Agencies, which is so new and we weren't able to get into the slides before the webinar. So, that does cover some of these topics as well which I think is important to check out. But generally, with on the open data side of things, most agencies today are trying to take a like the least restrictive approach, like the more attorneys you throw in some type of data license agreement, the less likely consumers are to use it because then you got to go hire a lawyer and figure out everything.

So, do the simplest you can keep it is the best and typically, that's like the no warranty is a big one. And then like just don't pretend that you're the transit agency. Don't use the agency's logo, that type of thing, or cannot use common terms. Other than that, you really want to try to put up as few barriers as you can to open data. You want to leave it at an open URL. Some agencies will have a registration page where you can go in and type in and say, "Hey, I'm a developer, here's my contact info," but the link to the GTFS is still open there where anyone can really go download it. It's more of a computer communication mechanism than an enforcement mechanism.

Sheryl: So, I'll ask this of all of our speakers, because we talked about open source software and updating and open data maintenance. And that's when my eyes glazed over a little bit because then it seems like oh my gosh, as a consumer, not only do I have to figure out what kind of software or data, I then have to think down the road, what's going to happen and how am I going to deal with that? So, do you almost need staff looking at that and taking care of those future needs from the very beginning so that you don't hit a wall somewhere?

Carol: So, I'll jump in just overview that absolutely you do need to be thinking about that because we all sort of alluded to this issue of data quality. And the fact that you can't simply open up your data and let everybody look at it. If you have not looked at it closely, and don't know, for example, how accurate it is and whether it's really ready for developers and consumers, sort of on the open market. And so, you do need to factor that in.

And that's one of the reasons why I always say, open data looks like it's free. And I mentioned this in the TCRP report, it may be free to the developer and a consumer of the data. It is absolutely not free when an agency has to spend some amount of labor to ensure its quality. And that part, a lot of people in the early days this that-- That's why I like that question a lot because you really need to factor that into how often you're releasing data and there needs to be at quality checking. Whether you're just starting or you're looking down the road, that's going to consume someone's time.

Sheryl: And for smaller transit agencies, even though there's a benefit of both open data and open source, it almost seems like there's a set of skills that are needed either at the agency or that the agency outsources, if you will, almost on a permanent basis to make sure that these new needs, these new ways of communicating with passengers and making sure everything is up to date, that that gets taken care of. It's not something that you buy or you arrange for and then you walk away.

Carol: Right, right. I think that sort of brings to mind when do you decide that this might be too much for your staff, especially if it's very limited staff? And there, like you said, you might have a consultant on board that helps you take care of that on a periodic basis. I don't know what my colleagues think.

Sean: Yeah, I certainly see especially for smaller transit agencies, where you have one person that's already doing scheduling/customer service/other things. You just don't have the hours in the day and there are a surprising number of, I would say, very affordable consultants out there that will help you create and maintain like full-service models, where they can-- If you find somebody you can trust that can take over kind of that entire picture and there are agencies that have been very successful, I think, working with those types of partners.

Sheryl: And do you see statewide approaches to these are regional or mostly, is it really a transit agency by transit agency figuring out all of this almost like reinventing the wheel on their own?

Carol: So, let me take a crack at that first only because for better or worse, I've been involved in numerous statewide procurements of software. Now, all of those were not open. It was not open software. They were all procurements of proprietary software packages, mostly of more like demand response or paratransit scheduling and dispatching software, and they have all run the gamut from not being successful at all to being very successful. And so, at the state level, there's a lot of guidance that needs to come with looking at one particular solution for all agencies in the state.

So, there are a lot of the details that need to be worked out because all of the agencies in this state may not have exactly the same needs. So, there might be different levels but the statewide approach, like I say, runs the gamut from being extremely unsuccessful for a variety of reasons to be very successful.

Sean: I agree. There certainly isn't always a one size fits all approach. I have seen a number of successful models where the regional or state body provides technical assistance and resources to help agencies. So, those that are kind of struggling. So, I know Oregon DOT has a really good program for that as well as more recently Cali ATP and FTRT providing some funding as well. So, the region or the state level just to provide resources even if it's just to help bootstrap the agency find-- Just spend the time to find the consultant that they can then fund themselves going forward, it's just kind of getting them a step up to getting started is, I think, super helpful.

Sheryl: Sarah, can you talk a little bit about updating the software? From what I've heard, that's a real benefit of open source software. What's involved with that? Why is it a benefit to have open source at the point where you have to update or change?

Sarah: I would say one of the benefits is that the updates, enhancements, new features are driven by the community, and the community are the users. So, the product can evolve organically and naturally to fit the evolving needs of the people ultimately who are using the software. And there's also usually some sort of triage process around what is the prioritization of these things. And there's usually a governing body. Sean Barbeau sits on a number of them.

