

Washtenaw Equity Partnership (WEP) Data Subcommittee

May 11, 2022 4:30-6:00pm

Attendees:

MaryAnn Sarosi (co-chair)
Eli Savit (co-chair)
Grady Bridges
Judge Darlene O'Brien
Linda Rexer
Angela Tripp
Trevor Bechtel
Linh Song
Jeff Rose
Truly Render
Sandhya Kajeepeta (Vera)
Alex Roth (Vera)

- Vote to approve last meeting's minutes
 - Minutes unanimously approved.
- For the majority of the meeting, the subcommittee's two subgroups (mapping and modeling) split up into separate conversations.
- Mapping subgroup:
 - Survey of government agencies about criminal legal data
 - After last subcommittee meeting, we sent an online survey to government agencies that may hold criminal legal data and asked about the types of data they collect and how it is stored and managed.
 - The subgroup reviewed responses thus far and noted important observations, such as:
 - Most organizations reported not having the capacity to analyze their own aggregate data and assess trends.
 - The organizations that did collect a personal identifier all collect a different type of identifier. It should be a priority for our recommendations to develop a common identifier so that data can be linked across units.
 - Discussion of who we want to follow-up with to hear more about the data/metrics that would be most useful for identifying and addressing racial disparities
 - Suggestion to begin with the agencies who said they have the ability to analyze their own data. This would help us understand what they are analyzing, for what purpose, and what their main challenges have been.
 - In this follow-up conversation, we should clarify that we're not just asking if they have historical data. We want to understand, if you do have historical data, can it be queried, and do you have staff capacity to analyze it?
 - Next steps:

- Sandhya will draft an email that would go out to stakeholders to invite them to have a conversation about data. Sandhya will send this draft email to the subgroup for their feedback.
- Models subgroup:
 - Subgroup meets with Joe Brookins from Multnomah County Public Safety IT who works on Multnomah's data warehouse called DSS-J.
 - Everyone introduces themselves
 - Joe provides his background – he is a database developer and has been around since the beginning of the DSS-J project in 1998.
 - Subgroup starts asking Joe questions
 - Q: Could you tell us a bit about the timeline of the project?
 - A: In Oregon there is a mandate that every county has a local public safety coordinating council (LPSCC), that includes every public safety agency and private nonprofits—it has about 80 members and includes representatives from Portland, the biggest city in the county, as well the other cities and smaller towns. It includes the sheriff, circuit court administrator, presiding judge, and city and county commissioners. For the data warehouse project there was subcommittee of the LPSCC, which had about 13 members, including the sheriff, the DA, the public defender, and was chaired by a presiding judge and a county commissioner. Back in the early days, there was one judge who had a vision for a project for smart sentencing so he could determine what the best sentence that would serve the person in front of him. He became a champion for that back in the mid-90's, and I think that's why the data warehouse idea wasn't a hard sell for the LPSCC—members saw the benefit of having data sharing and a central repository of data. That decision was made at the top policy level. The next thing that was put in place was an operations committee, made up of researchers/data scientists from all of the agencies that supply data for the warehouse. The operations committee was involved in creating policies about how they would share data, who gets access to the data, and who gives approvals for the cross-agency analysis that gets done. And then at the user level, we have the research and evaluation teams from each of those agencies—a smaller group of people that digs into the data and disseminates it to a wider audience. When this project started in 1998, they hired a consulting firm that was really experienced in creating data warehouses and created the intergovernmental agreements among the groups involved. It took about two years to get that first version off the ground. At one point there was about 11 staff on board. I think, in some ways, we overdesigned it to begin with, and data warehouses are expensive. The original version was bit more than we could maintain, so over the years we've scaled it back and now we have two FTE's to maintain the warehouse. When a big project comes along, we end up hiring two or three extra contractors to help out. Over the 25 years of the project we've rewritten the thing several times. The initial stage of creating it took two years, and then we rewrote it in another two-year project in the early 2000's, and then again in 2012 we moved things to different platforms. We've been on the same platform now since 2012. We've five source agencies for the warehouse—the sheriff, police, DA, courts, and Community Justice [probation]. Over the years, each of these agencies have changed their data systems, and we find that it usually takes four to six

months to get those integrated; that's usually when we hire some additional help. Then we've also had technical issues with the Portland police data. They changed their data system and we've never been able to connect that up to the warehouse. I think they're redoing it again, so maybe we'll be able to reintegrate their data in the future.

