

RAW TRANSCRIPT

SOCIAL POLICY RESEARCH ASSOCIATES
(SPRA)
RI'S RESPONSE TO THE COVID CRISIS:
HOW SERIOUS DATA SYSTEMS WORK
PREPARED THE STATE FOR THIS MOMENT
MAY 12, 2020
12:30 P.M.(CT)

Services provided by:

Caption First, Inc.
P.O. Box 3066
Monument, CO 80132
1 800 825 5234
Www.captionfirst.com

This text, document, or file is based on live transcription. Communication Access Realtime Translation (CART), captioning, and/or live transcription are provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings. This text, document or file is not to be distributed or used in any way that may violate copyright law.

>> MAUREEN SARVER: Great looks like we have more folks joining the join. Wonderful. We're going to start at the top of the hour but go ahead and introduce yourself in chat and let us know you're here.

Wisconsin is in the house.

We'll just wait a couple more minutes as people join.

Welcome if you are just joining now, please take a

moment and say hello and introduce yourself in chat. We're so glad you are here. We'll get started in just a minute.

Okay welcome, everyone. It's 11:00 here in California. So, I think we're going to go ahead and get started.

As you know, we reached out to many of you in the beginning of February and we learned that although the responses were different from state to state, you reported a high degree of uncertainty and interest in how other states were responding to Covid. So, we shifted in March to focus on working with FTCs in light of Covid related modifications then again in April to focus on emerging Midwest collaborative in the ADRF and in the near real time dashboards that the Illinois team is piloting. Then again, this month in which we're highlighting Rhode Island's response to the pandemic unemployment assistance program. Specifically, how they worked with their long time WDQI partner RIPL how they built in days. How they anticipated helping the Governor with the state emergence from Covid related economic stress.

So, welcome. I think Eduardo, if you are on the slides if you can move to the third slide. Awesome. Thank you.

Before we begin, we've muted your lines so only the presenters will be able to speak. But we do want to hear your thoughts. So, please share your questions that you may have throughout the call into the chat feature. For those that have colleagues that are unable to make it, we're recording this session and we will share it afterwards. A big welcome to Caption First who is doing a live transcription. So, we'll also make that available after the call.

As you can see in the chat, there's a link to where that live transcription is happening.

Eduardo, if you could leave to the next slide. So, we wanted to alert to you resources coming directly from the US Department of Labor. This website has a page with frequently updated information. They have over 20 resources findable via search. It too is updated ongoing. You can sign up for the DOL newsletter and those of several programs in specialty areas in the newsroom.

As the federal staff shared during our last webinar, there is work going on within the department to provide guidance and aggregate across grants and programs rather than individual level program level guidance. Information about this forthcoming. We anticipate by our June webinar if not before.

So, in the meantime, continue to contact your FPOs for specific issues or questions that you may have.

To kickoff us, Kristin is going to invite us all to

take a collective deep breath as we found that doing that has been really helpful for us lately. We'll then take a moment to check in with you all using the chat feature. So, make sure you have that in view. We'll then pass the microphone over to Amelia Roberts from Rhode Island and her partners at research improving people's lives, RIPL, to tell us about the remarkable Covid response they've developed. Then after that we'll open the floor to Q & A using the chat feature. Finally, we'll share what's coming up for the WDQI community in June.

>> KRISTIN WOLFF: Good morning. I wonder if you can back up to the transcription slide for a minute.

So, so glad you could join, everybody. Just a quick word about the transcription. As you know we're experimenting with WebEx here and it's a new platform for us. It turns out if we host live transcription right on the platform everything is very, very small and difficult to see. So, we've included a link here and in the chat. And on your own screens you can go ahead and click into that link then you can manipulate the size of the screen on your own computer. And that allows much more customization. So, I just wanted to make that clear to everybody. You can just click on to the link, open two screens. You can watch the slides on ours then manage the size of the screens on your end. And that should allow everyone to see a much larger image on all sides. Just a quick word about that.

Go ahead, Eduardo, back to the-- I think it was two slides from here.

There we go.

So, you know, just wanted to, imagining that many of you have already been on quite a number of Zooms and WebEx's or Meets or Hangouts or Google Classrooms or whatever you've done today and welcome you here and invite you now to just take a deep breath and just invite you to be present. We're really excited about some of the information that we'll share today and some of Rhode Island's experiences. So, we just wanted to kind of give you a minute to not rush from a Zoom to an X. Just breathe a couple of times. If you could, use the chat window to share an example of how your agency's data has helped your state respond to the current crisis. So, be sure and indicate what state you are in. But just use the chat window to talk a little bit about your experience. How have you been called on to share your data in this time of crisis. And we'll just give everyone a minute there to use chat.

