
Di [00:02:10]: We have talked a little bit about some of the business drivers and the business capabilities that information sharing provides. What we'd like to talk about now are the CONNECT project's impacts on actual users in the field and how you help them become engaged in the project, take advantage of those capabilities, and how did you help your users through those changes. First question, for Maury or for Mike or for John, at what point did you engage real-world users in the CONNECT project?

Maury: After Phase 1, each of our states, as part of the phase, opened up the search capabilities for the other states.

[00:03:03]

So there was a training curve that we needed to explore with our user base. We went through and showed them how to use this system that's different than the traditional way we were doing search just because of the results. Certainly for Alabama they returned differently than they would have in our traditional way of receiving data. But again, that was still more of the end of the pilot in terms of could we get this done, so we showed, yes, we could. Now, the user experience is better now. I think the most practical effect has been not in Alabama because certainly our state doesn't border the other three states, which Nebraska does. I think Mike has a much more practical experience there.

Mike: I'd say for us, though – again, a lot of it goes back to the whole notion that we all had our own portals. We all had our own data search capabilities.

[00:04:00]

So bringing the users on to the notion of searching against anything was fine. We had been letting people know that we were working on this, letting people know that we were trying to move to that, so they were kind of ready for it. We started out with a smaller user group, a lot of our state patrol and CID [Criminal Identification Division] folks. Then once we got it tested through and brought them live, we opened it up to a broader law enforcement community. They took to it fairly well. We have out in the panhandle of Nebraska – they're close to Wyoming, close to even Colorado, and then along the border we share with Kansas, we've got parolees, probationers, so people certainly wanted to look at those capabilities.

People took to it. The first thing with driver photos, identification – there were things that we didn't know about what we were going to be using. When we started throwing out data, at least for us, there are scenarios that we don't imagine what people are going to be using. They're talking about somebody coming across for a drug deal, and they have a name or they have a photo. We didn't know why they were trying to identify them, but we're hearing about these scenarios that made a lot of sense.

[00:05:00]

So bringing people under a standardized search mechanism that they were kind of used to, and with a standardized display, which is a good thing that the standards do as well. We're able to map the data and then display it in the same way. Part of the difference is translation of meaning and terminology across the states. Things don't mean necessarily exactly the same thing, so we have to be a little cautious with that, but I think it was fairly easy because, again, because of the technology, and people were used to that, so they get used to it. And then one of the comforting things for us, anyway, was if it went down or we were doing an upgrade, we started getting e-mails and phone calls asking, "Where is CONNECT?" That was a good sign

Di: That's a very good sign!

Mike: It was gratifying and good to know.

Maury: One of the approaches that Mike and I have taken in the way we display this data is we try to make it intuitive. You do not need a huge learning curve in terms of the user. It's pretty straightforward. You get results back from what you – generally the way you expected. There are tiny little things, like I think all four states displayed height differently for a person.

[00:06:05]

So we had to say, "Which one do we want to go with? That's the way we're going to display it," so that might be a little bit of a difference for the end user.

Di: I don't even know if I could come up with four different ways of displaying height. What are you saying? Like feet and inches, total inches?

Maury: Exactly. Some said, five feet, three inches. Others said 5'3". Then one said total inches, and that was it.

It was just a different way, so we're all sort of exploring these nuances. But ultimately, that's still intuitive to an officer or whoever is using the system can look at it and go, "I know what that means." That's the nice part about our portal technology though. It can take the data almost any way. We've translated it so matches up if it meets a certain definition, and then we throw it out in a common way.

[00:07:02]

Mike: But it's also changed and evolved as we got used to the data. We started with a more traditional approach: you start with a data set, and you list the 20 elements that you're going to share, and then you go from there. With LEXS [Logical Entity eXchange Specification] and a lot of the IEPDs [Information Exchange Package Documents], when we got to the next step, we realized we didn't need to all agree to share the same data. Or if somebody could serve up court data, that's fine. If somebody else couldn't, that's fine. And if somebody could only serve up 10 elements, that was okay. If somebody else had 120 elements out of the IEPD, we can serve them up in a standardized format and make that migration.

That's a real, real strength in being able to get at the data and get at the data that people are able to share, are willing to share, and they think is a need to share. People can conform to whatever they need to do. If they don't want to share a certain data because it's restricted, they don't have to.

Maury: Or if they're not technically capable yet. We bring it to them as, "Give us as much as you're comfortable with getting too." We put it on your sophistication level.

