Council on Open Data

October 3, 2014
9:30AM - 12PM
Governor's Reception Room, 2nd Floor
The State House, Annapolis

Introductions (Greg Urban)
- Matt Power is not able to attend today
- CIO from Montgomery County is here today to talk about the data cataloging process which addresses much of what we want to accomplish at the State level
- Want to talk about a piece of legislation and the approach for the next legislative session

Open Data Council Website (Barney Krucoff)
- Thanks to the Web Team at DoIT for putting together the website
- http://doit.maryland.gov/opendatacouncil/pages/default.aspx
- Key links on the homepage are to the Socrata and MD iMap data catalogs
- Information about the next meeting and agendas to be posted as they become available
- Meetings Tab
  - Will have documents available from each of the meetings, so local copies are not necessary, can always reference information from this location
  - Information will also be emailed to each member of the council
- Anticipate adding FAQs and other links as this website continues to grow
- Open to suggestions as to what should be included on the website

Possible repeal of Section 10-901 through 10-905 = "Subtitle 9. Automated Mapping - Geographic Information Systems" (Barney Krucoff)
(Presentation Slides Available)
- DoIT is proposing repeal of this law
- Law has been around since 1992
- Goal was to fund the development of GIS since there was no basemap at the time
- Funding for basemaps was through the sale of data
  - Governments were the only ones making this data and making it available
  - Now many providers along with the government are providing this data
- This law was unique because it also allowed for overhead charges related to the building of the system
- Fees were high at one point, but have dropped significantly
- Required a contract or license agreement also between the government entity and the requestor
- Want to find out if entities are still using the law
- Governments/Agencies would still have the right to charge based on the cost of distribution, labor for custom requests over two hours, CDs, plots (sq. ft.), etc. That authority is under the Public Information Act.
- Are state agencies or local governments recouping money that is spent by using this law? If so, please help us quantify the financial impact of repeal.

- DoIT's rationale for repeal the law:
  - MD Department of Planning (MDP) sold the MD Property View (MDPV) maps for 20 years using this law to fund it
    - Incoming funding had been falling for years
    - MDP does not have to sell their data anymore because money has been put into the budget by the Governor to cover the costs
  - There are legal costs for each agreement that has to be written
  - There are bureaucratic costs of time and energy to get the agreements in place and get documents signed
  - Opportunity costs from making it harder to work together
  - Not consistent with the open data law, which no longer defines GIS data as special from other data
  - There are other ways to recoup funding from the freedom of information act, but couldn't charge for the overhead of the system
- Asking the council to come to an opinion, specifically about the financial impacts
- Are there areas where this is successfully being administered that we are not aware of?
- What other information does the council want prior to either co-sponsoring with DoIT or sending a letter of endorsement to the legislation?

Question: Is there a one sheet that could be sent out to the locals to solicit comments?
- When we prepare the draft legislation, we put something together which can be massaged into a single page with a link to the original legislation.
- Barney agreed to provide such a sheet.

Question: Are we talking about a repeal of the law or a replacement for the law?
- Repeal only, other costs are covered in other statutes

Comment: Some County Councils are requiring sales to recoup costs and to maintain a perception of gaining revenue back whether it is a net income or not.

Question: Did MSGIC do an analysis years ago to determine if the amount of money that could be recouped was near what is out there?
- There are decades of support and proof that the costs cannot be recouped in this manner
- Stifles some economic activity because of trying to continue to charge
- Started with TeleAtlas sharing their data with SHA
- Overdo and an excellent idea, vote the group should support repeal (Tim Baker from State Archives)

Comment: Spend significant amounts of capital and then the State agencies are getting the data without contributing
- Geographic data definition has changed over the years
- Baltimore Metropolitan Council (BMC) still charging to connection to map services for permit data, is this GIS data or not?
- Is this type of data being addressed?
  - Will need to check in the law

Question: Does this law shield Public Information Act (PIA) requests from getting them for free?
● If a PIA was requested, they would have cited this law and provided for free for requestor's use, but requestor could not distribute the data to others

Comment: Baltimore County charged a nominal fee for planimetric data mostly for engineering firms
● Were not making money
● Changed to free, the call volume went down and allowed for refocusing onto other county needs rather than just selling a small amount of data

Comment: Washington County sells very little data
● Amount charged does not cover the costs
● In favor of this repeal, but good idea to poll some of the other counties and municipalities (Bud Gudmundson, Washington County Government)
● Volunteered to poll those in Western MD

Comment: Montgomery County used to sell GIS data. We changed that policy (Sonny Segal, County CIO).

