Session 1: Thursday, September 14th 4:15-5:15pm

Location: Ambassador II

Session Title: Data management/Open data

Organizer: Greg Jordan-Detamore

Primary Notetaker: Katya Abazajian

Attendees: Nancy Jones, Randy Rosso, Eamon Johnson, Matt Richardson, Ali Carella, Greg Jordan-Detamore, Cherie Chung

* Introductions
* Nancy Jones: Challenge is data management, BNIA is not the main user of Open Baltimore’s open data
* Matt: Interested in linking community open data portal with other datasets and portals, so want to formalize metadata standards; want to apply that as a community database as a communal effort uniting with other cities and regions while establishing best practice
* Ali: Working on getting leaders to come up with measurable indicators
* Eamon: Poverty Center has an open data portal since 1985, but don’t use recent technology. Have experience but haven’t seen a great open data portal yet. Common are CCAN, Accela, DCAN, Esri, Socrata
* Cherie: Digital C trying to create an open data portal for Northeast Ohio
* Greg: working across jurisdictions has issues with formatting where different precincts work in different file formats, etc.Example of common standards is of course the Census Bureau. Transit data with GTFS standard is now the defacto global standard which is why you can go on Google Maps/other for directions
* Matt: Anytime anyone attempts to do that in Milwaukee, the city, district, county, and school all count attendance differently. Getting them to agree on that is the focus of a whole working group. Why can’t there be an API connected to a matrix with an intermediary that can submit their own data.
* Eamon: Processes create data, but there can be totally different reasons for those processes being different. There has to be some target schema that everyone else’s data is fitting into.
* Nancy: In Maryland, enrollment is standardized, the number of children at that school. The end of the year is the day after school ends in every school no matter when it is. But the number of enrolled vs. attended at the end are different, so there are different fields. Then put together a “switchers” indicator.
* Randy: Issue with truancy definitions, absences can confuse the results.
* Nancy: We have inputs and outputs and Socrata does the outputs well - you can get a variety of formats. But what people are putting in is a mess. If you want to instill standards across the country, we can’t even get the same fields from the same department year after year.
* Matt: County, city, district everyone tracks gender differently. People have different standards for asking about gender and different reasons, so that’s not even the same. There have been established standards that have gone out effectively, and need somebody really safe and big has to start using it and then get critical mass going. And those are fairly simple data sets. But individual, programmatic, education data is extremely complex. Help people transition to match the matching schema that maps various definitions to other definitions.
* Spike: The problem is who pitches this change to who? It’s a standard if the world uses it, but it’s an idea or specification if it’s brand new. There are good transition tools like Open Referral, the Human Services Data Specification. Not enough work in actually promoting the standard.
* Eamon: Is there any relationship to Medical interoperability standards?
* Spike: Those are driven by big system players who need to do it and then get big companies to enforce the standard/built it. Like the ICD codes, which you have to pay to access the back end. They can afford to institute these things.
* Nancy: Good example of data that is same over time?
* Greg: Police/crime logs would be a first across region.
* Nancy: The police data field changes every year and it goes through Esri at some point because the variable names are cut off at 10 characters every time, so who knows.
* Eamon: The question is what is the role and responsibility of the regional intermediary - who would do that? Who would do the standardization work. Someone needs to force the vendor.
* Greg: If you have a standard that’s documented, that’s documentation that the people entering it don’t have to write. Value add.
* Spike: It seems like it’s hard to push metadata where there’s not open data. When it becomes open to the public, it becomes transformative when it is standardized. Where could the standards come from - the people using the open data could be the large force.
* Nancy: I’m responding to the ad hoc data requests, because the open data portal is the focus. Can still focus on standardizing the portal.
* Matt: There’s process to put the data in and take it out, and helping people through that may help.
* Eamon: Getting people to comply with a standard means that you can help people through that. They may not see the benefit of the standard police law. If the benefit is sporadic maybe it’s hard.
* Spike: CHange from within is faster than 1000 people pushing against it, 10 internal and it’s done. There’s no internal leadership because there’s no one to hold them accountable.