John: This is one of the things CONNECT was looking ahead for.

[00:08:00]

By having a single portal, the user only had to deal with one format of presentation. Whereas before, you'd have to go and look at the presentation of every hosted system, and then you're not really sure how to use it, etcetera. This particular design of providing a

single user experience, and then behind the scenes doing the queries to the various systems, it goes a long ways towards usability and a better user experience. Now, we took our criminal history system, and we don't have a portal of all the different resources.

Di: You're speaking now about Los Angeles County?

John: Los Angeles County. In our criminal history system, we have a version which is a light version that we came up with so that any user who comes that's not a member already of LA County and currently using the criminal history system, they just get like two functions – some identity and name searches, and a rap sheet return.

[00:09:06]

Whereas the users in LA County, when they come in, they have a number of six or seven different functions. Because we didn't want to have to deal with training people, we didn't want to have to deal with help desk calls and all of those things. So one approach is make one standardized interface and have multiple resources all displayed consistently. That reduces all those variabilities and help desks. Another is to take an existing application and kind of restrict the capabilities that you're offering to avoid a lot of confusion about that interface.

Maury: I will say that this was a big discussion early on because – and we've changed our mindset – because at first we all thought we were going to just absorb the data into our own portals, display it the way we always knew how and in the way that our customers were used to seeing.

[00:10:00]

It's technically capable and possible to do that, but you get more bang for the buck with the new approach we're doing, and I think we all came around to the idea that after seeing, after a year or so of us really going through the first phase and we evolved was this really is a better way to come across to present the information.

Di: More predictable, more scalable.

Maury: Exactly. If we had some specialized needs, since we are using the GRA [Global Reference Architecture] approach, I think we could

allow certain sets of data to rifle directly in, in a different interface if necessary, but I don't think it's necessary.

Mike: Well, I think it's good though to remind the user that it's not the same data source that they're used to. If you push it into say your own portal display, someone might assume that it's Nebraska data. I really like the notion of being able to label that data as being Alabama data or Kansas data.

John: Oh, there's source. You display where this data is sourced in your portal when you do that.

[00:11:01]

Maury: Absolutely.

Mike: And using standardized displays that are the query results coming back.

And instead of pushing them into the way I normally would display corrections data. When I show corrections data, it's going to be obvious it's coming from Alabama.

John: The difference is that when you're displaying data, you can do that in a consistent fashion. When the application that you want to go to is actually performing some function for you, then you really are going to go to that interface. For example, the N-DEx application – National Data Exchange application that's hosted by the FBI [Federal Bureau of Investigations] – they have a lot of link analysis and all kinds of software and tools that you want to have the ability – and we're getting the ability – to do that interface for our law enforcement when they want to use and take advantage of those tools. But we also have, just to do a check to see if N-DEx has any information, we use a little service interface just to get some key data elements that would say, "Yes, I found this name."

[00:12:03]

Then when you want to drill down, you go over to the N-DEx browser application because it's got a lot of features and functions that don't come with data. They're resident at the application. So you'll still have both scenarios. The one that's being described here is primarily data retrieved and displayed.

Di: Viewing only.

John: Yes. And not actively trying to drill down and find link analysis and all that sort of stuff.

Mike: I think the method you described, which is analogous to what we're implementing now, is great because it doesn't totally rely on an application. Somebody might not need to consume all of those N-DEX resources. But they do need to know if there's somebody there – or maybe three potential people there that they might want to look at. They can choose which one, and then get the power of N-DEX or whatever the application is.

John: Absolutely.

Mike: It centralizes a little bit of that up front work.

John: It's very flexible.

Mike: It's a really good approach.

[00:13:01]

Di: Help me understand again, at what point did you start engaging your user community? How are you managing users' requests for enhancements? I think there was some talk earlier about managing expectations. Once people are used to getting this kind of information, they want more. Can you speak to those ideas? Do you have a prioritization process that's part of your CONNECT governing body, for instance?

Mike: We've discussed a lot of enhancements and a lot of things that we want to do. In terms of our users, they bring ideas to us, things that we haven't talked about. We tell them what our priority data sets are, what we'd like to bring to the table, what we'd like to share with them, what we would like to share out of our data sets with our other users. To a great degree, it also goes back to financing and the funding that we have available and what we can do. But now for us, I think we're in the last stage. We've been able to implement things which should make adding other data sets easier in the future.