Comment: The PIA statute that allows for charging if spending more than 2 hours preparing data
● All collected fees go to the general fund
● If charging for the data, people think harder on what they are requesting before making the request
● Under PIA can charge for custom requests
● This statute has allowed for set price sheets and standard products.

Comment: Repeal would get us ahead of case law coming out of California (CA) about how their law to charge for this kind of data is not in keeping with open data and public information laws in CA
● Good thing to get this out of the way before someone uses this law as a base for a lawsuit and fighting against something we don't believe in anymore

  ● Come back to this group after contacting the counties at large and those who support in the minutes. Bring this back in the November or December meeting for a more formal vote

StateStat Data Request & Tips for Using Socrata (Josh Exler)
(Presentation Slides Available)
Instructions for Account Creation (Josh Exler)
● Email Josh Exler (joshua.exler@maryland.gov) to get an account for Socrata
● The online account request form is for public users
● Josh will make sure you get the appropriate permissions levels
● Also contact Josh if you have an account, but it does not have the appropriate permissions, for example, if you do not see the "Create a New Dataset" button on your user profile page

Responses to Identify 5 High Value Datasets for Public Use
● Some State agencies did not provide responses, please submit ASAP

- Keep thinking about datasets that would be a good fit

Question: Have we considered publishing a few datasets that have been forthcoming and tick them off as published?
- Great model from Montgomery County, who will demo today
- From this demo, we will talk about what will be published and when
- Talked about statewide inventory at the last meeting, but not determine the process/timeline to get published

Tag Cloud
- Slide which displays all of the tags that are currently being used in Socrata
- Gives an idea of who is participating, most agencies are represented at least a bit
- Josh has made a view of all of the data that has been uploaded in the past month
  - Will determine which agencies were able to upload requested 3 datasets before this meeting
- Shows the name of the dataset, owners and keywords/tags
- Actively keeping an eye on the new datasets coming onto the Portal
- Appreciate when the datasets are being uploaded
- MDE, MDA, DHMH, MDP have been significant contributors in the past month
- Those who have not yet been able to upload, there will be a training session on October 10 @ 10am in StateStat conference room, in Baltimore
  - Training on how to upload data into the system
  - Perhaps something in person would be more helpful than just the online instructions
  - Start with the basics of formatting, upload and editing on the site and then move onto specialized topics, as requested
- Training open to anyone in attendance to today's meeting
- For more information contact Josh Exler (joshua.exler@maryland.gov) directly

- Barney: Data can also be submitted on the GIS/iMap side
  - In September agencies added Sports Venues, Green Infrastructure, Flood Data, and updated the geocoding engine (includes parcel centroids, highways exits) and more.

StateStat Tutorials Available Online
- [Formatting and Uploading Data](#)
- [Editing Data on the Portal](#)

Building/Maintaining a Public Information Act Request Registry (Josh Exler)
(Presentation Slides Available)
- Drafted an electronic form which would be standardized for all agencies
- Will not replace the submission of written letters from the general public
- Can supplement for quicker response
- Preliminary form up for discussion and edit suggestions
- Office of Attorney General has given *tentative* approval
- Associated database where some of the fields will be shown in a Socrata database, so can be seen if something else similar was also requested

○ However, this will only capture those requests submitted through electronic form

Comment: Form needs to clearly ask what data they are referencing and asking for, whether current data or not, etc.
● Need to be more specific about asking what data referring to

