

Council on Open Data

December 18 2014

12:30 - 3PM

Governor's Reception Room, 2nd Floor

The State House, Annapolis

Introductions (Greg Urban)

- Welcome and thank you for coming

Thoughts on Next Year (Barney Krucoff)

(Presentation Slides Available)

- Next meeting scheduled for mid to late February
- Not sure what will happen with the transition team
- Hope to have Council on Open Data meetings every other month in the future
- Looks like Josh Exler will come to Department of Information Technology (DoIT) and focus on Open Data full time, administer Socrata and work on the Council for Open Data future agendas
 - This will allow for getting out more deliverables
 - Not yet a done deal

Draft Council Report to Legislature & Discussion (Barney Krucoff)

(Presentation Slides Available)

- Report due to the Legislature in January
- Have created a draft and have distributed for review
- By January 10, 2015 the report has to be turned in, it does not have to be the report that is available, willing to work on follow ups directly with individuals
- Report has four (4) parts
 - Background - Describes the Council and provides some summary
 - Status of Open Data in Maryland
 - Findings and Recommendations - Section of consensus findings, which do not have to have, but do not want to have to have a vote where we count
 - At the end of the meeting, if we have a consensus on how to move forward, that is how we will proceed
 - Legislation - Edited legislation, mostly on sale of data
- Looking for general comments on the first two sections and then more detailed comments on the last two sections

Section I: Background

- On the act, on membership and other related items

Comment: Include previous success stories that could be from global sources?

- Help for those who are not familiar with open data, to provide context
- Does not have to be in this section, but recommend this is included somewhere in the report

Comment: Important to recognize that state agencies have established open data portals on their websites (For Example: MDP, DNR)

- Include a statement recognizing these facts
- Perhaps include some references to local governments that also have established portals
- Link included that goes to the Open Portal page on MD iMap, along with list obtained from Maryland State Geographic Information Committee (MSGIC)

Comment: Do not like the word “encourage”, perhaps “deliver” or “provide”, which is a little more firm of a word pertaining to the commitment of delivery

Section II: Status of Open Data in Maryland

- Brag a little bit about what is going well and what we have been successful with to date
- Some examples about how this has been successfully achieved in the State, even if not connected directly to the Council
- Includes the Center for Data Innovation study, Maryland is tied for the top spot, 6 states at our level, we are at the top level
- Talk about the two executive orders, related to a later recommendation
- Two portals
- Links to maps and dashboards we are aware of, this is not a complete list, but attempts to be
- Links to existing open local sites, both GIS and other (general)
- Links to particular study that MDP did, to be discussed later in this meeting
- Links to financial transparency data that is on Department of Budget and Management’s (DBM) website

Comment: Additions from presentations that were had in the past, the environment dashboards, with outright links

- Some discussion of the progress being made at the local level, that we are trying to learn from their lead
- Might include some pictures as well into the draft

Council Deliberations

- Talks about the data sales, mostly describing what goes on, not the findings yet, which is later
- Talks about the open data and how this crosses into more traditional forms of government transparency
- Inventory, our portals

Comment: Recent PIA hearing from Senator Ferguson, might be some interesting feedback that might be included in this report?

- Have talked about the PIA inventory request
- The hearings have talked about how the local municipalities have not been following PIA laws and perhaps abusing the laws
- Is this appropriate to include in the discussion
- Should include something about the value of the PIA process related to open data

- In this section shows that there is a connection between open meetings and PIA, later recommendation which is hands off PIA, that Legislature should think about and talk about open meeting calendar, not just in register, but in online calendar
- Nervous to call out local governments when we have not talked about that here
- Committee on Transparency has been joined to Legislative Data Committee to create a single committee
- New flavor and the Committee will be renamed with new session
- Overlap with open data and transparency of policy and implementation, made sense to pull together on legislative side
- New committee does not exist yet

Question: Addressed this report to Senate President, Senator Ferguson, should this new, combined committee get a copy?