[00:14:03]

There was possibly an expectation that it was going to be very easy, very transparent. That's probably not going to be quite as

easy as we had hoped, but we're just going to have to keep moving and keep adding things. Users like to see anything, quite honestly, but expectations grow.

Maury: Certainly early, when we were assuming we were all trying to contribute the same data, so we're all – what drove us in those directions were what was the common data that we all had access to? What could we provide? Certainly adding the LEXS SR [LEXS Search and Retrieve] capabilities and the new way we present the data gives so much more flexibility and freedom. Now we can be driven more by local requests of users about what they might see, what they would be comfortable with, what they would like to see. We've got this brainstorm list of quite a bit of elements and data sources and direction of where we're continually seeing the future going.

[00:15:07]

We have shaped that up somewhat by bouncing it against the user base. If this was your possibilities – just because of financial obligations, we can't do everything – what would you like to see? I think our driving philosophy historically has been bang for the buck. We've got limited dollars. What can we do that will absolutely touch the broadest audience, even if it may not be quite as useful in some areas that we would like to see? But if it has a broader appeal, then that's where we decided to make the investment currently.

Di: How do you provide technical support to CONNECT users – if anyone has difficulty with the portal or their ID is not – they're being denied access?

[00:16:04]

What's your help desk structure like?

Maury: It falls back on the states themselves, on the partners. Hopefully for every state or partner that joins us, they have this capability built in for their existing applications. Again, it falls back too on the idea that we try to make it as intuitive as possible. We find that the search capabilities we have provided internally in the state or through CONNECT do not have that much help desk needs. They're pretty straightforward in terms of "I put in a name, I'm going to get a result." We've made it as easy as possible to do.

Di: So what I understand you saying is that Kansas supports Kansas users.

[00:17:02]

Mike supports Mike's users.

Mike: And the Kansas CONNECT point people or help desk people, if there's a problem, they can identify whether or not it's a Kansas problem or it's a CONNECT problem, or maybe it's just a Nebraska problem, and then get in touch with, say, me if it's a Nebraska problem, or we get in touch with our technical folks if it's the central portal or whatever we might happen to be using. The users, again, they're used to going to their primary website and clicking on something. They don't have a clue what may or may not be working, so they have to rely on the next level up for them.

Maury: We decided, I think, part of governance structure is at the partner level, at our level, we have designated points of contact between systems, between states. Say, for instance, a law enforcement officer runs a query and somehow they determine it is an error – that it is bad data. What do they do?

[00:18:01]

Well, generally, we have mechanisms within our own state for correcting issues like that. Certainly, I know we get calls as soon as there's something that comes back wrong, so we deal with it. What we've allowed is taking into account that this is a very likely possibility among our states, so we've designated points of contact for data issues between the states that only the partners themselves deal with.

Di: This probably goes without saying, but just to get it on the record, no one state has the ability to overwrite the other state's data, right?

Maury: No.

Di: So that's why you need a data quality reporting mechanism to just bring to somebody else's attention that they might have a flawed record?

Mike: Sure.

Maury: Sure.

Di: Okay. Now what about – the question that I have in my mind is whether you're actually replicating data.

[00:19:01]

So for instance, if I'm a Kansas user, and I run a search against Wyoming data and I get the results, am I actually downloading that Wyoming data in Kansas? Or am I just viewing Wyoming's data?

Mike: You're just viewing the data.

Di: Okay, so there's no problem with version control of data because you're not replicating data.

Mike: No.

Di: Okay. Very purposeful, right?

Mike: That's different from a centralized data warehouse, which gets incredibly complex. That's what N-DEx is actually doing. Replicating and maintaining that stuff, even at a state level or within a county actually, is incredibly difficult and a resource drag. So we basically just go out with kind of an index where we reach out to the states, find out the results, typically an intermediate result, and then go ahead and drill down and display that data back to the user. But, again, it gets back to who is the custodian of that data, who maintains that data and just shares it with a valid and trusted user.

[00:20:03]

Maury: There may be some legal waters to test here, but what we've said in our bylaws is this view is the law wrapping around how the dissemination is possible is the originator of the data. So if Alabama looks at a Kansas record, that officer is held to the standard of whatever Kansas allows for the dissemination, not what Alabama law says.

The idea is some other states have very open records laws. And that was another reason there's no replication of data locally, because we don't want there to be any question about the records

being transferred to a custodianship of another state. They're not. They get a view, a window into whatever that state has.