Thanks, Andrew. Yeah, all UI. That's great. Thanks for sharing that, Kentucky. I also wanted to raise, as you know we've been kind of checking in on you. And one of the points

that Kentucky made in the call was the importance of-- the new and I guess renewed importance of communication in terms of press releases and communicating data. It's just a very important time to get data and language around it right. And so, Kentucky had shared they were spending much more time on the communications itself. So, thank you. And everybody just keep chatting.

In the meantime, Eduardo, can you advance the slide. Thanks.

So, as we explained during the April webinar, and also in the newsletters, and I just reminded you, we've been talking to you to learn about how you are responding in the moment to the crisis. We're learning that increasingly all of your work is Covid response. We've had several people say, is there anything else but Covid. So, we know that's top of mind, top of work and top of everyone's priority list. And it's changed the work that many of us are doing. Whether you've been called on to process UI claims or to provide new reports or to share data with new people in new ways, a lot of the work that we've all been doing in the last couple of months has shifted.

And we learned last month that the crisis is prompting more states to collaborate with neighbors and with other partners. It's accelerating the emergence of the Midwest data collaborative. With Illinois, as they shared with us last month, is amongst the first states to begin producing new realtime reports using the UI claims data. While this work may not be exactly what you imagined doing under WDQI a year ago or so, it is clear that the capacity that you've built for improving data quality analysis and generating new insights is paying off. And today we've got another example of what data-centered, rapid-response looks like.

The Rhode Island team has long partnered with the organization they call RIPL. It used-- we used to know it as the Rhode Island Innovation Policy Lab and now it's Research Improving People's Lives which turns out to be a fantastic acronym. But RIPL's helped the state turn administrative data into insights on a whole range of issues. From child and family health to snap, nutrition, college enrollment, college completion, a whole range of things. You've seen many of these on the Rhode Island data hub. The Rhode Island WDQI I team is going to share some of their experiences today. Especially in response to the CARES Act which was the federal response under Covid. And it's posed new challenges you're all familiar with now like building systems for accepting UI claims from people with non-traditional work arrangements.

So, we're going to learn about that today. Amelia Roberts and Justin Hastings and colleagues are going to share what Rhode Island's experienced and identify some of the challenges, how they were overcome and point to new ways to use the state's enhanced data capacity going forward. So, we're thrilled to have them with us today. Maureen will provide a brief instruction to our speakers then we'll get started.

Thanks so much.

>> MAUREEN SARVER: Many of you are familiar with Amelia Roberts. She is the Director of Data and Performance for the Rhode Island Department of Labor and training response for analyzing and reporting workforce outcomes to assist executive leadership in making informed data driven investment decisions. In addition to analyzing and tracking DLT workforce programs, Amelia is responsible for coordinating statewide workforce data projects working with data teams from education, post-secondary, health and human services to analyze federal and state workforce funding. Amelia has spearheaded the effort to transform Rhode Island's traditional workforce reporting methods into interactive tool that's allow policy makers and stakeholders to easily access workforce data and better understand the impact of workforce development investments.

Eduardo, if you could move to the next slide, I want to make sure everyone can see their lovely faces. There you go. Thank you.

Justine Hastings is a Professor of Economics and International and Public Affairs at Brown University and Faculty Research Associate with the National Bureau of Economic Research. Her expertise includes research in industrial organization and public economics which address important economic and public policy questions. She has conducted academic research on topics such as market structure and competition, environment and energy regulation, advertising and consumer protection, consumer financial markets, healthcare, social safety net programs and the markets for higher education

We have Mark Howison who is the Director of Research and technology at RIPL. He has pioneered design and deployment of the RIPL research data link in collaboration with RIPL researchers and government partners. He leads predictive modelling efforts at RIPL and has published research in top general interest journals such as the proceedings of the National Academy of Science. He's contributed important research on topics such as predicting opioid dependence before

prescription is given, identifying geographic partners in the spread of infection diseases to inform Public Health policy, using AI to nudge low income students towards college.

Without further ado, I will pass it over to Amelia Roberts

>> AMELIA ROBERTS: Great. Thank you for the great introduction and I'm happy to be here and part of this great webinar and share all the wonderful things Rhode Island is doing in response to the Covid crisis.