Question: Would an agency fill out the form if they receive a written letter? If so, this could be centralized registry of PIA, going forward.
● May move this direction in the future, but not a catch all database for now, database is the bonus
● Cannot log all forms ever received, not past, but perhaps can do so moving forward
● Those agencies that want to digitize their written requests could use this form

Question: Request type, can that have a dropdown, looking for specific types for cross referencing?
● Anticipated not to be the type of request being asked for, check with Josh Exler about the meaning

Comment: Catch data requests and track data requests
● If we take written requests, would want to attach written request, need to add document attachment to electronic request
● Need ability to check on the status of the request (received, processed, responded, etc.)
● Can add these as additional internal only fields

Question: Will this be publicly available what has been requested?
● Yes, but will hide contact information

Comment: Be aware of media PIA requests, many for internal communications, especially around policy decisions that are made
● There might be concerns from the media if a request from a competing agency would be searchable

Question: How would electronic requests be routed to the agencies?
● StateStat would be the initial POC and would distribute as necessary for requests coming in without receiving agency known/specified

Question: Agency would be responsible for responding to the requestor and StateStat so that the response can get into the database or can directly enter response into database?
● Do not anticipate changing the workflows of PIA requests, but might want to establish a POC for new requests coming in through this database

Question: Is there are automated way instead of having to get filtered through StateStat to get requests to the appropriate agencies?
● Some requests, where they know which agency has the data, this can be automated/streamlined
● When receiving agency is not included in the request, StateStat (or another party, TBD) would need to be involved to get the request to the proper agency

Question: When does the clock start counting for the request?
- 30 days from when the written request has been received
- User hitting submit needs to understand that the clock starts when the proper agency gets the electronic request
- Automation is important so the process can be streamlined and the agencies have ample time to respond

Question: Anticipate the electronic form will make it easier for people to make a lot of PIA requests, do the agencies have time to handle the larger number of responses and the feedback from the end users who have comments? Does StateStat have time to route requests which are missing receiving agency information?
- If there is a significant increase in requests, we would reconsider this process
- StateStat has designated agency contacts that the StateStat team routes questions to in order to get the request to the right person in the agency

Question: Why not allow for the requests to be made directly when the requester knows who to ask? Informal requests can be easier for staff to handle, example provided DHMH and Archieves. Once request is PIA there is a more formal process.
- Trying to make the PIA request more accessible, since there is no electronic option at this time. The online form could include a POC at each agency, so the requester knows who to ask.
- Suggest having an upload of a formal request which can allow for determining the validity of the request

Comment: Important to have the requestor indicate what agency they want the request to go to because if this is misinterpreted by a State employee, this cuts into the amount of time that is available for the request to be fulfilled
- Want to have public access, but you have to make sure the agencies have adequate time to respond and redirect
- Public more apt to misinterpret data owners than are State employees

Comment: Make agency a *mandatory* field with drop down list and tie an email address to each agency, which would allow for automated routing to the designee
- Modern workflow that requestor says where it should go to, along with documents attached electronically
- Submitter is not required to specify the agency that the request is going to at present, as per OAG's guidelines. Form could request this information but not make it mandatory.
- With a written letter they have to send the letter to someone, which is essentially the same as requiring them to pick an agency to route the request to, while not do the same here

Comment: Has significant implications for some agencies' workloads and workflows. MDE suggests a subcommittee be setup
- A PIA subcommittee meeting will be held before the next Open Data Council meeting. As a follow up item, please email Josh and Barney if you are interested in being on the subcommittee.

Question: Will the public understand the difference between a request for data and a PIA request?

- Some of these answers can be provided quickly without having to fill out a form
- Have to make clear that requesting data is different than a PIA request
- Need to understand more clearer the line between the two requests
- A PIA request has to be handled in a specific way, because of existing law, even if it is a data request that could be fulfilled without this designation

Comment: Senator Ferguson has received comments about how these requests are being handled differently by different agencies
- Interesting process question, especially with appeals process and there are details to be worked out

Subcommittee Volunteers: DoIT will send a request for volunteers who want to work on a PIA subcommittee.