[00:21:01]

That way, we believe that it is legally binding to maintain the rules and regulations from that originating state.

Di: You've also avoided problems with prohibitions on re-sharing, right? Once you've shared it with Kansas, then what is Kansas doing with it? You've avoided all of those policy issues.

Maury: And one of the standing principles currently the way we've set this up is our baseline is going to be the FBI's CJIS [the Federal Bureau of Investigation's Criminal Justice Information Services] security policy rule. That makes it fairly straightforward in terms of what can be disseminated and what can't, the rules for what a state is going to be held accountable to if they do disseminate a record from one of the other state's partners. Fortunately, that has already been tested and proven at the national level in other information sharing systems. So we agreed and are binding each other to that standard.

[00:22:02]

John: It's not to say that the collection of data from multiple systems isn't part of what is supportable within this architecture, because with the rise of fusion centers throughout the country – I think there's like 72 of them now – most of the data content that they get is data based on queries or actual publishing of information to them with the appropriate security mechanisms. Law enforcement will send incident reports to the local fusion center, and they'll have access to photos and all these other things, and they bring those things together and have their own link analysis tools to solve local crime.

There is a place for view only for that single transaction versus bringing all this information together for some investigation and analysis.

[00:23:07]

Then you get into all the rules about how long can you keep investigative data. As long as it's an active investigation, you can keep it for five years, is one of the rules I've heard. Then after

that, you have to purge it. All of those are just another use case beyond what's being described with CONNECT that still is applicable that you want to know the data that's sent to you, you know who sent it, then you want the credential on that data to know that it came from a reliable source. Then when you share it with your fusion center folks or with other agencies – that's secondary dissemination, basically – you know who they are and what agencies they're representing and what their roles are.

So the model doesn't change based on the – the security model doesn't necessarily change based on whether you're aggregating data or you're doing transactional queries or you're just a recipient of a notice.

[00:24:04]

An Amber Alert in California that some child's been abducted, that's kind of broadcast to a lot of different systems, and they all get that. But still, you want to know that that's a valid broadcast and where it came from, and you need the credential that goes with that. I would just add that to the discussion.

Di: That's a good point, John, that these standards that we're talking about today, federated identity, technical privacy, they are adaptable to a wide variety of business needs.

John: Right, and interactions between different information sharing purposes and models across jurisdictional boundaries.

Di: John, did you have anything to add on data quality issues? Have you seen some best practices out there about how suspected flawed records are reported?

[00:25:01]

John: I can only just share with you that one of the things that happens with our local criminal history system is because the data is coming from the state criminal history as well as the court records and the arrest records, as that comes together, we do regular analysis for looking at data quality issues. The fact that you bring it together actually improves data quality in terms of discovery of where you didn't think you had a data quality problem, and you really do. We actually use that system as feedback to some of the source systems when we see that there's some data that just isn't quite matching up the way it should.

Di: Maury, you sounded like you had some recognition that information sharing has a way of surfacing data quality issues?

Maury: Well, yes, and that transcends the CONNECT project.

Di: Yes, it does. It's like magic.

[00:26:02]

Maury: That's the 800-pound gorilla in the room to me for everything related to justice sharing is the data quality. The solution – I really believe the solution is the Global tool set. If we just keep on pushing the whole community, I mean across the board, toward adopting these standards, we are going to – I do see the horizon when data quality issues are going to be much less significant than they are today, but today it is a big problem. Because the silos do exist in the states, in the counties, in the cities, even in the feds, where they don't talk to each other when we're building our data sets. So even if we share them, that's when it becomes evident – wow, this one doesn't match with this one, and they should.

John: For example, we have five different formats in our county for charges.

[00:27:03]

Maury: I believe it.

John: And it's five different systems, and some of them are the sheriffs' systems, some are the courts, and then we'd have ours, and then the state. And when you try to bring all these charge codes together to make sure we're all talking about the same charge, you'd be amazed at the diversity of just how they record a charge.

Maury: Well, then the disposition of the charge. Even after that, even if you were to solve the charges, then applying the dispositions, the way the different courts at different levels might review that is just a big problem.

John: So by using a standard vocabulary and a standard breakout of how you represent a charge, it leads to everybody kind of mapping their local format into that standard format so that you can enable information sharing that makes sense to the other party when they get it.