Before I hand it over to Justine and Mark to get into the nitty-gritty details of the innovative solution that they've helped Rhode Island build, I just wanted to provide a little more context to Rhode Island's journey through WDQI. And really that starts with how Rhode Island is still realizing the value of integrated data. WDQI has really helped us through the years through additional funding, supporting the infrastructure of our state longitudinal data system which is pretty mature at about ten years old. And recently with this most current round of funding we've focused on how we can demonstrate the value of longitudinal data system and identify opportunities to leverage existing state systems so we can increase the accuracy and the data we are collecting while simultaneously streamlining the way in which we're collecting and analyzing it. And so beyond utilizing our SLDS system we're also looking at how we can implement innovative solutions to build realtime responses to specific needs such as this Covid crisis.

So, in the slide you are looking at right now, these are really our high level WDQI goals. A lot of this I have hinted towards improving the quality in the breadth of our data and what I was hoping to do in these initial slides is really help connect the dots for the participants on this call. And then also give a high-level overview some of some of the other innovative solutions we're working on as far as different analyses and different technological solutions that we've developed under WDQI.

So, I think we can move on to the next slide.

Great. So, I think many of the folks on the call have also been part of some previous presentations on our talent dashboard. This is really the tool that helps make our longitudinal data accessible to stakeholders and policy makers. One of our major buckets of work, which demonstrates our commitment to using link data, to value our long-term outcomes is this dashboard. So, we link data across K through 12, postsecondary and the workforce. Just to give you a quick example how we're using linked data is we're looking out

outcomes for technical and educational students and that would not be possible without the use of integrated data. Those have actually been pretty positive outcomes. And this tool also allows us to look at the different outcomes by equity lenses and by different districts, schools, institutions and industries and allows users to drill down into the areas of opportunities and success that we want to point to.

So, this is really all about making data accessible and actionable.

This second example of how Rhode Island's using the data is thinking about new solutions that can help Rhode Islanders is this new data management system we developed under the WDQI umbrella. Real jobs Rhode Island is Rhode Island's marquee sector-based program that workforce program that meets employer needs through strategic partnerships, that trains and places individuals in jobs that actually exist. So, as the program has gained momentum over the years it became very clear that we lacked the technological tools to officially manage and administer the program. What is complex about this program is that we are braving both state and federal funds to administer the program. And as you may know the reporting in order to break funding can be overwhelming.

This system was actually-- this is going to help kind of connect this discussion to what Justine and Mark are going to go over is this was our first endeavor in utilizing Amazon Web Services cloud-based storage. So, this system is fully housed off AWS. As well as the storage we rely on the AWS analytic platform, quick site, to quickly pull and access and query the data.

So, again, it goes back to making the data accessible to both external users but also internally at the DLT.

So, here we're illustrating how the system along with other components of our performance management workflow create this continuous improvement cycle by bringing realtime data analytics to this process in really one central location.

Just to give you a little background.

We used to collect data on several different spreadsheets that were not linked. It made it very difficult to do any kind of robust analysis. You know, our data wasn't connected to any of our other administrative data from our other divisions. And we didn't have individual-level data for any of our partnership or activities that were going on.

So, with this system, we will provide to our friends on the federal side of the house the data they need for reporting purposes and provide, you know, researchers such as RIPL with

a comprehensive data set that includes accurate participant level data.

So, in addition to developing that data management system Rhode Island identified the need to use integrated data to evaluate the effectiveness of its workforce programs. So, what we also identified was that previous too much of the work we've been doing in the earlier rounds of WDQI was connecting data across multiple agencies but there was still this need to better align our own systems internally, so we better understand the effectiveness of our programs. So, we focused on connecting our internal administrative wage, unemployment, TDI and TCI data in an integrated AWS environment so that we could measure the return on investment for real jobs programs and its participants.

This was an extremely innovative solution because it provided RIPL researchers secure access to the matched data used wage data researchers were able to identify a control group of individuals in a similar industry with similar wages that could be compared to the treatment group of our real jobs participant.

Just to give you a high-level overview of what you are seeing here. This slide is showing some of the results of both the treatment and control group looking at pre and post training wages. Where you see the lines, split is when the wage with individuals start to vary between the treatment and control group around the time of training and post training. So, these are the solutions and analyses we're able to do now that we're linking data within the DLT.