- Yes, have to get proper citation

Comment: Merits and feasibility of statewide data inventory

- This is already a request for naming the data for systems that are already in place
- Suggest it says expand or further refine initial efforts that have already been initiated by DoIT
- Inventory of systems and not the data in them, we can strengthen the recommendation, so that DoIT does not ask agencies twice
- IT master plans are focused on larger projects which may or may not include some of the spatial data components, good idea to leverage what is already being done
- Add to the recommendation that is below on the document

Section III: Findings and Recommendations

Finding: State Government Section 10-901 through 10-905, Automated Mapping-Geographic Information Systems

- First finding is about the law, long list about where data sales is not a good thing
- Last finding is that agencies still want to be able to sell the physical products which have a cost and a cost associated with preparation of physical products (DVDs, Paper Maps, etc.)
- If all requests for a standard product as PIA request, you could charge, but put through more process and put the agency through more process, since these requests need to be reviewed by lawyers, panel reviewing
- Legislation trying to work to no longer sell to recoup overhead costs of system, only costs to prepare and distribute the physical product
- Would not charge for online, eliminate the need to establish a contract, retain the ability to sell and recoup costs to prepare system products, definition of system products included

Comment: Eliminate the need to enter into contracts

- There are companies that want to resell crash data, if you provide data to those companies, without a contract
- There is no way to control how the data is reused or preventing them from reaching out to those who have been in the accident, citizens have been contacted in the past
- The law says now that you must have a contract, another agency that is distributing something innocuous must have one, but does not exclude having one, just the must

- The privacy of a person's information is allowed within the confines of the new open data law, to be excluded
- Allows for our data to be web-enabled without the requirement to have a contract
 - For Example: Take property data and make available through other applications, do not have to a contract to do so, can, but do not have to
 - Baltimore County Example: Get data from MD Transportation Authority, long contract that has to go below county executive level, but cite this section of the existing law in their correspondence that this has to be done, government to government sharing of data, agency want to follow the law exactly
- This would give the option to not go through this rigorous process to share data, hopefully they would choose this option if available

Finding: Executive Orders 01.01.2012.04 and 01.01.2012.18

- There were two Executive Orders that established the program we had before the law was passed
- One is for MD iMAP, to have an Executive Committee, this Executive Order and the legislation with the Council on Open Data are almost identical
- Would be giving up some if we have the order repealed
- The GIO title is setup by this Executive Order, but the Secretary of DoIT can choose to still call the person in the position by the name of choice
- Also set up Open Data Working Group, data.maryland.gov, setup Chief Data Officer name and position, which is something else that would be giving up
- Leave all three in place, there would be three panels that are heavily overlapping, do not want that many meetings or the questions why these groups have not met that often

Finding: Gubernatorial Transition

- Ask Greg Urban to submit the report to the transition team who are already speaking with DoIT
- Ask to look at the websites and the committee, singles out Statestat as a heavy contributor to open data, both as a user and a contributor, not an actionable item, just acknowledgement

Finding: Open Data and Open Meetings

- Web team from DoIT working on open data meetings calendar, kept in Socrata with calendar interface and a user can see all open meetings in a single calendar
- Already required to tell to open register which is on Secretary of State website as. pdf
- Want to automate this process, allowing for friendlier interface and offer the same information in both places, can also create a service/website from the data being kept in Socrata
- Recommend this to the agencies, each agencies report their open meetings, hope to do so in one place, Secretary of State will get it into the calendar
- Can also do links to documents, map of where the meetings are, which is beyond what the register list can accommodate

Finding: Statewide PIA database

- Talked about PIA database, two things:
 - Should there be a database that tracks PIA requests and how they are resolved
 - Should we make it easier to make the request, backed away from automating the request, easy now, send a letter/email

- Idea of the database is a good idea, but not ours to act on, the Legislature, Governor, Attorney General are involved and should act
- Think about, agencies have databases already, Maryland Department of the Environment (MDE) for example has an extensive database and process, tie to existing databases
- Include existing databases to the recommendation
- Think about who we make the recommendation to or are we just throwing it over to the Legislature?

Question: Is legislation required? Would it be beneficial? Is the timeline of implementation recommended to be included?