- Q: Are any of the police agencies involved?
- A: They're not at this point. We had data from the Portland police until about 2015, when they changed their system and we have not been able to get reconnected. I think there were technical issues more than politics, but politics does come into play.
- Q: How many police agencies are there?
- A: In Multnomah County, I think there are four. The sheriff's office has a law enforcement branch; there's the Portland police; the City of Gresham, which is another large city that has its own police agency, but they tend to use the Portland police data system.
- Q: Do each of the public safety units that are involved (e.g. courts, sheriff, prosecutors) have their own research and evaluation personnel?
- A: Yes, there's generally one or two research/evaluation people in each agency, though some of that is fairly recent. A lot of those agencies are mainly doing analysis of their own data, which they can share with other people. But when you need to do cross-agency analysis, that's when the warehouse comes in.
- There's a prosecutor's dashboard for Multnomah County. Does that run off of the warehouse or is built on the prosecutor's system?
- A: I think that's off of their system. The Portland police also have a lot of deidentified data publicly available on Tableau, and you can download the data.
- Q: So, the warehouse isn't being used to put data out to the public? It's a research tool for behind the scenes work?
- A: That's right. For the most part, what we use it for is internal and it's never really shared out.
- Q: You have these publicly mandated coordinating councils, are warehouses a part of that in other counties or does each council sort of set its own agenda?
- A: I think that Washington County, which is another big county, may have something, but most of the counties in Oregon are very small and don't have the resources to create something like that. At the state level, the courts, state prison system, and Community Justice have centralized state databases. They have kind of an old system, but they are working on some dashboards there.
- Q: Do you have data from other agencies like health and human services or schools in the warehouse?
- A: Not in ours, we haven't done that yet. It's always been a dream to include those. It's something that's been talked about for a long time, but there's always been hesitation about combining HIPAA data and CJ data. There are some efforts to do person matching across systems, but the LPSCC and the data warehouse subcommittee seem to have shut down the idea of actually adding health data to the warehouse. Some people on my team do work with health data, but we haven't combined it.
- Q: Does the LPSCC have analyst staffing to do analysis on cross-system data, e.g. looking at cohorts or individuals over the span of their involvement in the criminal system?

- A: The LPSCC has about five staff people, and there's a research and evaluation person in that group. That person is really good at coordinating with the research and evaluation people from other agencies.
- Q: Could you tell us any more about the genesis of the project? How did it become publicly mandated and how did it take shape?
- A: The data warehouse was not publicly mandated, just the LPSCC. I'm not sure when those became something that the state required. But having the LPSCC did set us up to have a place to go to work on the idea for the warehouse project. Multnomah County having enough financial and human resources was also a key factor in the project.
- Q: I read somewhere that there was bond issued to fund the initial warehouse?
- A: Yes, I believe it was about a \$4 million initial bond. I'm not sure what the annual cost is. I used to think it was around \$750,000, but someone said seven figures now so it might be over a million.
- Q: If we were going to ask someone for that information, who would we ask?
- A: Abbey [Abbey Stamp, director of the LPSCC] would know that. Actually, we're in the middle of budget season right now, so I could probably dig that up in a budget.
- Q: If a judge wanted to know something like whether sentencing agreements are only being offered to certain groups of people or whether they're being uniformly offered to all, would the judge be able to go to the warehouse and ask for that sort of information?
- A: I think we've actually done something like that. Our source systems do track race and ethnicity and we've tried to track who's doing the data entry—is it self-reported or someone's perception? You might have people recorded differently in different systems and there may be changes to that information over time. When we bring that data together and do the person matching, we look at how race/ethnicity was reported in each system and then try to weight that. It's very rare that there's no agreement—that each system recorded something different for race/ethnicity. We did create something that shows that weighting system and how you can use it to match race/ethnicity data across systems.
- Q: Is the question about whether a judge could ask the warehouse for that type of data something that we could also ask Abbey?
- A: I would think so. Also, David Mitnick is the evaluation person on our team. He's only been here around eight months but he's really good, so he might be another person you could ask.
- Q: Right now, is it a private company that holds the data warehouse? Who is actually responsible for holding that data?
- A: It's Multnomah County and the LPSCC. The LPSCC is still a part of Multnomah County government but it has its own separate funding. The LPSCC is the one that pays me.
- Q: Is the LPSCC state funded?
- A: I'm not quite sure where all the funding comes from. I think most of it is from the Multnomah County budget, but there are so many different sources for the budget.
- Q: I'm curious about how electeds use the data warehouse. Have you seen them use it in crafting policy in a timely fashion that's responsive?
- A: What we see a lot is what we call "drive-bys," where an elected will walk by the researcher's desk and ask if they can tell them what they're seeing a particular area. I

think the research and evaluation teams in each agency get ad hoc requests like that. I'm not quite sure how that affects policy. There are also specific programs where the state asks for data. There was justice reinvestment program about five or six years ago where the state wanted to invest more in Community Justice to avoid building more prisons. For that program, they wanted to closely track a specific group of people in Community Justice programs and they provided funding to the county to get that program off the ground and track the data.