And this collective unit of that group of folks who are responsible for ultimately coming up with that roadmap. And then sort of the other benefit is that the cost can be shared across different agencies who don't necessarily reside in the same state or even in the same country. So, you can reap a ton of benefits from that sort of collaborative approach to keeping the system up to date.

Sheryl: So, it becomes almost more off the shelf.

Carol: Yes, certainly over time. It has both of those positive dimensions and I'd love for Sean to talk about that. But the features that become standard can very much be off the shelf. For example, one could go and deploy OpenTripPlanner and use sort of the standard features relatively easily and quickly. Yet, you can then also say, okay, I'm comfortable with this standard core set of features that have been around for 10 or 15 years used by 15,000 other agencies. I want to experiment with this thing or there's this new challenge I want to solve.

Let's experiment with that. And then maybe that experiment goes back into the core set of features that someone else can then take as off the shelf. So, there's this very organic living process to open source. I don't know if Sean, do you want to add anything to that?

Sean: Yeah, sure. I think the example you mentioned earlier VTrans, where they have federal funding to implement the flex and they took the existing OpenTripPlanner software that was out there and then kind of went off on their own for a little bit and were like, we want to experiment with this. And then Trimmed did something similar by adding train or Lyft, Uber, TNC type routing to the base OpenTripPlanner software and kind of went off on their own and then they both I think other agencies realized like, "Hey, both these things are really cool. We'd like to do all of that and we think we want to bring that all back together." So, I think that's the great part is that you have a lot of the key stakeholders driving the open source software forward, deciding which features are important, what we want to collectively invest in.

And then also, like Sarah mentioned before, without kind of the resources of having one gate holder that you have in proprietary software, that you're kind of stuck with that company, and it's not as easy to drive their roadmap. They're a for-profit company, they need to make money and they need to make decisions based on that. I think open source software is a little more community-focused and allows, I think, for the public sector to play a role in the decision making process and what is being prioritized and what we want to invest money in.

Sheryl: So, I'll ask I think one final question based on that. Are those for-profit companies now coming around, now that kind of the open movement, if you will, is progressing and changing their business model a little bit so they're kind of they're joining this?

Sarah: I'm happy to defend the for-profit business if we want to have a totally different perspective and that there is totally an always will be a place for a licensed proprietary off the shelf tool. I do not think open source is the only solution or in some cases the best solution. So, I gave examples of business cases where I think open source makes sense. But if it could be and it is very much the case that in some cases, a licensed product that's good and has a long track record of solving a well-understood problem or the agency has no real need to join a community or collaborate. They're not trying to solve a unique or interesting problem.

There are great solutions out there that improve because of the competitive nature of that particular market where they are competing against other proprietary vendors, and then that's what's driving the product to be better. Not sort of the community that's driving the products to be better. So, they're very different use cases. I think public transit because of the regional differences state different and nuances in how we get funding, the changing landscape of transportation, all of those reasons, I think, make open source very fertile ground in the transit industry. But I think we're all better off in some ways, for we all use proprietary products in our own life. I don't think there are many of them probably would not give up without a fight. I don't know, Sean, do you want to comment on that?

Sean: Yeah, no, I 100% agree with that. Yeah, I don't want anyone to take my previous comments as proprietary is always bad. That's certainly not the case. And I think a lot of the other early proponents of open source and transit and Trimmed included would certainly echo those statements. So, yeah, I agree that each has its own role in place, I think, because of the public funding and similar roles of transit agencies. Open source does lend itself well to solve certain problems in public sector space. I think also, just as a trend in the industry, the industry software industry itself has trended towards using open source software. So, even if you're a proprietary product, you're most likely using some type of open source software somewhere in your software ecosystem. So, I think that open source software can also play an important role in the proprietary software sector and for-profit companies in terms of designing and creating innovative products. So, it's certainly not a one size fits all, and I agree with what Sarah said.

Sheryl: Well, that seems like a great place to end because it seems like we could spend a whole day just on the last couple of sentences there. I want to thank all of you Sarah, Carol, and Sean, for excellent presentations and so much information. Our webinar series is really can be thought of as building blocks. As I said, our first webinar was a looking at a framework and now you're seeing information almost like from that foundation level of how do you look at these products that you're thinking of purchasing?

How do you look at your data and what should you consider? And then we'll go even more into that as we go on in our webinar series. Our next webinar will be next month, and it is on GTFS-flex. And then July, we go into mobility as a service. Carol will be on that webinar as well. So, thank you very, very much for being here. For your questions, don't hesitate to reach out to us. Here, you see everyone's contact information in case you want to contact any one of us individually. And thank you again, and we will see you soon.