So, similar to many other states, when the Covid crisis hit we saw a huge influx of UI claims come through. We were seeing our call centers inundated with inquiries as we knew our antiquated system couldn't handle the volume of claims. We also recognized that in order to implement PUA benefits it would take weeks if not months to implement it using our current system. So, we had to act. And thanks to the visionary leadership much our director Scott Jenson as well as Justine, her amazing team from RIPL, we were able to create a new application that was hosted through our existing AWS Cloud environment.

What you are seeing on the slide here are some pretty interesting statistics on the claims we received. So, since the start of the crisis, we've beaten our previous record of 5300 claims that came in during the banking crisis in 1992. And that happened twelve times. In just the first five weeks of the crisis, we had already 144,000 UI claims.

So, looking at this data we knew that we had to do

something, and we needed a quick innovative solution and we needed help from our partners at RIPL to build something that could accommodate all of the new claims coming in.

So, to actually go into, you know, the architecture and describe the solution in better detail than I ever could, I'm going to hand it over to Justine and Mark to describe their organization and how they helped us tremendously in building this great tool. Thank you.

>> JUSTINE HASTINGS: This is Justine. Can you hear me. Great. Can you advance to the next slide please? Thanks.

Thank you, Amelia, that was really a great overview of the important work that Rhode Island DLT has been doing. And they've been such a wonderful partner to work with because of their focus on really wanting to embrace technology, data and science to improve policy and lives. Which is RIPL's mission.

So, RIPL is a tech nonprofit, a social good nonprofit, that's led by faculty scientists and policy experts. And we work with state and local government across the country. Rhode Island being one of our closest partners. And we work closely with government leaders. We also form cross-sector partnerships. For example, with AWS or other partners that bring really important skills that in combination with our skills can really solve people's needs with government through improved policy. And we really specialize in solving policy problems with data, science and technology.

And the approach that we take, which I think will become apparent as we talk through our current projects and our partnership with Rhode Island on the pillar response is that we really integrate research with data infrastructure creation and technology so that we can deliver a scalable, sustainable, measurable policy improvement.

So, coming from a research science background, one of the kind of traditional models of research in academia and often in public policy is that research kind of happens with a static data set that was maybe formed in isolation for a particular partner on a project. And then that research is done with that static data. Meaning that it's kind of an extract from a data that's handed to a researcher and it's in a flat file. And there's some analysis done. And there are research findings that are published. And they may be important, but they don't have a way to create a virtuous cycle of research, technology and policy improvement, testing, measuring and innovation and more improvement. And if we were to look, for example, in the private sector when data is used intensively, it's typically used in a process research and-- research and technology and data are used with the intent of

actually creating a policy change that then be measured, feedback, improved, and this creates kind of an constant innovation and improve cycle. The way that you do that is that you actually have a data infrastructure that research can happen on. And that data infrastructure is optimized for insight. Lots of typically used variables and definitions are created in higher structures, in structured tables. There's code books and definitions. And those are created, and they live, and they're constantly updated. And researchers work on that, develop insight, and then whatever that code or that production is can then-- whatever that code is, or science is not only can be a research publication bit it also can be taken directly into production so that it actually becomes part of the policy.

So, to give you an example of what that looks like we have National Science Foundation funded project on the future of work. We have four state partners with a handful of others that we're bringing on board for this project. Essentially what we do is measure the causal return to the impact of training and re-skilling programs with administrative data, linked administrative data. And then because we're doing that process in this cloud-based environment, those measures actually can be constantly updated now as software and they can be pushed into an easy to use web and mobile interface which allows workers and job counselors to make informed choices about which programs are likely to add value that deliver skills and gainful employment. And policy leaders can also access these and know which programs work then invest in them.

Government in this system owns the data and the platform and the software going forward and they are able to use it and extend it and share insights through wonderful forums like this one where they can gather together and learn from each other and through that ability to own and learn we've built capacity so that those results are ongoing. They're ongoing improving people's ability to make wise investment decisions. They're leading to improved programs and innovation and the result is better opportunity and outcomes for all workers per dollar spent on training. And the returns that Amelia showed you on a slide a few slides ago were a great example of the type of output that can be generated not just as one static result but actually as continual results and then easily pushed through an API so that workers can use them, policy makers can use them, and training providers can use them to measure their own success and look for ways to innovate and improve.

And all of that is possible because of the Cloud technology that we've developed that runs on AWS and the research data links system. And it was essentially that system that we had set up for this future work project that we were able to quickly pivot to a new system for handling PUA claims. And so that's the other nice thing about setting up this type of environment is not only can it be used for partnering with researchers, scientists and nonprofits to help development science services, to develop science informed policy but it can also become a way for departments and agencies to quickly add, expand, adapt to new programs as circumstances or crisis demand.

