NNIPCamp Columbus, June 19, 2013

Shared Data Infrastructure/Disseminating Data

Led by Eric Busboom, San Diego

Notes by Brianna Losoya

Intro: What kinds of things can we use nationally to promote data sharing, and how do we do this on specific realms. Different from IDS because it’s more focused.

* Data dissemination
* Developer Hacker Friendliness/Consumability
* Data Standards, for data and open systems standards

How are people currently sharing data? What would you like to improve?

* In Detroit, as open data has grown we’re trying to figure out how to expand, because we’ve previously taken individual requests. Trying to move in API direction.
* Putting data out in API format, but after spending so much time on it, are you shooting yourself in the foot? Then people are just taking it and not giving you money or credit to maintain your operations.
  + That’s the nice thing about API
  + It’s hard to get paid for that hard work that nobody can see, data development.
  + Has this happened yet, or are we worried about it happening?
    - Anthony Galvan, we had to stop doing it.
    - People using our data for proposals
  + Our own staff has found data in the open data catalog, it means that every job we do is faster and costs less. When we get data requests, we add it to the site. –Spike
    - Does this contrast an open data policy?
    - Spike: We’ve adapted to do both, we publish our data. Data.openoakland.org

Who openly published the data we have available? Who has funding issues?

* A few hands for both
* Anthony: If we can’t pay for the process, or even if we can, we don’t necessarily have the resources to do it in the future.
* Pittsburgh is in a tough legal situation, where they can’t share not already public data.
* Oakland knows enough about legality to choose, but the lawyer default is always no don’t
* CKAN is currently one of the most funded sites, it’s a bear to set up and run but pretty well designed. Gives a good organizational base. Hearing about these issues is stressful as a new organization, but very useful. The idea is that this has to build once and then other organizations.
  + How many have formal funding to support data system in general?
    - Few hands
  + None of us are near the level of support we should have
  + This formal funding not equal to amount of effort
  + It’s a long term thing
  + When Casey was around in large cities provided funds, interesting intersection between NNIP and Open Data, to leverage what we’re doing. My analyst not responding to simple requests because it is available, and how we’re viewed though community important.

Why is there not more funding support for this?

* Our major funder, Cleveland, said we’ve been funding the same thing for 20 years, but now find someone else to pay for this.
  + Normally much shorter cycle 4-5 years
* Funders want to see results, you put the data out there and don’t see what happens with it.
* This is why it’s important to find stories of people using data, but these are hard to gather. Part of this is in marketing, and how to gather “Data Driven Stories”
  + Once you get them have to celebrate them
  + Hard to find new stories once you have them
  + Have to really seek them out
* We have funding to do Community Need Assessment for local hospital, in the past, any data they want to track that comes from a state agency, we made them do it, now we’re getting push back saying you have access you do it –Anthony Department of Health Services paid to do this. Funders will pay for a report using the data, but they don’t want to pay for the final product
* When it comes to quoting technical assistant jobs, it’s a stab in the dark, this would be a good workshop session
* One thing Oakland banks on is that being open means that people are doing things with our data that we know or can find out about. Every city is gun-hoe for data driven decisions, but nobody is funding the infrastructure, some cities are doing it. We’re all getting by with this stuff that IBM is charging millions to do, this is cause for concern, and ammunition to nationally talk about the state of data infrastructure. If this is the basis, this is week.
* CKAN is an attractive option, but what are the costs of getting and maintaining it?
  + It’s very hard to manage, which is why a coalition would be nice. Once running it works reliably. Data.gov.uk uses it.
  + Spike, took us about a week in a rush to do it, the newer version is better, Chicago (Smarter City Chicago) working on an Amazon mirror of it. They’re close. Sacramento had another similar program up in half an hour. Drupal module.
  + Greg Sanders knows about this
  + CKAN is open source
  + CKAN also allows for multiple users
  + This has been done by OpenColorado, one system for the state. Cities who want to be a part of can add data. Can even add skins to look like a city website. Oakland looking at this, but potentially California-wise as well.
  + Can the system manage aggregate statistics for more private data?
    - Not good for that
    - Ability to make things private
    - Better to aggregate ahead of time
    - This seems like a direction we would like to go
      * Data queues have been developed off of private individual data
      * We’ve seen a small backlash
  + Do these systems have feedback?
    - Socrata does, CKAN does but not very robust
  + I’ve seen some data visualization tools, how far from being ready are they?
    - San Diego has them all disabled
  + This seems more like a file server, but can’t be developed on top of. Any developer ready API things?
    - Spike, we’re working on one
    - San Diego want to store in a data warehouse, Amazon has one that can handle a lot of data, users would use data aggregates that would be linked in relational tables. Tableau and other business intelligence tools. Problem is Red Shift stars at $1000 a month. Could do the same thing on a smaller scale, RDS, amazon version of SQL. Skeptical of APIs, easier to get started but more fragile to changes in technology.
* How to help less data intensive people?
  + - * For publishing open data in needs to be about getting it out there instead of forcing connections. This appears to be the trend.
      * Data dumps serve a very limited set of people, and not the people we want to help
      * The audiences we’re working with want the raw data
      * Putting data out there isn’t making decisions happen, doesn’t guarantee anyone has the capacity to use it.
      * I started out with journalist. They acquire and clean data many times. We should tell this story about why this should be funded.
* Does anyone have other ways of getting money for funders?
  + In Providence, we’re doing data story development where we’re working with funders to curate and come up with the data, funders get appreciation for time and effort required to develop the data
  + Piton has recently done data connection of free clinics that will be important in health care reform, hospital funding for locating clinic and publishing user data and location data. We have long-term funding for it, we’re teetering with how successful data will be. Very sector specific and crises driven. Not the best way.
  + The sector view came out in San Diego. We have a small number of funders in the area and have trouble getting these funds. One reason I like working with journalist. If we can line up with their stories, we get more recognition. They’re even poorer than the non-profits.
* How many people have seen the work we’ve done in the last few years has been rendered unnecessary because of data driven journalism?
  + To some degree in Detroit, they’re publishing data we don’t have
  + Milwaukee journalist as well
  + Oakland doesn’t do homicide reports because journalist do it for us. This threatening work, but the more you help them the more they promote your organization.
* Places where people can learn more about these types of things?
  + Most tools coming out of OpenGov realm, a lot of things
  + Sunlight Foundation, start there
  + Book coming out later this year
  + Who do people encounter that are obstacles?
    - Requires a push to host and to cloud, which is barrier for some. Sacrata offers a subscription which helps.
* Thanks for joining, if you’re interested in CKAN, talk to Eric Busboom. Interested in creating a hosted site that people can just take.
* Email notes to Eric Busboom, eric@sandiegodata.org