StateStat Data Request & Tips for Using Socrata (Josh Exler) continued...
(Presentation Slides Available)
Sensitive Data Handling
- Spelled out in memo, many options for what is public/private and how users can be provided access
- Uploaded data is private by default and can make it public later
- Manage Menu > Sharing can email the dataset to anyone with a Socrata account (need their Socrata email), can allow another user to view, become a contributor or transfer to another owner
- Can share with others even if the data is kept private in nature

- Ability to make data series public/private, this includes specific data fields with a dataset
- Some background information about working with HIPAA data is also available and will be detailed in the follow up to this meeting

Data License Types
- From one of the Socrata pages, included in the memo
- Either "Public Domain" or "Have Not Selected" are the greatest options
  - Public domain is the best option, but please go in and assign a license type for all data in the system

Question: Can some of the license types be removed?
- Not an option with Socrata, might ask for this feature in the future, have more choices then we need

Montgomery County Data Inventory & Open Data Prioritization (Sonny Segal & Team)
(Presentation Slides Available)
- A little over 2 years ago, the County Executive launched Open Montgomery Program, includes 4 components
  - dataMontgomery

- ■ Data will get consumed and something will result from that
  - ○ mobileMontgomery
    - ■ Data to become mobile and consumability on mobile devices
  - ○ engageMontgomery
    - ■ Wanted to know what datasets were available to consumption outside of government
  - ○ accessMontgomery

- ● dataMontgomery Team, supported by Data Technology Services
  - ○ Governance is diffused with involvement of County Executive Office and Council Members for Digital Gov
- ● A internal partnership
- ● Also became a part of county staff program, performance measures whether Departments are contributing or not, have inventory or not
- ● dataMontgomery Team: John Gillick, Victoria Lewis, Melanie Wainger
- ● Started a pilot program for open data in 2012, goal with having single source site roll out
- ● Later on in December 2012, the County Council passed Open Data Act and County was already prepared to take action
- ● Key factors of Act
  - ○ Prior to Act had to ask for data between agencies
  - ○ Act mandates all groups must publish all public data, do not have to ask anymore
  - ○ No longer if data will be shared, but when
- ● Mandated publish PIA responses on open data board, put PIA requests out to be reviewed by public
- ● dataMontgomery launched about December 2012
- ● Now up to 50 datasets representing most of the Departments, including employees salaries, food inspections, etc.
- ● Open Data Act also mandated two documents:
  - ○ Technical Standards Manual:
    - ■ What data formats are accepted
    - ■ What formats can be distributed
    - ■ Answer is what Socrata can support since they are the vendor
  - ○ Open Data Implementation Plan:
    - ■ What dataset inventory contained
    - ■ 550 datasets identified
    - ■ Contains timelines of when data to be published
    - ■ Contains many other components beyond requirements
    - ■ Factors for prioritizing data
- ● Lessons Learned:
  - ○ Looked for published data already to streamline review to first get started
  - ○ Factored in data that would be most compelling to the public
  - ○ Have to clearly define purpose for publishing data
    - ■ Why are we doing this? Checking a box? Form some specific outcomes?
    - ■ Transparency and accountability in government was a County Executive bullet point, which this initiative falls under

- ○ Formalize charter signed by CIO

- Three-tiered Governance Structure:
  - ○ Executive Steering Committee
  - ○ dataMontgomery Working group
  - ○ County Departments
- Help navigate the road blocks that come up when trying to accomplish the mission of the dataMontgomery initiative
- Executive Steering Committee has enough authority to help work out some of the issues that arose amongst the various Departments and their varying missions and workflows
- Departments all required to publish a dataset, all Department Heads must sign off on all publication of data coming from their Departments
- Best practice:
  - ○ Thoroughly define the datasets before they are published
  - ○ Have sign offs and reviews in place
  - ○ Everything documented in metadata, so citizens can pull up data and understand
  - ○ Document data up front to avoid calls once published
  - ○ Have data definition, data schema specified
- Review process is another best practice, have a legal review process to make sure privacy issues are observed, not violate any existing laws, attorneys must sign off on all datasets that are published
- Emergency calls are all documented and reviewed and then provided as a published dataset
- Datasets influenced by multiple Departments or controversial datasets are raised up to the Executive Steering Committee to get processed

