

Council on Open Data

April 14, 2015

9:30AM - 12:00PM

Maryland Transportation Department (MDOT)
7201 Corporate Center Drive, Hanover, MD 21076
Harry Hughes Conference Room, Lower Level

Introductions (Barney Krucoff, State GIO)

- Mix of folks in attendance
- Some new people joining us today
- Share if you are new/returning
- Quick briefing on the Council
- Onto the published agenda

- Council on Open Data was created by law with legislation signed by previous Governor
 - Went into effect 7/1/14
- First meeting occurred in late August 2014
- Overall purpose is to promote Open Data Policy
- Replaced existing groups of MD iMAP Executive Comm and Open Data Working Group
 - These other groups were set up by previously signed Executive Orders
 - Groups overlapped in purpose and are now consolidated
 - Council on Open Data is scheduled to meet four times a year (Quarterly)

- Council on Open Data Chair: DoIT Secretary David Garcia
- Council on Open Data Vice Chair: StateStat Director Kevin Conroy

- Two Legislators are Council Members: Representative Cullison and Senator Ferguson
- 23 named State Agencies as Members, either Secretary or Designee
- 5 Members from Local Government and 5 Members from Private Sector
 - These members were submitted for appointment to the Council
- Appreciate the time that those coming from the private sector and local governments dedicate to coming to the meetings and providing advice

- Law created the policy that open data be machine readable and made easier to find and accessible via portals
- Different from PIA, not dealing solely with documents and analog data, making data easier to use in spreadsheet and for developers
- Not replacing other laws, this new law is in the same spirit as other laws
- Two Portals:
 - <http://data.maryland.gov> - Socrata, cloud provider, has alpha numeric data to date
 - <http://imap.maryland.gov> - Esri, cloud provider with most of the geographic data

- Council on Open Data has a web presence on the DoIT Website (<http://doit.maryland.gov/opendatacouncil/Pages/default.aspx>)
- Annual report was completed based on requirement by law
 - A lot of the recommendations in this report have been acted on, specifically the ones given to the General Assembly
- Next steps are to determine what should be on the next report and how should we track the progress from the recommendations in the previous report
- Website always includes information on the next meeting and other future meetings
- Meetings tab includes all documents generated for a specific meeting, all in one place, including agenda, slides, sign in sheets, minutes, etc.
- Council has been in violation of Open Meetings Act
- There is a compliance board and we must receive these instructions
- We are created by law and we are subject to the Open Meetings Act
- First meeting was in August 2014 and we did not give adequate public notice
 - Now we have registered meetings and posted the website
- Staff people of the Council have taken the appropriate training
- Those employed on the public side might want to take the training also
- We strive to be open and follow all appropriate laws related to open government

Question: Where can you take the training?

- Online http://www.igsr.umd.edu/VLC/OMA/class_oma_intro1.php, certificate is available at the end of the course
- Takes about 30 - 45 minutes to complete

January 2015 Annual Report Recommendations Status (Barney Krucoff)

- Every year we have to provide report by January 10th
- Current Annual Report made several recommendations:
 - 1) [SB0094/HB0353](#)
 - Change two things in an existing law that was set up in the early 90s
 - a) Organizations with geographic data could charge to cover overhead costs of the system, can still collect a small charge based on cost of distribution (plots or CDs), but not for the overhead, cannot prorate the cost of system, servers, staff, etc.
 - This worked for a while generating revenue
 - As other resources came along the base data was no longer a government only and the revenues were falling off
 - b) Contracts had to be entered into to protect the ownership of the data, contracts no longer required
 - This was keeping the data from being shared with partnering organizations and with the general public
 - Anticipate the Governor signing the bill soon
 - 2) Retire Executive Orders [01.01.2012.04](#) and [01.01.2012.18](#)

