NNIPCamp Columbus, June 19, 2013

**Session 3: Integrated Data Systems (IDS)**

Led by Leah Hendey, Urban Institute

Notes by Rob Pitingolo

Intro: there are a lot of different types of data to integrate (housing, services, crime) can be at county, city or other levels. Used for policy making, service targeting, other purposes. We think of IDS as a database of combined integrated data. (Hendey)

We released a request for statements of interest for the cross-site project. Info is on the NNIP website and Friday I will go over that project more specifically. Kimberly Spring will talk about why the Annie E. Casey Foundation wants to fund IDS. For that reason we don’t need to talk about that now. (Hendey)

How are people using IDS? What challenges exist? We’ve been connecting with some groups already to provide TA for IDS (Hendey).

How many people currently involved in IDS? (Hendey)

* (About half of people raise hand)

How many people don’t think about IDS (Hendey)

* (Nobody raises hand)

Any intervention funded by the county will be involved in the IDS. One thing we suggested to get from it is a characteristic of the Shelby County population. People generalize from people in poverty nationally, but poor people in our area may be different. We hooked up with the state to do integrate with state agencies having to do with kids. (Betts, Memphis)

Grand Rapids involved with school district to be the hub linking program level data with school district data. InFocus is the vendor for the data. Share back data from the school district from the non-profits. We do daily transfer of raw data and quarterly transfer to outcome data. Another IDS links juvenile delinquency data to program data. Hope to have pilot completed by end of this year. What we learned from the first project makes us think maintaining record level data is expensive so we want some kind of ‘translator’ to capture data from any format into the system. We will meet people where they are with their systems rather than force them to use one software. (Gustavo Rotondero)

Allegany Department of human services is known nationally for IDS work. We have a ton of school districts in the county and they are working with about 6. We are doing the community profiles like providence plan and using their IDS. Combines what we have now with human services to create indicators. A lot of it will be with school district data. It’s an exciting and ambitious project. (Sabina Deitrick, Pittsburgh)

We have a data warehouse in Pinellas County and have an exemption from public record law in FL. This makes us a natural place to do IDS. We have a data system for what we fund, plus data agreements. School district sends of school records. We get child welfare data, court data, some health data. Health data has to be aggregated before it left the health dept but that was a one-off. Other data comes on regular intervals and is uploaded to the data warehouse. Some comes in monthly or quarterly. We use it for longitudinal, like whether children stay out of the justice system post-program. We also use it for planning purposes county-wide. We work with school district and county to provide that info. (Denise Grosbeck, Pinellas County)

We’ve got about 9 unfunded efforts. Partnering with a charter school to develop the IDS and some logic layers to provide some guidance. We’d like to use real-time data, like how did the kid do on their test today, not months later after they get their report card. Want to not just share the data but build programming suggestions so that the time becomes more valuable. A group in Dallas is launching a volunteer management system. You tell it what kinds of volunteers you need, details about it ,etc. Kind of like facebook for volunteers. Can use this system to real-time tweak the volunteer demand. A lot of people utilize data systems but don’t share their data. Lots of orgs buy ETO but don’t share their data with each other. Fear is that they didn’t ask the same things the same ways so after-the fact the data won’t merge. Down the road orgs will be forced to share to fundraise, we need to get ahead of that. (Bray, Dallas)

What Shelby county did was make the orgs be part of the process. We developed indicators like hardship index and everybody has to do the same indicators to come into the system. So that gets at the problem that Tim Bray just mentioned. (Betts)

Orgs buy these systems to do their own performance measurement. All it does in some case is take crappy forms and make them electronic. Some of the vendors just sell the customers what they want, not what’s best for the greater good. Too many orgs have good systems that don’t track stuff. (Bray)

San Antonio has a system to integrate person data and place data for early warning. Figure out who needs services and also how they’re doing. As many indicators as make sense will be open to the public – transparency is important. Right now there’s no data entry screen, each provider has their own system and we don’t want to make them switch, so we just plug into theirs. We make it an indicator with a red/yellow/green indicator for easy tracking. There’s not just an indicator but a trigger to an event (kid missed so many days of school triggers a home visit). One school district and a few other orgs are contracted under promise neighborhoods. There’s also Choice that we’re trying to loop in. 45-50 agencies will eventually be part of this system. (McKieren, San Antonio)

Who is supporting all of this (Harkness, MacArthur Foundation)?

* We originally had funders, but then Promise and Choice came along and all kinds of other funders started putting money into the pot. Now there are sophisticated requirements. (McKieren)

There’s a person viewer so that info about every child can theoretically be seen. Different people can see different things based on level of permission allowed. (McKieren)

Social workers are the ones who have code of ethics not to share client data (Betts). It seems to me like when I go to the doctor I know that insurance people see my records, I know the specialist sees it. HIPAA doesn’t seem like much of a roadblock. I don’t want to be down on social workers, but they distinguish themselves as the protector of client data.