- Q: The Portland police aren't sharing their data with the warehouse, but do you know if they've ever asked for data analysis, even though they're not part of it?
- A: I think that they haven't recently. They were very involved up until they time of their system change. They do have their own research and evaluation group, and the manager of that used to work as the LPSCC's analyst. I think they have so much work they're doing on their own data that they haven't requested anything from us recently. We were working with them really closely when they did the original data migration to the warehouse from their old mainframe system from 1983. Then they moved to a new system and they had a tight deadline by which they had to go live, and they had so many technical challenges that we weren't able to reconnect them to the warehouse. We've also had a bit of a lull in the work because of the pandemic. The LPSCC has still been meeting regularly but the subcommittee for the data warehouse has not been meeting during the pandemic.
- Q: I was going to ask how often you meet with your colleagues. So, that's sort of dropped off during the pandemic?
- A: At the policy level it did. Our users at the research level, we meet once or twice a month and we have a lot of interaction throughout the month as well.
- Q: I was wondering about security breaches—has data ever been compromised? I know you said that people are deidentified, but have you even had security breaches?
- A: I think in our situation we have so few people accessing the data that there's not as much concern about that. But somebody was saying when we last met two or three weeks ago that the newspaper reported something that looked like it came from our data that shouldn't have been given to the paper. I don't think it was our researchers, but they might have shared it with someone in the group that somehow made mention of it. But that wasn't a data breach so much as a misrepresentation of aggregate data. So, it wasn't showing people's identifying information, but it was kind of a misrepresentation of what the data really maps. One big concern we have is with expunged data. If the judge says a person's record is clean, it needs to be deleted across systems and from the warehouse, and all of the agencies that are sharing data with us want to know that expunged data is being expunged. And expungements are successes so we're actually deleting successes and aren't able to analyze those.
- Q: Do you use two-factor identification?
- A: Not within our database. You do have to be logged into our network before you can access the database. There are also security issues with our reporting. We use Tableau in house, but our Tableau server doesn't meet the state's Criminal Justice Information Services (CJIS) standards—we've got a CJIS officer who says what standards we need to meet. And that Tableau database doesn't meet those standards, so we're not allowed to

use Tableau for sharing reports. I think we're starting that conversation about if we get the data fully deidentified and only use it for aggregate reports that maybe we can publish some of that, but that's been an issue for us. But, as far as two-factor identification, our users are analysts going straight into the system to run SQL or database queries. They have to be logged in to the network before they can get to the database and even then, we have an active directory that really limits who can access it to a very small number. I think we have fewer than 12 people that have access to the database.

- Q: You talked about issues having to be hashed out and figuring out how to address things like data security and other basic issues. Those are threshold questions, I imagine, or protocols that you had to set. Who was at the table in setting those?
- A: Those have evolved over time too. It seems like in the early days, that it was a lot looser. At the operations level, the middle managers, we worked on a policy manual for a couple of years that's still not quite signed. There's a group of about six or eight of us that have gotten together. That group includes an IT manager and the research and evaluation manager from the sheriff's office, the research and evaluation manager from the DA's office, the training supervisor from the court, and Abbey Stamp from the LPSCC. That's the group that is really hands on and hammers stuff like that out.
- Q: Did those people or anybody else do a data dictionary? For example, if the court has one definition of recidivism and the prosecutor's office has a different definition, how do those get reconciled in the warehouse?
- A: That's really tough, because everyone has different definitions and that's lacking. It's something that we've talked about quite a bit. Now, David Mitnick is starting to bring that together so we can get working on it again. I think the DA's dashboard has some really good definitions of data on that site and that's a really good model for what we'd want to do. For now, with the projects we do collaboratively across the groups, that's when that gets ironed out. That's where people are making sure that they're agreeing to what's been represented.
- Q: You've stewarded this from the beginning, but of course the world evolves, and you learn as you go, as with anything? What do you wish was built differently from the beginning, or what would you want another group to consider when starting a project like this?
- A: I think that there's always a lot of technical debt[?] that we're cleaning up a lot of that now. We had a contractor on board the last year and he's done a lot of cleanup to bring us to industry standards. Our deployment, code management, and deployment methodologies are up to industry standard now, instead of haphazard like we had before. I think one thing when you design a data warehouse is that there are almost two purposes. You've got ad hoc purposes where the data analysts don't want to wait for IT and they're good at getting data from anywhere and making something out of it. When you try to design a data warehouse with a star schema or something where you have analytical data models, that takes time the researchers don't want and they basically code the stuff they need and then you have to try to implement that in the database. So, you wind up creating a whole bunch of copies of code instead of a central database design. What I would like is to reduce our database design. Over the years we've had