- Want to clean up and make Council on Open Data the only Council
- Would eliminate the MD iMAP Executive Committee and the Open Data Workgroup as separate entities
- Leave only the Council on Open Data which has many of the same members as the other groups
- Not practical to have all the required meetings for all the existing groups/organizations
- Since January, no new developments on retiring the Executive Orders

3) Online Open Meetings Calendar

- Meeting on register now, but this is a paper based environment
- Hope to work with Secretary of State to have online calendar on government site which rolls up all open meetings in one place, which people could subscribe to
- Idea is proceeding slowly, but DoIT and StateStat will meet with the Secretary of State's Office to discuss further

4) Statewide Data Inventory

- Ambitious, working on tool to allow agencies to list databases, log in
- Hope to roll out over course of this year and operational before next annual report is due

5) Open Data and Geospatial Portal Synchronization

- Current existence of two data portals, two searches
 - One for alpha numeric data and one for geographic data
- Looking to federate the searches, talking with Socrata and Esri
- Might set up a third portal, which would include functions from both that we want to retain, merge the searches into one of the portals
- Have some of this working, but there are things that are getting lost in the transfer, such as metadata, field aliases/names
 - Still working to improve so we do not lose fidelity of the data during the data transfer

- PIA was talked about a lot last year, but no recommendations were set forth in annual report

StateStat and Its Relation to Open Data (Kevin Conroy, StateStat Director)

- StateStat has been a valuable tool in the past, want to still use it as performance management tool in the future
- Will change how things will be done, using open data portal to collect data, rather than templates coming to StateStat analysts
- Want to be able to perform simple queries using the available portal tools
 - Man hours that were used to create reports in the past were heavy, want to streamline this process
- Group will be renamed in the near future and focus will change

- Now that the budget session is over we hope to be Governor's Office of Performance Improvement, head by Lieutenant Governor and Kevin Conroy who will report to the Lieutenant Governor
- Goals of this Governor's administration are coming and are going to change what will be tracked, which will be discussed with the individual agencies
- Will work with agencies who are using old technology, implement use of new systems and hardware, in conjunction with DoIT
 - IT consolidation to be done to save money and be more efficient
 - Doing spot visits now and seeing green screen computers, which is not necessarily the most efficient technology
 - Improve systems such as the handling of checks, budget constraints will require implementing new systems in a smart and budget efficient way
- There are some agencies who believe they might feel as if they are being bullied by initiatives from this office, hope this is not the opinion moving forward
- Want to improve the customer service experience, average citizen of Maryland, we want them to have a much more positive experience
- Once rebranded, we will work with agencies about templates, ask for opinions from the agencies about what is valuable to track and what is not
- Intend to bring on about 7 more staff, mix of Auditors, Analysts, etc.
- Open data: Lieutenant Governor wants citizens to be able to look at data in an easy format
 - For Example: Potholes in their area, submit a simple query to get results of location and status
 - Make the more user friendly experience

Question: What is the timing of when meetings will start?

- Need to hire more staff, determining number of pins expected to receive
- Late May/June is time frame to get the office functioning
 - Requires training, since these are unique positions than found in most agencies or states
 - There are not many states that are running an office like this one

Question: What is the vision for working with counties? Many have open data programs already in place.

- Work together, especially when locals have data that the state does not have
- Reciprocal relationship anticipated, there are some rules that will restrict this relationship, but want to work together
 - For Example: Heroin Task Force
 - Trying to collect information from health agencies, most coming from the state
 - Some of the state agencies are trying to get data from the locals
 - This is a top priority of the Governor
 - He is relying on counties and StateStat to collect this data and share with the citizens
 - Once the Governor comes down with a directive, expect to see all hands on deck and immediate response to get good data and make it available to all appropriate parties
- Many times the data that is received is old, want to work to receive data sooner, identify problems, with new open data rules, moving forward this can be successful

- Any ideas of a better way to do things, please bring this to the attention of the StateStat staff, who are open to these ideas

Question: Is there a mechanism where citizens can provide input to the open data portals?