We wrote a big report about regulations that affect datasets. I think it is complicated when you are trying to share across sectors at individual client level. The permission management layer is definitely important. Ideally it can be changed as rules change. (Coulton, Cleveland)

Our most difficult group has been the school district. (Grosbeck)

Research is really needed now on what the influence of observing this line of things about an individual on the providers that they’re interacting with. If teachers know that the child’s mother has problems, they might tolerate poor performance. On the other hand, certain information could boost provider’s capability. It’s probably proven that specialists need primary doctor’s records. Not proven for case management. (Coulton)

I don’t think the social workers believe that sharing the data can work. (Betts)

Doctors all do (mostly) the same thing, but teachers and social workers don’t – it’s harder for them to talk to each other. Providers risk providing a bad service because they don’t understand jargon or communicate with each other. We use IQ as a data point, but it’s becoming clearer that it doesn’t mean you’re smart or dumb; not everyone yet believes that. (Bray)

We discovered things like days absence not important but chronic absenteeism matters. Maybe it doesn’t matter if a kid’s parent has health problems but it is if their parent dies. Who benefits from knowing that a child is from a military family? Not everybody needs to know that. (McKieren)

In the health arena, some things aren’t stigmatizing, but a kid in HS in a foster home may impact how people react. We have to be sensitive. (Grosbeck)

Internally we created a risk score from the various datasets that we have. Kids have been in trouble with police, in school, etc. but decided not to use it because of some ethical issues. (Grosbeck)

In WA they have a state-wide IDS and modeled what level of case management they needed to provide. If they could predict if the client needed hand-holding or not, they could tailor it to them. (Hendey)

That is a layer that is missing; the business processing layer (Bray). The risk assessment isn’t correlated with the services they received when you look back on it. We have to both have a process and follow that process. We don’t know someone is going to take a number and do the right thing with it.

Most models have huge MOE of prediction at individual level. (Coulton)

Maybe you learn that people who get lots of services don’t get them and vice versa. We are trying to figure out who people are that fall through the cracks. Need to share data back and forth to answer these questions. (Betts)

I think we assume doctors do great things by sharing data, I don’t think it’s definitely true. We are overestimating and what do we know about doctors anyway? (Betts)

User consent trumps in the system. It’s actually very complicated. User permissions allow you to see person level data or aggregate. Then there is a parental consent process about seeing the data and sending it around. (McKieren)

We have hospitals trying to do work with the health exchange and 3 years later I haven’t seen much of anything it’s all very complicated it seems. (Grosbeck)

We have an ETO based system and doing a pilot targeting the families with most complex needs. First phase is getting community outreach role organized. We have a data bridge to the school system. Next phase is extending that to mentoring programs. We’re adding more complexity – combining city and county. The 3rd level will be the out of schooltime programs. Using national guidance to create a true ‘network’ and the data model is yet to come. Once we do the systems work the data model seems straightforward. In the middle of that is the parent permission problem. (Millea, Austin)

We have to document the relationship of people to each other in the database. (McKieren)

Getting started is hard because we get data from different places in native format. Example, no shared id numbers so we have to match on name which is a pain. (Grosbeck)

The problem with matching is that you don’t know what you’re missing. You have to set a threshold of matching and that threshold has to be arbitrary. Over the years you get better likelihood of matching because there’s more evidence about the person. What you don’t know is whether there are children who should have matched but didn’t because they have so many changes or they don’t have enough records in the data. If you can’t find matches, you have to start making assumptions. (Coulton)

Do the agencies help with the matches?

I am working with agency data that they collect and then supply, so they are never involved in the matching process. They get reports back about what the data says. We work with specific projects like Promise and Choice Neighborhoods where we are setting up new systems. Add to that the state data and it becomes an even bigger mish mosh. (Coulton)

In Grand Rapids we are moving toward the San Antonio type of development. It’s too much to ask people to adopt a new system or change what they are collecting. We need to think about mixing and matching complex data. Is it more effective to have daily reports or quarterly? There’s a cost difference. Some people can become overwhelms. (Rotondero)

Some states are motivated to do this to catch cheaters (Betts).

Yes, that’s how they justify funding (Hendey).

Is anyone asked to do IDS for pay for performance or something (Harkness)?

United way wants this instead of asking grantees to self-report (McKieren).

We work with a tech company that got into it because of pay for performance. They’ve gotten involved in using the system for documenting pay for performance. (Bray)

There’s a difference between pay for performance and catching cheaters. (Betts)

We are doing it for pay for success investment fund. Promises to save the government money, if they do well then they make a profit. You have to determine how much money can actually be saved. You have to base it on past evidence. If you intervene in something, how much can be saved for homeless services, jail, etc. the burden is having to be really accurate because someone is going to lose money if the accuracy is bad. After we prep and tabulate the data, then it’s fed into a financial model. (Coulton)

All of these things hinge on wide margin of error on the cost estimate. People with money don’t want to lose it and want to know you’re really accurate. (Bray)

We could do some webinars on this, especially what San Antonio has been doing on the consent process. How you do consent determines a lot about the kind of data you get. (Hendey)

Other webinar interest in probabilistic matching.