- Give agencies opportunity to get fiscal note
- If we did it now it would be inside agencies existing budgets
- Not a huge task in theory, put these results into the database would be more considerable
- Legislation might get it moving and a nice example, since the legislation now is a paper oriented thought process, give people a chance to talk about fiscal impact
- More on capturing the activity and not the results, topic and disposition, not a repository for all data, acknowledge receipt, how it was acted upon and the authoritative source of the results

Comment: MD PIA does not require a requestors name or address, anyone can send without this information, how do they remain anonymous?

Not sure how, but apparently is an option

Comment: At MIEMSS, get requests for data for researched purposes that still have to be reviewed by the panel before released, is this the type of data looking for or just PIA requests?

- PIA request could be for data, for additional analysis and review
- Wording for PIA requests under the PIA Act, if there is another law that should be covered by this recommendation
- Research requests are a different beast than these requests being suggested

Comment: Anytime someone comes into an open data portal and needs one more piece of information, this would become a PIA?

- No, but how are these defined?
- In discussions, difference between informal requests and formal PIA requests
 - For Example: Want data that is public data, cuts CD and makes delivery, other side letter goes in with PIA request
- Legislation as-is, 13th edition of PIA manual, does not make it clear between physical and electronic requests, general email through open data portal requesting data, even though not labeled as PIA, could also be treated as PIA
- Need to clear whether electronic data requests do not express PIA, whether these would be considered a PIA regardless
- As open data becomes larger and more common, need to be ready for responding to these questions
- PIA database would allow for determining where to focus for open data to get that data out there without needing a PIA request in the future

- Appeals for PIA to open data would need to be prepared to handle, cuts out a level of first line interaction, more information gathering to a degree
- Each agency has discretion on how to appeal, sometimes the appeal is binding or advisory depending on agency, there is no common place for appeals
- Could change to state PIA appeal database from State PIA database, cuts out a lot of the information would not want to start there, keep as option
- Montgomery County there has been a substantial drop in costs related to PIA, once more data was made open
- Might want to get some concrete numbers on these drops
- NY did a report on open data findings, decreases in FOIL requests, additional stats that could be used
- Important to inform what data is liberated, recognize it is tedious work of open data, but important for informing on the process we are going through
- Hard to capture what employees are sharing based on discretion, we would have to ask each employee to log these interactions, would be good information to share the availability, but hard to track
- To keep this simple, work within the PIA process, when threshold is made, consider tracking date, resolution, leaving aside appeals, which appears to micro manage, want to discover the types of data we want to put out as open
- PIA manual specifies that requests do not need to be written or formal, do not need to id who, just asking something of an agency qualifies as a PIA request, this is not as formal of a request as we might think
- Focus on the PIA which has a predefined notion of what it means
- Add recommendation DoIT should survey agencies to see what is typically requested or replace with something similar, in addition, in replacement or none of the above
- Also want to be able to capture the data moving forward, a regular survey or should this be something that is decided by the council on open data?

Question: Are we talking about finding out from requestor did they get what they asked for?

- Like this idea, follow up
- Front end is 311 for PIA, dial in, funneled, tracking code, fulfilled/not fulfilled, more regular you can say what requests were avoided through open data, can make the case for why open data is successful, can be difficult to implement this process
- There are different agencies that do these requests differently, want to consolidate and aggregate PIA resources for each agency, good idea to have a central place to find where to be routed for your requests
- There is a list on most agency websites to connect to PIA, but want to have a centralized place to go to all PIA requests
- Create a class of PIAs which does not exist in practice, but in theory
- Are we not ready to make this recommendation? - No, remove from report
- Do not think this one is fully baked yet
- Requires further study, no recommendation, Council on Open Data will debate this further

Question: History of the request?