[00:27:58]

Mike: And not just queries. I think it also helps in terms of workflow. I think people need to agree that they're going to benefit by automating the process. Maybe moving filing from the prosecutor to the court, it needs to be in a standardized way so you have a natural way to enforce people adopting the standards. You don't have to get them to agree for some nebulous reason. It has to be that way at a technical reason, then you get the benefit, so it helps.

Maury: And ultimately a positive benefit of having search done so well through the portals that we've described is now it makes it more evident. Policy makers can see, "Wow, we certainly do have a need for this," whereas if you look at your own data all the time, you don't necessarily see that there's a problem.

John: That's right.

Di: That's very interesting. Have there been any very difficult – they're not data quality issues, exactly, but they're just data varieties, like different ways of marking up data.

[00:29:02]

Did those cause you difficulties at the beginning of the CONNECT project, and how did you overcome some of the lingo that you were talking about earlier, like with height? What processes have been particularly effective in reaching a consensus about what the right one way is going to be?

Maury: When you're talking about the specific data sets, I think first we looked to see if any national standard had been created related to the data type, whether it's corrections data and looking at the warehouse at NIEM [National Information Exchange Model] to see if an IEPD had been created or if we had to create one from scratch, which we did for the driver's license data. The CONNECT project put together three different IEPDs related to the way you query driver's license and search results return. And that's a very small data set, by the way, your driver's license. That was an exercise in and of itself to learn how to do what you just suggested.

[00:30:02]

Most of it is just sitting around rolling our sleeves up around the table, the practitioners' side of the house as well as the technical implementers, and just finally coming to an agreement. It takes some give and take, people willing to say, "It doesn't have to be exactly like I've always done it," in the room. Ultimately, I think it worked very, very well with what we did come up with.

Di: Mike, do you remember those days, the data harmonization process?

Mike: They aren't that far behind us. They're still with us.

Di: Maybe they're perennial.

John: They're in everybody's future, still.

Maury: Yeah, they still are.

Mike: The good thing about having standards here for identifying an IEPD or a style sheet for how you display the data is that you know what the end result is going to be. You know how you're going to transfer the data. You know what it's going to look like. All the work comes on the front end, as Maury said. Doing that mapping, no matter what new data set you come into, you've got to do that mapping.

[00:31:00]

That's where Global is really, really valuable in having that data set, having those data elements, having mappings, having translations because once you do that once, once you do that that initial time, you can reuse those mappings for other transactions, for other types of queries, related data sets. That's a real benefit. But that first time through is incredibly difficult. We worked with our corrections folks last year to do that, and then they're working on some other stuff now. It's "roll up the sleeves" time and just try to grind through it.

Maury: And you need some expertise, some competency there. Again, I touched base on that earlier, it's a learning curve for so much of our staff that's never done this. There's a patience that needs to be involved with the leadership, a patience that needs to be involved at every level, but an encouragement and to really make sure everybody's on board and knows the end goal, and that's what

we've got to achieve. Yeah, after the first time, it was just like one of those "woo" moments, "We got there."

[00:32:03]

Di: Mike, one of the things I think I might have heard you imply is that it you plan your documentation – for instance, your mapping documentation – if you plan that ahead of time with an expectation that there are portions of that that will be reused, that that would be a best practice, because it *will* be reused.

Mike: Absolutely. Yeah, and you *want* it to be reused because you don't want to redefine it and, even internally to your own state, define it in two different ways for two different applications. You want to be able to reuse that. As we see staffing turnovers, we see the technology definitions change. Having that kind of consistency is essential, and it is going to help us on the cost curve and everything else. Yeah, I absolutely agree. Making the assumption to do that is also related back to what people were talking about before. People want to do things the same way they've always done it. Going to a NIEM mapping or mapping all of your data elements is new for people.

[00:33:00]

And it's time consuming and costly, and it's a lot of overhead, so it's difficult, but if people can agree that there's going to be that reuse and that benefit, I think it's very, very useful.

John: And information exchange specification – and all programming technical folks understand writing a spec, what that is, but this is really an information exchange specification, which is different than writing a spec for a particular function in an application. The Global Reference Architecture, it really gives you a template for writing an information exchange service that includes the data elements and why you're doing it and what's the workflow that's involved. But ultimately, you've got to do the specification and the planning, and you've got to get it at the right level of data exchange so that it is reusable. If you make it too small, it's not going to get all that reuse.

[00:34:00]

If you make it too big, it's too brittle. There's a little bit of art to finding what the right chunk of information is to model.

[00:34:09]