So, I'm going to turn it over to Mark Howison who is the Director of Research at RIPL who is going to walk you through details of how this type of approach and system works.

>> MARK HOWISON: Thank you, Justine. Can everybody hear me okay?

>> Yes. Thanks, Mark.

>> MARK HOWISON: Great. Can we go to the next slide, please?

So, here is a detailed system diagram of what Justine mentioned earlier. This architecture we call a research data link. I'm not going to go into all the details of this but I'm happy to follow up with everyone who would like more details afterwards. But I want to explain this to you because this was really the foundation of our ability to quickly pivot and address the PUA claims. So, this is the system that we created under the virtual work project that Justine just described to you. And the key features that I want to highlight for everyone is that it is a secure Cloud environment that includes a lot of features around governance and controlling who accesses which data for you know what purpose. And also has the ability to integrate and anonymize data. But in the research setting that we were just talking about, often you don't need to be able to identify data. You can conduct research and gain insights without knowing the actual identities of the data you are working with. Obviously to process PUA claims you do need to know people's identities because you need to pay them. But what I want to highlight is this secure Cloud infrastructure was already in place. So, we found a way quickly to start up the project to process PUA claims. And you can almost think of it as just another project that's going to happen in this research data link. Just like the future of work is one project, you know, with specific governance around which data we use to calculate return on investment, measures for training programs. The PUA claims

system is another project. The goal of which is, you know, less research one and more of an operational one which is to collect and process PUA claims authorized under the CARES Act. So, having this facility already in place with all the security and auditing and logging and encryption and governance processes and, you know, the ability to create user accounts with restricted access, these are all really central to the speed in which we're able to respond to director Jenson's request that we help stand up this PUA claim system.

Here what I'd like to point out, there's a lot of information on this slide, but I think the most important thing is that third bullet point which is emphasizing how the system is designed to it's actually owned and operated by Rhode Island DLT. It's not owned and operated by RIPL. We're just a contractor who under a specific contract with Rhode Island DLT are conducting, you know, basically work within some scope defined to do the future of work project, we're able to just quickly put an addendum in place to that existing contract to say we're also going to do a second project. Which is to assist you in quickly setting up a pro system to process PUA claims. So, this was much easier because provide DLT actually owns and operates this platform they can just bring us in as a contractor to help with this urgent need resulting from this crisis.

And this has a lot of advantages for both an operational and urgent need like that. But also, just longer term and supporting research and data driven policy and insights. It's much simpler to bring in, you know, the right partner to solve the policy problem you have if you have the data already. You have a system where you can grant them access under your improved project. As Justine was mentioning where you've already done a lot of the data cleaning. You've generated code books and organized data in such a way a new research can step in and immediately understand what is there and what they have to work with. That gives you a lot of power to be able to, you know, work with kind of the best researcher to answer the problem that you have and to get up and running quickly. And it also simplifies a lot of the legal complexities to arise if you take this more traditional model where you are going to partner with an external researcher, and you are going to send them data. You are going to share your data with them under what is called a data sharing agreement. Because at that point you now have to trust all of the security processes, they're going to put in place to protect your data. And, you know, that requires a lot of, you know, for a legal instrument.

In this scenario here where Rhode Island DLT owns the

system they're just simply granting access for somebody like RIPL to come in and work on their data and they're maintaining control over that data at all times. And all the audit logs and all the security controls are available to them and everything's very transparent in a way that it wouldn't be if they had to sign a data sharing.

What this meant was because all these security controls were in place and because we're able to work with, you know, sensitive data such as PII, SSNs, bank account information, et cetera, in this environment, we were able to develop the PUA claims application, which, you know, collects that type of information from applicants or claimants and then stores it for processing. We're able to do that all within the same account.

And I think the key things to point out for-- I think it was about ten days that we spent in development to get to the production launch of the system. The key things I want to point out are that that was possible because of a lot of what are called managed services that exist in Amazon Web Services. So, these are essentially services that Amazon has developed to solve problems that they've seen before, you know, through their position as a major retailer, web retailer, you know, on the internet and through supporting all of their customer's needs, they basically identified there's these common technological solutions that people often need. So, one of those solutions, for example, is to on-demand spin up some computing power to handle requests from a website that somebody's interacting with. So this is a service called lambda that allows you to basically as, you know, a customer in maybe the retail setting or here a claimant in this setting with the PUA claims, is using a website, wants to submit something to you and you are able to quickly spin up compute instance that basically processes and accepts what they're submitting. The nice thing about it is they manage all the scalability for you. You don't have to worry about how many servers do I have, how much memory do they have, how many processors. Is it enough if I suddenly get a spike if 10,000 people access it in an hour, do you have enough resources? You don't have to worry about that. Amazon's service is handling that for you. It's automatically scaling as more people start using that site. So, that was key in being able to do that in ten days.