- Citizens can interact with the site by providing feedback, but for uploading we would be paying for cloud hosting, we buy capacity of Socrata which is allocated amongst agencies
- Validation process would need to be in place, some of the best information is from the public
- Something to be looked into, data could be submitted to the agencies
- There are feedback mechanisms in place on the system, ways to comment, no way to upload new datasets, can create views and queries of what is already on the system

Question: The expected process moving forward is to start with the templates and meetings and then uploaded and then analyze data?

- Want to create new templates over the next month
- Data will be uploaded to the open data portal
- Expect some overlap with the templates from the last administration, but some aspects will be new
- Some data in the past was cumbersome to collect and did not happen until the last minute, want to improve this process

Automating StateStat Templates (Josh Exler, Senior Analyst for Open Data, DoIT GIO Office)

Advantages of Accessing StateStat Data Through the Open Data Portal

- Discuss change in how data will be extracted, put into template and how would it get to the portal, do not want to change agency workflow too much
- Envision as GOPI (StateStat) is asking for new templates that support comes from DoIT
 - Both spatial and non-spatial support personnel:
 - GOPI Analyst support
 - Josh Exler and additional GIO support for data transfer determination
- Josh Exler would only come to initial meeting and then Analysts would take over support moving forward

Status with SHA and DHMH

- Taking data out of templates, putting into a format that can go into Socrata and automate the upload process
- DHMH has been the test subject, understanding what workflows exist and how the data is available for submission, take advantage of existing workflows and submission processes
- Discussion more direct data submissions for the future
- Do not want to ask to change processes already in place

Technical Considerations as Template Data Is Added, Removed and Changed

- Writing ETL scripts, automate analysis and correlations are being run on the data as part of this process

- After data is loaded into Portal, graphs, charts and other visualizations are available, data coming from GOPI templates can quickly be rendered as a chart
- Many automatic analyses can be provided for all data that is submitted to the system
- Plan to have one graph for each metric for the templates, all will understand the data visualization being applied and analyzed before, during and after meeting
- These will not need to be regenerated each month in a manual fashion, but will instead be consistent and automated
- Available in the Open Data Council memo: all analysis that will occur each month
 - Open to discussion/recommendations, the scripts can be modified based on need
- Included in the legend, default color is grey, data that are 1, 2 or 3 standard deviations from the mean are those columns which are colored (light green = 1 Standard Deviation, medium green = 2SD, dark green = 3SDs or more)
 - Can easily allow for visualization from month to month as these values change
- There are one, two or three asterisks based on correlations, levels of significance, noted at bottom of graphs

Long Term Plans (Automatic Graphing, Dashboards and Analysis)

- Copy the data out of the templates and paste into files that are Socrata readable, created Excel files, which could be a new data template, columns for data, analysis and correlation
- Maintainable in the long run, agencies add data directly to the Socrata readable templates, more sustainable, might want to shift to in the future
- Loading data correctly into the open data portal is the next focus
- Scripts point to particular cells and pull data out
- Loop in DoIT and GOPI if you want to add new metrics, but can just hide the row if you want to remove metrics within the template
 - Scripts need to change as a result of changes to the data structures in the agency spreadsheets
- Each template is different, unique cases with the data, but are working to acknowledge and make considerations for each
- For some of the data will have to sit and think how best to represent it and pull out of templates, most of the time, can just point, pull the data and copy to data.maryland.gov

Question (DHMH): This is good for those who are not as familiar working with data, and correlation analysis is positive addition, but data is not always stable month to month, can the data be kept as annual for testing platform?

- Response: Are the monthly and annual count being reported separately? DHMH: Some monthly and some annual, also quarterly, great to look at counts month to month, but significance would not be valid month to month
- Response: Could exclude some analysis for individual agencies, but might not want to submit month to month data, this would then not be included in the analysis for monthly GOPI meetings
- Some automatic analysis could be valuable for monthly data, some are not as valuable for month to month

Question: Can you change significance?