different standards and things don't look the same. We have the booking number for the jail which is also called an MCL number—I'm not even sure what MCL means—and it's also the custody number, so that's all these different names for the same data element. I've tried to build a kind of semantic layer that would make all of that uniform, but it hasn't been finished. I guess what I'm trying to say is that you should build a lot more of that structure in place so that it supports a wider range of functions. The researchers don't want a black box, something that they don't know how it got there and so there's so many different varieties that they can get when they have the raw data. So, that's a challenge to get something that's reusable and also flexible enough that they get what they need. One other thing to consider is, whether you work from the top down or from the bottom up, where you have a couple of agencies sharing data, it's important that you get the buy-in from the top level so that everybody knows what's going on.

- Q: You said that, when you have special research projects then you hire a contractor to help with that. Where do the contractors come from? Who requests them? Is it to look at the system as a whole? What are some examples of what those special projects might be that might need somebody to step in for?
- A: Actually, that wouldn't be on the research side, those projects are more about technology and maintenance. When we rewrote the system, that was a two-year project and I think we had five people on board for that. And when the court changed their system, that was a big change and it took us quite a while to get their data back up. So, those are what those types of projects are about, when we have to rebuild all or part of the back-end database, getting the data coming in and then automating the load, which takes a lot of work.
- Q: Do you get data live from the agencies you work with or are you doing some cleaning?
- A: It's a daily load. We actually do get live data from the courts, so as soon as they save a record it gets sent, but we are loading it once a day. We're loading it to a staging database. Our flagship product is being able to match people across systems, so after the data gets loaded to the staging database, we then run our person-matching process and then load it to our reporting database for the analysts to access.
- Members thank Joe for his time and the information he provided. He leaves meeting.
- Models subgroup discusses potential follow up. Will compile follow-up questions to send to Abbey Stamp by email.
- Member who was unable to attend the interview the subgroup did with Katy Collins from Allegheny County asks how the information she provided compares with what we heard today.
- Other member explains that that project was driven by the health and human services side, where the analytic capacity was, and it was a story about an administrator who had a vision for the data warehouse and made sure that they had the resources to manage it. So, their data warehouse is really strong on health and human services data and then has some data on aspects of the criminal justice system, but not the prosecutor and only uneven police data. What's really interesting about Multnomah County is that 20 years ago they probably had full system-wide data from the criminal

justice system, and there aren't many places where that exists. It seems like their data warehouse is something that was probably built pretty well at the turn of the millennium and has gotten worse since then. Also wonders whether Joe was possibly underselling the political disagreement in losing access to Portland police data. In terms of getting agreement from system players to releasing their data, it doesn't seem like either place has figured that out long term. Also, in Allegheny, it seems like social workers are using their data warehouse to inform their work on a day-to-day basis, whereas Multnomah's warehouse is only used by a handful of analysts. Allegheny seems better in terms of cooperation, but Multnomah seems better on transparency.

- Another member says they disagree about transparency, that Allegheny's analytics unit has dashboards for both health and human services data and some criminal justice data. They also feel like it would be good to take the framework and the approach that Allegheny did, where they did things in the beginning like creating a data dictionary and figuring out things like how to reconcile different race/ethnicity data from different systems, and that that made for better data and then analysis looking at the system across the board. Also thinks it was interesting that in Allegheny the county executive or county commissioners were involved in the creation of the data warehouse or the impetus for it and didn't see that necessarily in Multnomah.
- Another member says that they liked how Allegheny had internal advocates to help explain their work, and how Allegheny's warehouse seemed like it was more accessible and was more of a systems approach.
- Meeting adjourns.