* Ali: Our push to get leadership buying internally has been a push for efficiency and to get data leaders. All of our systems are data-in but we could be using it to inform our practice and our grantwriting. The idea is we could be saving time if we had less things to report out, and we should have a leader in the organization who is the data person. If we could internally train one person.
* Matt: In some scenarios, can people even collect the right information?
* Ali: Service provider working with partners who all have different funders and standards, so hard to standardize. Standardization isn’t possible, so we’re looking for translatability, a measurement indicator tool, like well-being. Then choose indicators at each site, adn then choose tools at each site.
* Spike: Department of Labor at the state level is starting to develop the idea that reporting into five systems is maybe hard and that people could report in more consistent formats that would be less burden on the reporters. Lots of funding streams necessary. DOE, Byrne, Promise. Locally its vendors who determine how you default. Marketing reach works which vendors have.
* Matt: Convening work is expensive so the collaborative effort of creating a standard… Getting even internal staff to adopt a standards is hard. But if fundees are driving it and there’s effort over time. People want interoperability but it doesn’t work. If you made the standards a requirement of your procurement process for all your vendors, you could create.
* Katya; Promise top down indicators is hard because the programming doesn’t fit the federal standards.
* Spike: You have to accept some fuzz in the data. That happens at the local level too because a third of providers may collect homelessness in detail, or a third just track the number homelessness just yes or no, and the definitions themselves are different. There’s no funding to make them report it or collect it differently. They have to be OK with that fuzz.
* Ali: In Chicago, that definition of homeless vs. chronically homeless, if someone is housed for one night, they are no longer eligible for a voucher. So people don’t come in for services because they want to stay “homeless” and be in the voucher.
* Matt: I wonder if we could get out of the idea that there will never be any standardized data. SKOS is a schema that says that your data and my data are linked, but at a certain level, so it incorporates the fuzziness. So you can link fields, but it may not be direct. Librarians in Milwaukee are super interested, but there’s no funding, and then convening the people to do that is too expensive.
* Spike: There’s huge power in it. The 211 data people in Oakland are intrigued in it. We want to get them to just publish the data, but ideally they would be linking data. One day, linked data. Need a different publishing mechanism.
* Matt: Someone needs to manage the matrix.
* Spike: The APN has a known web address for a parcel number so the url is exactly the same so you can use them the same everywhere. You replace what you’re using in your system with the reference.
* Eamon: So how do you represent fuzziness.
* Matt: In the schema, it’s defined as the relative certainty.
* Matt: Using Marquette’s metaschema, layering CCAN (PostGres), RDS, SKOS could turn it on from a starting place. When getting data into the portal, not always populating SKOS. Got a vendor where in his proposals, puts open data in the line item. Put the schema option in the line item on a piece of software (management software). Affecting procurement change would be a long-term benefit as bad systems phase out. Could do that at an immediate level at the foundation or human services level. That all data systems should have schema.
* Eamon: Need an influencer to include these kinds of line items.
* Spike: Nonprofit sector has very little accountability.
* Ali: Part of that is that the populations we serve are so vulnerable. There’s an element of exploiting stories.
* Spike: At the simple level, how many GEDs are getting county level, adn that’s not sensitive but no one is collecting or publishing data. Without the simple level stuff you can’t dig down. We’re not giving people jobs and GEDs, so who actually is. Published data and reports can say “this is what we’re doing” so what’s the workforce money giving you - look at organizations that are working through this data and look at how much budget.
* Eamon: Coding bootcamps for example may be open to publishing data and standardizing.
* Spike: Pushing partners to publish dashboard of all different organizations data streams. Because what you achieve is public you can have more accountability. But would need to standardize across those partners. Look at the weakest link. In the benefit of the funders.
* **WHY don’t we have a guide for how funders should use NNIP partners. Data standards help you do your work and reduce the technical costs, will help measure results in faster time.** Standardization could start with a foundation. Can’t pick the fields, need to just pick standardization. People can’t agree on fields. Need an interoperability standard.