**Thursday, October 27th (2:15 PM ET)**

Session 2 – Open Data Platforms

Notes by Donovan Harvey

Attendees: Hayley Raetz (NYU Furman), Joe Andre (Case Western), Steve Saylor (WPRDC), Amy Nelson (AISP), Ethan Hug (MORPC), Tim Stallman (DataWorks), Jay Dev (ESRI), Annie Calef (Boston Indicators), Cheryl Knott (BNIA), Jeremy Pyne (Community Information Now), Jeff Williams (Grand Rapids), Ross Riley (WPRDC), Jordan Grace (Data-Driven Detroit), Jenna Losh (Data Center of New Orleans)

*Why are we here?*

Hayley Raetz; Furman: Always thinking about new opportunities for data

Joe Andre, Case Western: Dealing with data has been a struggle, especially watching people communicate data with high margin of error.

Steve Saylor, WPRDC: Use C-CAN, curious about

Amy Nelson, AISP: Thinks a lot about data access, intrigued about what is open and what shouldn’t be open

Ethan Hug, MORPC: Want to increase the amount of publicly data they produce

UNC Charlotte: Open data makes analysis sharing easier; learn from others

Littlefield, MKE: No open data platform, want to level up their interactive tools

TIm Stallman, DataWorks: Maintain Neighborhood Compass, between portal and compass. Curious about what works

Jay Dev, ESRI: Work for [ ]-HUD team

Annie Calef, Boston Indicators: No open data tool, hoping to include an open data component in Racial Wealth Gap research

Cheryl, BNIA: User and creators of open data. Over 1000 different items on ARCGIS account. Appointed member of MD Council of Open Data, workgroup for the state.

Jeremy Pyne, Community Information Now: User of lots of open data portals - recently launched a data tool for Bexar County. Had someone ask if they would set up an API

Andrew Carr, DataHaven - Always curious about data sources

Jeff Williams - Published data viz tool on top of existing data portal. Want to learn about scaling and APIs

Ross Riley, WPRDC - Use C-CAN,

Jordan Grace, Data-Driven Detroit: Has open data portal, curious about metadata, maintenance

Jenna Losh - Was using Socrata and realized it was a poor tool for the kind of work they were doing.

*Span of Sophistication in the room?*

Jordan: Trying to determine how relevant open-data portal was to users. Deciding between maintaining portal and focusing on tools with clearer use cases.

Steve: C-CAN, big selling point was that it was open source and community maintained. Building plug-ins was not a far reach. Alot of their tools use the ODPs API. Have direct-res access

Jeff Williams: Use SAS (soon R) to pre-process data before loading it into the table. Database queries to access information, no calculations are happening in the platform proper. Pre-processing data makes error analysis easier. Download states or country at a time and then run scripts on it → faster than querying city by city. Less marginal effort to load the whole nation instead of just Michigan. CDC, SVI

Joe Andre: Using legacy platform (since ‘88?)

*What are the use cases that you have in mind for your ODP? Initial intent vs. where you’re at now?*

Viktor, MKE: No ODP, some MKE orgs have data but don’t know how to present data or make it user friendly. In midst of strategic planning process → focus groups surfaced that there was desire for MKE NNIP to make data accessible, user-friendly. What would it look like to have data portraits visualized. Value in being part of a network, no need to recreate the wheel. MKE big fan of Data-driven Detroit (DDD).

For group: *do you collect identified data? if so, how do you deidentify and secure it?* Currently conducting cool project on housing but are collecting lots of sensitive information

Cheryl: BNIA was started 20 years ago, data outputs were put in a physical black book - occasionally PDFs on website. 130 indicators that are continuously updated in Excel - but sometimes in formats that aren’t accessible. One day, Cheryl just put data tables on the BNIA website. Baltimore noticed and asked for the data too → City was in the process of creating Socrata-based ODP (OpenBaltimore). Challenge was that BNIA did not have access to data and had to submit requests. for corrections to the city. City eventually transitioned from Socrata to ArcGIS portal, BNIA took initiative to … [not captured]

If they (the communities) want that neighborhood lens, they come to BNIA because they have the track record with it. Beginning to create datasets that can be used for indicators e.g. community-managed open spaces in Baltimore. Now working with ESRI.

Jeff: Similar experience to BNIA. Filtered data published with an option for users to download the raw data. Thought they would sell 4-5 dashboards/yr. Sell 1-2 but do 20-30 custom data requests. Data requests are now much faster.

Using open source software → generates

Call they get the most is about the neighborhood-level fact sheets. People are more interested in neighborhood information than tract-based.

Jay: Trends that I see from the field. I work on the tool people publish portals on. Multiple doors in → “not looking for a data portal, looking for housing data in my city.” Data portals are important as a one-stop shop but focus is on creating multiple doors in. We have all these things that are open data portals except in language and interface. Local partners can be very valuable for providing context. Value in having a few focus areas (or silos) so that local partners don’t have to sift through entire portals. Lots of microsites. Other NNIP partner mentioned that they started tacking on sites to budget proposals.

Cheryl: Look at how much staff time is required for a project ask what percent of annual budget is coming from this project. then pro-rate cost of ESRI portal based on that percentage and add it to budget. Way of offsetting costs so that clients get interactive sites. Recent shift to including online data presentation included in contract language.

CWRDN: User base is almost exclusively institutional employees.

Cheryl: Received funding from NSF and South Bay Data Hub. Hosted a series of webinars → convened online zoom session on BNIA YT channel where folks talked about how they set up various data portals. Included ArcGIS, C-CAN and HubZero. HubZero requires most set-up but very powerful.

Tim Stallmann: Public tools and internal database (PostGIS). Tech skill is higher to work with internal database (SQL coding), but substantive time savings. Pull data out of it and post CSV files for public-facing open data portals. Mostly work with static site. AWS offers a good non-profit credit - so hosted there.

Steve: Operate on the same level for their C-CAN database.

Jeremy Pyne: No database people - set up a ( ).

Jeff: Decided against using Python because city tends to change datasets in small ways. Prefer to use R so they can test breakdowns quickly.

Jenna: ODP process has been a whole revisioning of data infrastructure. Google Drive → One Drive → MS Azure. *How are you all storing data. What works?*

CWRDN: Secure cloud environment

Jeff: University Center as dev and test environment, shifting things to the cloud to avoid enterprise standards. Migrating production environment out and including it in contract language

BNIA: Have to go to physical office to access secure data, minimal IT support from university. Not yet feasible to do anything on the cloud because they would need to fundraise for it

Jeremy: Go through AWS to store things - non-profit credit covers the majority of the costs but may

Andrew Carr: *Do you all go through IRB approval? If not, what are the ethical and legal considerations*

Jenna: Most partners do not collect primary data. One project with PII had to go through

Amy: Core focus of AISP. Mainly work with govt agencies. Two NNIP partners that routinely collect govt data and communicate it at the neighborhood level. IRB is only necessary for "Research,” most public information. IRB approval is very expensive especially outside of institutional settings, independent IRB can be 5-10k. Prefers very clear data governance and stewardship. Highly encourages the creation of a table to decide how data is released. Laws also vary

Steve: Want to go as granular as possible but sometimes that risks privacy concerns. Begun looking into synthetic data

CWRDN: Work with quasi-public data, available upon request from gov’t.