The other thing I want to point out on the slide, this is the third bullet from the last where we're basically proposing science as a service model. When we think of these scientific analyses like the return on investment one Amelia

was showing earlier it's really valuable to be able to do these on an ongoing basis. It's not just a one-time analysis that we did, and it gets published in the journal article somewhere and nobody is ever to use it again. It would be so much better if we could take something, we figured out like Justine was describing and put it into production so it could be an ongoing capability. And what we imagine here is that, you know, it could be that there are other partners that governments have like training providers, for example, who might want to be able to use some of these results. But, again, it's traditionally been hard for them to access administrative data because it's sensitive and it has PII and, you know, because of the legal requirements for sharing data. So, we envisioned this way in which, say, a training provider could have a system at Rhode Island DLT where they could submit information on their enrollees. Which might include PII. It's processed in a service account where it doesn't reveal any of that PII to Amelia or anybody of Rhode Island DLT it's just used to do a match and return a scientific result. For example, a return on investment measure for that particular program enrollee in a program that that provider runs.

So, this is a way in which we see leveraging another real benefit of the Cloud which is that it's somewhat of a digital common. It's a place where different organizations can come to collaborate and share data in a safe and secure way. But it's not like one organization has to grant access to the other organization to their infrastructure. Which is typically hard to do with you have on-site infrastructure. So, it really is leveraging this kind of nice feature of the Cloud as this digital commons or this area where, you know, different organizations can get together to collaborate.

We prototyped this actually in kind of a maybe I wouldn't call it like a complicated scientific analysis but a pretty simple analysis where it was really going to speed up the processing of quick claims if we could verify claimants self-reported adjusted gross income. So, because they're goods workers or independent contractors they're not traditional UI claimants, they're not going to necessarily have a wage history. In the regular UI program, you'd take a claimant and you'd look up their wage history and you could verify how much they made through their wages. You can't do that here. So, instead what we're doing is asking the form we put up we're asking the claimant to report their adjusted gross income either in 2018 or 2019. Obviously, people are still submitting 2019 tax returns as the system went online but report

whichever one, they had. The Rhode Island DLT had to somehow verify that information was accurate.

You can imagine trying to do some kind of manual verification. People are going to send us photos of their tax return or mail us a copy. You can imagine how that's not scalable at all. We're not going to be able to quickly process and pay benefits to, you know, this large population that's in crisis and in need. So, we're trying to find an automated way to do it. We came up with a solution where the Rhode Island Department of Revenue who is very on board with being able to support this type of application they set up another AWS account they owned and operated and they exposed an API end point that the Rhode Island DLT system could query with this science is a service type of approach I was describing. The analysis is very simple. It's going to match a UI PUA claim to a tax record and then return the AGI from the tax record. But what we're able to do is use, you know, the minimal amount of PII required to do that match. We do it in a service account where it's not accessible to anybody at the Rhode Island Department of Revenue. Then we get back the AGI result. Again, the minimal amount of information we need from the tax record to do the verification. And as a result, we were able to very quickly validate and process a large number of PUA claims. I don't have the exact numbers on hand, but it was something like 80 percent of the claims were able to automatically find-- obtain a verified AGI.

So, I think that covers pretty much everything on this slide. I believe this is the last slide. So, I'm going to stop there. Amelia, Justine, is there anything you want to add?

>> Amelia: So, I think you gave a very extensive description of the solution. But I think we do have some questions in the chat, Mark, that you might be able to speak to.

>> MARK HOWISON: Okay great. So, I'm going to open the chat. So, looks like the first one is from Stephen Todd. How do we generate the global ID. To replace SMM with a random number. Yeah, you're essentially on the right track there. We used what's called a deterministic algorithm. So, we look at basically the combination of the person's Social Security number, the first letter or sound of their first name. There's an algorithm called sound X which tells you what the sound of the name is which is nice because it's robust to nicknames. Then the full last name and the date of birth. We use those four elements to generate exact matches.