Question: Do you know the average time to get the dataset published? How long in review?
- This is a rigorous process, there is a dataset of datasets
- Scheduled to publish 57 datasets for next FY, lean team, does take a month to over a year depending on the content and the intent of the Departments to get the data out
- Structure does slow down the inventory to date, but are not getting questions about appropriateness, questions on content, etc.
- Policy to automate all datasets, do not want to have to manually update the many datasets in the system
- Datasets are updated at all different intervals, annually all the way to real time, slows down the process because not asking for spreadsheets, want to set the data transfers up to be automated, so this slows down the initial publishing process because the original source system has to be understood and worked with to get into Socrata
- Email notifications built in to identify automatically if something was not updated on scheduled or other errors

- Implementation Plan recommended to include:
  - ○ Inventory all available datasets countywide, prioritize the list, compile into a schedule and overall plan
  - ○ Countywide effort that is a people project, not just a technology project

- - Objective way to rank and prioritize these datasets, part of the law that had to be followed, need to be able to respond to why datasets were scheduled in the order they are being scheduled, in case this is questioned
    - Deadline to inventory, compile and submit the plan for review
    - Include County Stat members, Chief Innovation Officer, Department POC (1 person per Department) to educate on open data and have an on-going relationship with, Workgroup and Executive Steering Committee and Project Team
    - All members of the initial committee to get the plan put together and ready for implementation

- Started an education process that started before any datasets were submitted
  - These were not IT people necessarily, but needed to be familiar with Departments, what Departments did and who would have the various components they would be looking for
- Define what a dataset is was established as a common understanding
- Tips on how to inventory had to be established to realistically accomplish within an existing full-time job
- Meet with the group more often than just during inventory process, also help with the implementation plans and other issues, open data advocate in each Department

- 4 main steps:
  - Establish data experts
  - Id data owners
  - Establish dataset list
  - Establish kick-off meeting

Question: Did you inventory all data, even if it would not be public or not?
- Inventoried everything initially because the charter indicated that just because you do not think it should be public there are not components that could be made public
- Looked at the data more closely, could omit specifics or roll it up/aggregate the data and then it was able to be published
- Sometimes this was not the case, but there were opportunities that were discovered because of this inventory process
- Gap analysis took initial submissions, compared with application inventory, looked at performance metrics and looked at Department websites, as external users, to see if the data matched the perception of what the Department does
- Main focus of Implementation Plan was from us to residents
  - List objectives
  - Explain how to prioritize datasets and comply with the bill
  - Engage public with the effort
- Community outreach was conducted formed of open data town halls, asked residents what data they wanted to see, had focus groups on some specific questions that the government could provide answer to
- Opportunity to extend the repository to engaged with residents
  - Using data to help manage data performance, to be released soon
  - Take budget books and replacing them with online budget publications

- - Taken datasets with complex financial data and provided guided views
    - Result is budgetMontgomery to see CIP and operating budget and see capital projects by neighborhood, see phase they are in and see the funding sources
    - Show view of checks that are cut by the county, who is spending, on who and what
    - Includes multi year view to see down to a list of transactions and down to the check level itself

- Question had been posed as to what data the State could provide Montgomery County and open up a channel of collaboration between the two entities
    - Data that compares cities
    - SDAT property data
    - School data
    - Uniform crime reports
    - Grants data
    - Transportation plans
    - Election data
    - State issued permits and licenses
    - Continue to get responses from Departments on other data that the county would like to consume from the State

- Matt Power and team from County Executive Office visited Prince George's County about open data, crime data and transit data
    - Had fruitful discussion about cultivating a partnership because there is a difference in progress of open data programs between these local entities, but encouraged to work collectively