- Similar to routing bullet, when did it come in, did it go to the right agency, now it is in appeal, track the life cycle of the request and final or current status
- A little clarification of what is meant by the history
- Do not know if one agency knows if the request then goes to another agency, might tell them if the first agency does not regulate what is being requested, but requestor would be on their own to determine
- Most requests come directly to an agency and usually respond or do not but do not usually know who to send the person to if it is another agency, closed at this point, no history beyond one agency
- Leave discussion in findings, further study in recommendations column

Finding: Statewide Data Inventory

- Already said that DoIT should consolidate with existing IT master plan process/request already going to all agencies
- Desirable to have an inventory, that would be the basis as to what is and is not open, we do not have one, we have what is in the two portals, recommendations to have a pilot with some agencies by June 30, 2015, report back to this Council, reach an inventory for all agencies by end of 2015
- Ambitious based on the presentation from Montgomery County, concerned about the timeline
- Originally put in as two years, but already received comments that that was too long
- Use the 6 month pilot to inform how long the overall process should take to complete
- Council would set a longer term goal based on what is found in the pilot

Question: Can we extend the length of the pilot?

- Is there reason to believe we cannot get a pilot done in six months, so why extend the length
- No committing to a specific number of agencies
- Quick job, less good, have not defined the final project, so lots of leeway on pilot
- IT Master Plan process is an annual process, due in August 2015 for report preparation

Finding: Maryland's Open Data Portals

- Typo was fixed
- Have the two existing data.maryland.gov and imap.maryland.gov
- Finding is that two is one too many, there was a good reason to have both, from Executive Orders
- Recommend we consolidate, does not mean we put everything in one platform
- We do research at DoIT and federate search, replicate the search and determine the best tool behind the scenes
- Asking DoIT to come back to the Council with a recommendation

Finding: Local Government Participation

- Report talked about the local data sites that are available for free, but do not have that many that have taken advantage of this resource
- Not overwhelmed with the take up
- Appeal from the Council that locals should take advantage of this free resources
- Contract good through February with option year which will be exercised, good through February 2016

- See what the take up is and reassess continuing

Question: Can it be a whole county?

- Yes, as long as population under 100,000

Section IV: Legislation

- Trying to meet the bullets with the changes in the legislation
- Bracket indicates being deleted
- Caps indicates being added
- Done what we said we would do, cut out sell for higher price, require contract, enter idea of system products
- Most of it is gone, but this is the format that it is required to be submitted in
- Follow through with comments, redistribute first week in January with new version
- Get these comments in do we have consensus report? - No objections
- Distribute for a day or two, submit on January 10, 2015

Selling Data vs. Open Data, MDP Data Usage Stats (Jim Cannistra)

(Presentation Slides Available)

- Implement open parcel data over the past few months, was traditionally sold
- Blog post was published, on MDP site and MSGIC site, discussing these topics
- With support of MSGIC, MDP, DoIT, Governor's Office and MD iMAP Executive Committee, just first started open parcel initiative, took some time to assemble all resources, soft launch in July 2014
- 300 subscribers as of this date, made them aware of the change
- Mid-August 2014 went live with press release
- Work on an addition year cycle, so subscribers for 2014, provided DVDs until recently in addition to having data on the website
- Last DVD that was sent out was on December 1, 2014, can still give them out, upon request, will be a charge
- Three primary products: MDPV, Finder Quantum, Web-based app (Finder Online)
- Hard copy and digital tax maps still being sold, as part of prices on website, link to GIS data download page
- Messaging offers a contacts database, so users can receive emails about updates, all subscribers and anyone else who has signed up for notifications
- Seems to be working well, over 1,000 people on the list
- Data organized on the download site, similar to how it was organized on DVDs, product and County with .zip file for download, some statewide datasets also available
- Looked for the best way to distribute, these requests get the greatest amount of traffic on the website, did not want to overwhelm the servers, put on dropbox for download, each file is about 1GB of data
- Worried during testing for the length of time to download, a lot of people who use the data are not as technologically advanced GIS users, expected a lot of calls about download, have not gotten any of these types of calls, happy with the process

- Data included: parcel points database, linked to SDAT tax account information, extracts of customized parcel data, tax maps, zoning, data layers, parcel polygons, getting either ArcGIS or Quantum GIS product which is stylized already and executable to utilize the open source software
- Bundled with the files, links to MD iMAP services, the raw data and service links
- Started downloads in July 2014, some statistics were collected, great impact after the press release
- 3,933 for MDPV and 802 for Finder Quantum to date
- Consist pattern to the number of downloads
- Interesting to see the consistency between the MDPV bundle and the open source software download
- Overall, almost 4,800 downloads in 6 months
- Last FY, shipped 1,275 DVDs, in less than 6 months been 375% increase over previous year of data distribution
- Great, make distribution more efficient, more accessible, all accomplished
- Does not include any data through MD iMAP that can also be downloaded, .zip files only

Question: 1,300 DVDs that were previously shipped? Percent FTE? Compared to now?