So, you can think of this as, you know, records across Rhode Island DLT's different data sources. So that includes

the UI system and also includes worker's comp and training programs and temporary caregiver insurance, TDI. All these different programs. We find records that have those same four data elements and then we assign them what is essentially a random number and that becomes that person's anonymous ID

>> JUSTINE HASTINGS: I also want to mention in the research data link system we also encrypt the SSNs with-- we hash it. So, we're matching on hashed value. We don't keep those types of sensitive-- basically the whole software is set up to run end to end. So, it does the matching, the creation of drive tables, the code books, anomaly detection in the data as you run the software instead of doing things by hand. So, essentially it runs all in one go and does all of this on the fly in one step. So, in that way it's very fast. It is robust and it is something that can be implemented by the agency going forward. And it takes extra precautions so that we're not leaving sensitive information for any period of time in a state that people could see it, or it would be exposed to, you know, the naked eye, et cetera, even within this super secure system.

One other thing I wanted to mention that I think Mark mentioned in passing is that, you know, you can't download data off of this system. It has to go through an export server that DLT has-- Amelia actually is the person who is the exporter, reviewer and exporter. So, that's really great because you, you can partner with researchers. They can't take data off the system. You have full audit record logs of what's going on with your data and at the end you can review the output to make sure it's not disclosing sensitive information. Then that final output can be taken off the system. So, that's another way in which there's real security in governance and complete transparency in the system

>> Amelia: To build on top of that, I think what I really love about this system is that if we were to bring in other agency's data that in a traditional integrated data system you still have some human interaction or somebody that is seeing that identifiable information. But if we were to have another agency's data in this, you know, I wouldn't be able to see the identifiable information. They would upload that and then the hashing and the matching would all happen behind the scenes so that we don't have to worry about any security issues there. So, that's what I thought was truly innovative about the data links solution

>> JUSTINE HASTINGS: I think somebody else asked any thoughts about extending the system beyond agencies.

We have additional state partners for the future of

work project. Those include Colorado, Virginia and Washington state. And we're working with more states. And we also have some new state partners that we're working on a bunch of health and human services, child protection and health projects with as well. So, the whole goal of making this entire system as software in a Cloud-based solution is not just all of these benefits we've talked about so far but that it's totally scalable.

So, our NSF funding provides funding for essentially productionizing the suite of Cloud templates and all the software that goes with it so that actually any state can-- or local government can pick this up and use it. I think we also forgot to mention that the PUA, the cares app is what we call it, all of that is available to any state in a package. All of the code and pipeline. We've provided it to NASWA, and they've made it available through their marketplace. So, it's open source for nonprofit and government uses.

>> MAUREEN SARVER: I think we have another question from Carol. Is Microsoft also deploying a [inaudible] approved infrastructure?

>> MARK HOWISON: Yes, this was specific to Amazon Web Services but other Cloud providers like Google and Microsoft, you know, conceptually all this is portable. There would be some re-factoring of our code to make it work with another Cloud provider but certainly it's possible.

>> MAUREEN SARVER: Okay. Then what is the success rate of the matches?

>> MARK HOWISON: Of generating the anonymous ID? Yeah so, we find that depending on the agencies we're working with or the quality of vendor link data we see between 95 to 99 percent accuracy. Which for, you know, research purposes and insights it's sufficient. You know, obviously if had an operational system where you need [inaudible] exactly you would drill down further. But the reason we chose the deterministic approach is because it's pretty consistent and reliable and has worked across a lot of different partners we've worked with.

>> MAUREEN SARVER: Looks like we have one more question from Carol. But we have a few more minutes. So, if other folks have questions, feel free to type them in the chat box.

So, Carol asks, did RDL build the software or use a combo off the shelf for this?

>> MARK HOWISON: So, there's a couple of components to it. One was something RIPL did build. A system called secure infrastructure for research with administrative data. So, the acronym is SIRAD. That's open source and does all the data

integration and anonymization we were just talking about. It's available open source for nonprofit government use like Justine was mentioning. That part we've kind of already built. Then the software that got deployed, the Cloud infrastructure, we built in collaboration with AWS and some of their solutions architects who also helped Amelia deploy all of this.

By the way we should give a lot of credit to Amelia because she single handedly deployed this entire user style

>> Amelia: Oh stop

[Laughter]

>> MARK: She did. She's being modest but that was all-- there was a code that's been packaged in these things called Cloud formation templates. And AWS you can take infrastructure you've built like in a test account, for example, and you can package it up with a template then Amelia used these templates to then deploy the same architecture over in the Rhode Island DLT account. And that's all through this NSF grant we're going to be making available more broadly.