- Perhaps allow for .05 rather than .1
- Can be changed individually, based on discussion with individual agencies

Question: Is every dataset going to be a bar graph?

- Column charts would be the best case for most of the data
- There are other visualizations available, most of the major graph types this would work best
- Can also do stacked charts
- Both would cover most of the requirements for GOPI analysts and meetings
- Would be available immediately and would allow for immediate review to assure that the data is accurate and being rendered appropriately
- These capabilities are out-of-the-box options
 - There are other options that will be available, such as can be embedded in websites and do not have to send everyone to the central portals for information
- Trying to determine best way to create a dashboard for an agency, allow for showing multiple graphs and charts in one visual environment, rather than one chart at a time

Question: How many years data will be displayed?

- Once hit certain number of data points, have to scroll left and right (horizontal scroll) to display additional years of data

Question: How will monthly data be valuable for 10 years worth of data?

- Charts can be setup to start with a specific month, only need to review recent months, but would also be able to access years' worth of data by month

Question: Is this aggregate data coming in or is data aggregated by the system?

- Raw data is submitted, copying raw data out of system

Question: How will the statistics be clarified for the general public?

- Can accomplish significant analysis for GOPI and separately, share more basic information for general public consumption, but this might not be able to be accomplished with the same graph
- There are hovering windows which provide additional details
- Can develop separate graphs to show significant analysis and general information, but this will double the number of graphs in the system and therefore make data.maryland.gov more difficult to browse and navigate
- DoIT will provide introductory information on how to interpret the analyses and graphs, allowing for those coming in to educate themselves to interpret the graphs accurately

Question: StateStat templates are very cumbersome, lots of data that is not used, mostly operational data, rather than customer oriented data, who determines what the graphs should be, do not need to graph all the data that is in the system, will we go through and determine what is important to the administration and isolate interested data?

- Have to get Governor's strategic goals first and then determine what will be discussed at the upcoming meetings
- Agree that a lot of data that was collected was never discussed at past meetings, want to limit the extra work of collecting unnecessary data
- GOPI plans to bring agencies in one-by-one and determine what is valuable to discuss during meetings and what is valuable to add to the open data portal
- Data does not have to include a graph, sometimes not helpful, will be determined case by case
- GOPI's request of agencies: Keep doing what we have been doing, so we do not have to ask for retroactive collection of data once the new goals and paths are determined

Question: With passage of Open Data Act and data elements required for federal grant recording, has discussion taken place around federal funding in Maryland tied to these reports?

- Some of this is currently reported in the templates, but not all, would be valuable to add additional information
- DoIT met with Mr. Hollister and discussed a few years out until this hits us at the state level, important to keep a close eye on, but not an immediate change
- More discussion is necessary

Question: Is GOPI to work with Management For Results (MFR) so duplication is minimized?

- Data collected annually
- Want to work together, but annual data collection is not often enough for the purposes of GOPI
- Agencies might collect monthly or quarterly throughout the year and then provide an annual aggregation to fulfill MFR
- Might be able to ease the workload by using GOPI data to fulfill MFR requirements
- DoIT and DBM are analyzing what is being collected for GOPI and MFR, and allow what is being collected by GOPI to become what is being submitted to MFR

Comment: DHMH has high profile data that is being put out, testing statistical significance, putting out some reports soon, data should match what is in the data portal, putting out the same message

- DHMH uses a little bit different testing, but will need to compare the two sources
- Attach data to the metadata of data portal, referencing one from the other
- Might not be testing the same data and correlations, if they are the same, want to make sure they match

Comment: From agency side, not overwhelming amount of work to transfer to new process, more on Josh Exler's side, keep templates up-to-date and clean, challenge will be for Josh Exler to make changes to the template for each agency moving forward

- Delegate Bonnie Cullison has joined us and helped to pass the bill changes
- Several groups came forward to support the submitted bill (Maryland State Geographic Information Committee (MSGIC), Maryland Society of Surveyors (MSS), several A&E firms, Common Cause), great to see significant support of the bill