- Glad to not have to call for paying bills annually
- 1/2 FTE subscription manager, manage database, mail DVDs, huge box of DVDs, shipping and mailing, definitely lot of efficiencies
- Send one DVD to the subscriber at an agency, but internally they would have to find how to redistribute to field offices, for example
- All analysis done based on Google Analytics
- Larger jurisdictions have the largest volume of users downloading data and vice versa
- Some variables effective by when the email is sent out
 - For Example: Sent out an email that Garrett and Allegany Counties were updated and then these counties bumped up the list, so the communication impact is being seen in the download frequency
- See how the messaging is impacting who is accessing the data, gets to the question of where the data is being used
- Counties with similar data on local website might have lower numbers than counties without this replicated resource
- Had subscription based web application, bought in regions, flat rate last year, now free
- Interesting to see the total and unique visitors since August 2014, from 40 to 50 to almost 500 total visitors
- Hope to have a mobile and tablet based version in the next few months
- Been successful to date
- Have saved some time and effort
- Gotten more questions about different data, which is hard to measure
- Objective to get the data out there and used and are meeting that purpose

Question: Do you know who the customers are? Businesses? Citizens?

- Normal subscribers were pretty diverse, most state agencies, counties, real estate development, planning, surveying, engineering firms
- Have not looked at the greater detail of who and where they are coming from

- Been an expansion of GIS community
- Expect in the future to be more and more occasional GIS users
- Number of open source folks have increased, not in the standard GIS community as much and now accessing the resource

Question: Use Dropbox, but have other resources been explored? Is it a State Dropbox Account?

- MDP license, State does not have a master contract with Dropbox
- MDP under the threshold of the allowable amount of data for an enterprise license (50GB) before having to pay more
- Looked at MDP servers and DoIT available servers
- Maintain excel spreadsheet that is linked
- Does not include the imagery, but the tax maps
- Imagery even at most compressed level, would be well over a GB, so imagery connected through MD iMAP, contact another means to get the raw imagery, if needed

Question: Was Google Drive considered?

- Did not look at Google Drive in part because of timing, this was done before MDP had started to use Google products suite
- Might be worth exploring in the future

Question: How often is data updated?

- Couple of counties every month, push to website and send email notification
- Process is the same, just the end of the process that has changed for distribution

DoIT News Items (Barney Krucoff)

MD iMAP 1.0 Decommissioning

- MD iMAP 1.0 retirement on December 31, 2014 at 5PM
- If you are still connected to 1.0 your application will break
- Go to <http://geodata.md.gov/imap/rest/services> to find the replacement links

Transparency Site Data Update

- DoIT to come around with spending data, now 2013 on transparency site (spending.maryland.gov)
- Can look up all the payments made to a vendor, can look up by ZIP code
- DoIT to come to agencies through CIO committee, need to review the spending data
- There are times when the data has been wrong
 - For Example: All Ocean City spending looked like they were going Sheriff's Office, but that was not the case
- All agencies asked to review
- Spending data is also available in the open data portal
- Quick link at the top of the page
- Before the data was just in the session based website, now available and previous years also
- Coming out with an appropriations dashboard
- <http://geodata.maryland.gov/budget> (test in Chrome recommended)
- Can drill into the various blocks which represent the size of the portion of the budget distribution
- Newly developed by dashboard team and anticipate putting on DBM website

- This is appropriations data, not spending, would not match up perfectly between the two, in one agency's appropriation budget, but in another's spending budget, which happens all over government

Question: Will appropriations data go on open data portal?

- Yes, once site is released

Question: DoIT will come around asking for spending data?