Then the next question from Carol is about the match rate. Is the matching stats available? Yes. That's part of the package. I'll get the URL and place knit the package.

>> JUSTINE HASTINGS: The next question: Do you have any data files that are missing data that hospitals don't collect birth dates how do you deal with them. So, the way we do this is the more agencies data that you combine kind of the more pieces of a puzzle you have for the software to determine who is a unique person. So, one agency might have Social Security number and, obviously, it's hashed so they have a hashed SSN but not a birth date. But another agency has an SSN and a birth date, a hashed SSN and a birth date. And another agency might have names and birth dates. And by combining all of those data you get closer and closer to having, you know, this very high and precise match rate. In Rhode Island, for example, Department of Education doesn't collect Social Security numbers. But, you know, they have very good other data on children and individuals. And by taking, you know, different types of matches and having different algorithms for matching that we may assign different qualities to between the data we're able to kind of piggyback off of other pieces of data in order to find the right match.

>> MAUREEN SARVER: Thank you, Justine.

Before we move on to just wrap up with some announcements, Kristin, do you have any lingering questions?

>> KRISTIN WOLFF: I actually have about 30 questions. And I don't have room for all of them. So, I'll just ask this.

So, you've all spoken to this, what is really kind of a

secular change in how we think about using data to improve programs on the research side and on the application side. And you've identified a couple of tools like the town dashboard and things before this crisis with the intent of changing policy, improving programs and also in an implied way kind of changing behavior. And I wondered if you could kind of speak to that issue.

When you think about the tools you've made available, the way that you've collapsed the space between data used for research and data used for programming improvement, and then built tools to have that occur on an iterative basis, I wonder if you can speak to what's been the behavior changes in terms of Excel rating policy, launching new programs, new people talking or working together. What does all that look like? I know that's a really big question

>> Amelia: It is a great question. I think it might be too simple of an answer but it's really increasing, you know, data literacy throughout the DLT. And for leadership data has just become part of daily operations. Before it was kind of segregated in terms of like okay the data performance unit or the data folks on a workforce set of things are doing it for the purpose of reporting to the Federal Government or, you know, working with outside researchers on key studies and what we're really doing here is bridging that gap and then taking that data and making it usable in realtime. And so, for example, you know, some of these results could be used in what we call pitch meetings. So, these are when new ideas come in from our partnerships it allows us to evaluate their performance in realtime and then think through our investment decisions based on that information. Are they coming in asking for additional funding for programs that, you know, didn't show a big return in previous cohorts? And if we do decide to fund those, are there changes they need to make before we make that final decision. So, that's just a way we can incorporate these analyses in our daily work and in our investment decisions in our evaluation process.

It's also a way we can use it internally for our performance around accountability and when we're looking at performance by, you know, our grant advisors who are managing these programs looking at performance, comprehensively for subsets of staff. So, I think there's, you know, we can use it for external and internal evaluation as well as part of our staff program that we have that I spoke to earlier in the presentation.

So, I think there are a lot of use cases and are more in the way we can expand on that. But as it becomes accessible

and we're able to respond to requests from executive leaderships very quickly, I think that is just going to organically enhance the consistent use of data in everyday operations and conversations.

>> KRISTIN WOLFF: Well this is very exciting. Thank you so much, all of you. Go ahead, Maureen, and close us out and make sure you show the slide that has the resources on it.

>> MAUREEN SARVER: Sure. Be on the lookout for a post-webinar survey coming soon. Podcasts. Those who have already scheduled we'll be in touch in the coming weeks. June webinar will be the round VII showcase. You don't want to miss it. Find a focus on the recent Covid related amenities. We have extended it, so we don't have to rush it.

This is where you can learn more about what you heard today. And we'll send out these slides so you will have these resources. And again, just thank you so much, Amelia, Justine, Mark, thank you for sharing your insights. This has been a fascinating conversation and I'm sure we'd continue if it wasn't already noon. Thank you to all the states for joining and taking this time out and for doing all the good work you are doing in your states to respond to Covid. Please reach out with any questions and take good care.

Thank you.

This text, document, or file is based on live transcription. Communication Access Realtime Translation (CART), captioning, and/or live transcription are provided in order to facilitate communication accessibility and may not be a totally verbatim record of the proceedings. This text, document or file is not to be distributed or used in any way that may violate copyright law.