Data Freshness Dashboard (Josh Exler)

- Is the data still fresh or does it need to be changed out?
- Not live online yet, wanted to run by Council before allowing
- Looking at date of last update for the data on the portal, compared to how often data is supposed to be updated, determine if the data has been updated recently enough
- Pulling the metadata in real time, but anticipate a static report each morning to be made available. Real time reporting will be practical once new Socrata backend launches later this year and API calls speed up by approximately two orders of magnitude
- Currently 350 datasets (approximately) on data.maryland.gov. All are included in the dashboard
- Metadata field (update frequency) was added as a custom piece of metadata, included for most datasets
- Options: “Daily”, “Weekly”, “Quarterly”, “Monthly”, “Annually”, “As needed”, “Bi-annually”
 - May remove “As needed”
 - May add “Static data cut”, which means the data will not be updated again in the future
 - *April 28th 2015 Update: “Static Data Cut”, “Biweekly”, “Semi-annually”, “Other” added as options. When choosing “Other”, a dataset owner must type in additional details about the precise update frequency.*
- Brings to a single location the status of all datasets
- Can search by different categories (owner name, users, agencies), adding tags and keywords field to allow for better sorting data and grouping
- Monitor this to determine what datasets are not being updated, which might result in deletion of the data, will be brought to the council before data is deleted permanently
- Do not want to continue to provide outdated data into the system
- Is this going to be public? Yes, but the name of the dataset owner will not be made public, not the intent to publicly shame individuals

Question: Can you vote for a dataset to be updated?

- This would be an effective mechanism to generate public feedback about desired data updates. If a specific dataset were to receive many vote, this would apply pressure on the data owner to update the dataset

Question: How will the decision be made for those that have used “As needed” in the past?

- “As needed” is a catch-all right now, have to add an additional field with specific information typed in, to allow for "every 10 days" for example
- The agency or the owner of the data determines the schedule of update
- Josh Exler will help monitoring what is listed versus how often the data are actually being updated

Question: Who determines the best update frequency?

- Agencies own the data and therefore the agency would decide, otherwise GOPI could determine if a different frequency should be practiced
- Would not want to have a date in the title, e.g., “2012 Vehicle Collisions”, want to have "Vehicle Collisions" as a consolidated dataset then append new data each year

- All data is in one dataset for easy access, also saves money because there would be a single dataset charge and not individual datasets for individual years

Question: When will the dashboard go live?

- No date right now - Plan to make live to Council a couple of months before it goes live to the public
- Allows Council members and agencies to submit refreshed data before released to public
- New Socrata backend will speed dashboard load time significantly and therefore make internal and external products more usable

Question: Future meetings will include what is the stalest data in the system?

- Anticipate creating aggregate visualization, e.g., pie chart showing datasets up to date vs. datasets out of date, rather than looking at individual datasets at each meeting

Question: Would it make sense to have a subcommittee to assist with this endeavour?

- If there is still a struggle to get data updated by the agencies perhaps, can discuss data curation, data updating
- First step: Making agencies aware of the freshness of the data and allowing them time to take action

Automating Large Dataset Publications (Josh Exler)

SDAT Example

- Pulling data from two agencies and using a join to bring data together into a single consolidated dataset
 - Specific example for property records
 - 283 data fields in the dataset, includes all fields from SDAT and MDP
 - Several million records, hitting the maximum of how much data can be included in a single dataset
 - Considered breaking out the data by county, but advantages of consolidating the data, e.g., cross-county mapping and querying, outweigh the performance hit based on the size of the dataset
- Anticipate some performance enhancements coming from Socrata in the near future
- Data being pulled from a server at SDAT
 - 2,000 character string with all of the data fields being represented
- Software transforming string of characters into a list of the data fields, aliases of the data field names and attributes of the property for each data field
- Data dictionaries have been applied and field names and attributes have been automatically translated
- Property data for all parcels in the state, in the past SDAT has published it for purchase which was updated monthly or could receive annually in GIS format
- GIS format will still lag because there will time required to clear up geographic changes
- Valuable dataset moving forward to have this information on the portal