- Data comes from DBM and each agency will be asked to review
- Not the first time this has occurred

Progress on Data Inventory Demo Using SHA Data (Josh Exler)

(Presentation Slides Available)

- Had call with Montgomery County and Prince George's County data sites
- Talked about data desired from state to locals and locals to state
- Preliminary list compiled
- What else is seen as the highest priority, what are you already trying to get data
- Opportune time to push locals to share this information with the state
- Monthly calls with Maryland User Workgroup which include Montgomery County, Prince George's County and Baltimore County and Socrata
- Additional layers that would be of interest in the open data environment
 - Vacancy data
 - Development pipeline
 - County educational investments
 - Centerlines
 - Address points
 - School surface areas/district boundaries
- Priority is to consolidate data sources and requests

Question: Addressing to be available?

- Right now only for state use, someone else asks, they are referred back to the counties
- Can obtain neighbor data for counties to obtain
- List was surveyed off of MD iMAP Technical Committee and need to send list to Josh Exler
- Montgomery County has fleshed out list that they will share
- List of items have been requested for the state to share with the locals
- Highest priority is: budget data and crime data
- Send out the list as part of the follow up, as we start to think about statewide inventory, criteria should be what data we are getting requested from the locals
- Started talking about standardizing the data categories between open data and MD iMAP portals
- Some overlap exists
- As we start to inventory, need to assign categories for metadata within the two sites
- Do not want to have data categories that fall in one or other, except some exceptions like imagery
- Review what would be redundant, overlap and what changes should be made
- GIS followed ISO standard which is map oriented, we would deviate from the standard when appropriate

- Might be some that are already the same data, but just with different title, could be quick consolidation
- Not clear on biota, structure, society contain, might want to rename or better define
- Drafting a list of data fields to be included in the inventory
- Agency that participated in the pilot would have more hands on assistance from Josh Exler and others working directly to make this process successful
 - Unique ID: Is necessary for number of reasons, removes duplicate work when each dataset is associated with a unique id, can automatically pull related information in the metadata
 - Dataset Title: Standard nomenclature
 - Whether Dataset Can be Public or Open Data: Two (2) options, programmatic code to check if five criteria are met or require a written summary which indicates yes/no and explanation why
 - For data already published on a site, the inventory should include links
 - Category: Part of required metadata
 - Contact Names / Email Addresses / Agency Name
 - Aggregate or Individual Records: Some data is secure if individual records, but could be open if the data is aggregated
 - Contains PII, Legal Descriptions?: Include some dropdowns that submitters can select from
 - Summary of Data Caveats or Restrictions: Beyond just yes or no not available as open data, help informed the process of which to upload first
 - Source Database: Is not a good reason to keep this private, would have use for the agencies logging in the data, public use could be to give them a better idea of what the process looks like from getting data from a source into the open data portal, do not see why this should be private in inventory - No objections
 - Time Period of Content and Gaps: Included in metadata on open data portal
 - Gap in the Time Period?: Could be barrier caveat to making data public
 - Subjective Data Quality Scale: Complete, accurate, well defined, etc.
 - Currently data type to select one to five stars in Socrata, which will be phased out now

Question: Advantage to have list on Socrata that there is a list, but you cannot get any of the data in the database? Is there an advantage to sharing this information?

- Legislation includes that if it does not meet the five criteria, then it is legally obligated to make them public
- Just cause you cannot publish does not mean you should acknowledge you have it
- If someone calls asking for it, it has been determine why it cannot be made public
- Most basic element of transparency is to tell people what we are collecting, acknowledge what we have, categorization said what are not public, what is and why they are not public and then work to make those that should be public available, make those open data elements open
- Want to know what data to protect
- We would collect this
- Get free requests if you are transparent about you have it and why you cannot get it

Comment: What about data that has some private information with it removed could be published

- Hope that this inventory allows for some of these datasets to be made known
- Retention schedule to be added, as part of time period of content, but perhaps add separately

Comment: DHCD might object to naming the source database, many of these are connected to database that are not to be shared, what exactly is being asked for, just want to have a common name to call it and a common place where it comes from that is understood by all potentially asking about the data

- Feel free to email Josh Exler with additional comments, not a finalized list, but good start

Future Meetings

February 25, 2015 @ 10AM - Noon

Location: Maryland Department of Transportation, Harry Hughes Conference Center