Question: How centralized is the process? For example the legal review, does one lawyer who is an export perform the review or does it go to various lawyers who support departments?
- One lawyer responsible for all open data reviews, they may collaborate with other individuals within their Department or other Departments, but a single POC
- Multi Department datasets have to be coordinated so that data when data is federated in a particular view it does not take on a different risk level
- Attorneys put all of the datasets together to make sure that sensitive data is not revealed
- Spending data had multiple lawyers involved for review to bring subject matter experts together
- Have 650 computer applications that are being tracked in Technology Department, so infer there are at least 650 datasets out there
- List that made it to the plan include 500+ because of the criteria review, can always add datasets to the plan, County Council has to approve changes to the plan

Question: Have you noticed changes to PIA volumes and costs?
- PIA requests have gone down
- Departments continue to receive PIA requests directly and Departments register to the PIA site and prefer to respond by publishing the responsive documents to the MPI site

Question: As CIO, do you have to balance security side and open data side, how to get these two sides to work together?

- There are security considerations at many levels, at network behind which all the automated processes reside
- Public facing website has to be protected from malicious intent
- All levels of security exist, networks, applications and data layer, this group participates in the review of all these levels, it can slow down the process
- Some of these processes are not the most agile, but are thoughtful

Question: Explain the ETL process?
- Extract, transformation and loading of data
- ETL process does not have an enterprise tool, but various tools to extract from larger databases and transforming them to the format for publishing and the loading is done in a number of different ways
- Sometimes staging of extracting data is key to the success of loading and minimize loading time
- Try to automate machine to machine and minimize handling of data, but there are cases where data has to be manually handled, valuable data, but not there yet for automation
- Starting to drive enterprise architectures and systems, at procurement time we require RFRs to state the data availability

Question: How to address the budget of this project?
- Fiscal impact statement had been done by 2012 and approved by county council
- Funding would be front loaded to get started and put processes in place
- Funding for out years is diffused through doing business in each of the Departments, there is no dedicated funding
- 2 years in and fully operational

Question: Key to success was having a good inventory, tend to write a memo to agencies requesting an inventory and get across the board responses, what was the level of effort in Montgomery County handheld agencies to get the inventory?
- Varied similar to for State agencies, at times received a quality product and other times the discussion involved investing time in sitting down with the dept to have a different discussion
- Some Departments thought it was a technology effort, not business-oriented, so the datasets must reside in systems they were not familiar with, but once business people and technology people engaged together through discussion that both sides together could become quite thorough in inventorying datasets
- Learned from what the State is doing and learn from what the other Departments are doing, they become examples for other Departments

Question: Are you archiving older dataset submissions?
- Defined with each dataset, is this kept, replaced, etc. this depends on the business opinion for each dataset
- Permitting updated daily and shows open permits
- Food inspections show last 24 months
- Archive for internal use
- Can go back if some archival data is brought into question

Question: Data requests for email, politics, procurements, what is the history of collecting and posting this data in relation to PIA requests?

- Example: procurement process that is competitive, PIA request comes in for information related to elements in the winning bid, does this type of information posted to the public, how is this handled?
    - Handled as part of the PIA screening process in Department of General Services
    - Make determination to respond or not directly or determine if worthy of posting to PIA site and post entire response
- Behind it is the approvals needed from the offerors as to what can or cannot be posted, case by case basis, generally includes an appeals process

Conclusion (Barney Krucoff)

- Know financial impacts of changing the law
- Follow up with volunteers for a PIA Registry Subcommittee
- Likely to be a Statestat follow up memo with some tasks in there
- Thanks to Montgomery County for coming and providing the presentation
- Will distribute dataset list from Montgomery County
- Think about what local data the State agencies might want, what would be nice in this format?
- Great to hear what some of the other local governments are going similar to mont co
- Expanding efforts to agencies "outside" of government, which includes schools, colleges, parks, etc.
- They do publish information which is not to be confused with datasets, not consumable, might be in a pdf for example, etc.
- Thanks to the Council from Montgomery County (Sonny Segal) for the opportunity to present to the group

Future Meetings

November 7, 2014 @ 1PM
Location: Governor's Reception Room, State House, Annapolis

December 16, 2014 @ 1PM
Location: TBD