- Can combine data from two agencies (SDAT and MDP in this case) to make the output more than the sum of its parts
- Viewing every single property in a single data table in one place, standard Socrata functionality available: search, filter, query, map and visualize and graph data, will include latitude/longitude data
- Can comment on specific points, fields, etc., embed data into a local website
- Should have first few counties information available in the next couple of weeks, sending to SDAT and MDP for final sign off shortly (Caroline, Kent and Anne Arundel Counties included in first cut)
- Can collaborate between agencies on the portal rather than putting out duplicate data
- Use similar process to take data internal to an agency, load the data into Socrata through automated processes
- DoIT is available as a resource for some of the larger datasets or more complex datasets to help load data into the site, if there are thoughts on what might be considered for loading

Question: What is the purpose of publishing this dataset?

- Very popular data, more hits than any other dataset in the state through a web interface
- Get a million hits a week through the web interface
- 2.3 million accounts
- The maps on the website themselves get 150,000 hits a month

Question: How do you protect sensitive information from being released?

- There are instances where PIA would need to be requested in the past, but this makes the data available to the public without having to go through PIA barrier
- Anticipate removing the owner names from the dataset
 - Owner names WILL be available however -- in an external location. Each record includes a link to Real Property Search (SDAT's existing web interface to view some of the data), which includes owner name and shows one property at a time.
 - Users will therefore be able to see property owner names one at a time, instead of bulk disclosure of all property owner names. Since data.maryland.gov data is indexed by Google, bulk disclosure would result in property owners' addresses listed prominently in Google search results.
 - The owner names are already public in Real Property Search. They do not fall under the PIA Act
 - There are some cases where the property owners needs to be protected, such as law enforcement personnel or individuals with protective orders.
 - Summary: Our tools will allow easy querying of who owns a specific property -- but they will not allow for the reverse query: seeing which properties a person or business owns

Question: Will there be a way to tag this information with polygon boundaries, for example legislative boundaries?

- County boundaries yes, already in system, could map it out with legislative districts
- Will have latitude/longitude, could map out the data using this information and overlay with known boundaries

- Can already do this now with property view data
- Looking to do this spatial join within the portal, the Socrata tools are not as robust as other geographic tools that are available
- Could pull data out and do the spatial join in an external system, such as Esri software and receive more up-to-date results
- Functionality is anticipated in the upcoming release of Socrata functionality for custom polygons

Socrata Functionality Improvements (Clara Zimmerman, Socrata)

- Tabular experience is the current view, some filtering, make charts, search data, built searches and filters can be embedded on other websites
- Rethinking the experience
- Currently requires a high interest and technical bar to use the tools within each of the buttons
- Understanding over the past year how to tell the story and how to best get answers to questions
- Currently in a technical preview stage, allowing some subscribers to test
- Branded "data lens", the original experience will not go away, just another option
- Metadata has been moved to the top (description, last updated date, etc.)
- Columns moved to the bottom of the page, a "card" experience
 - Card corresponds to a field and are filterable
 - Quickly drill down into the datasets through dashboard-type environment
 - Multiple tables which impact on one table will impact all visible tables (citizen experience)
- Publishers can create the experience that customers will have based on these new card designs, auto generate on page based on settings from publishers, tell a story with the data
- Access to become available within the next month

- Visualizations do not have to stay on open data site, can be embedded into an agencies page, hosted back end by Socrata
- ESRGC has dashboards being built and they have been building some customized dashboards and hope to continue to provide access to this resource through next year
 - These dashboards allow for multiple variables to be visualized
 - New functionality from Socrata allows for visualization of a single table (one-to-one relationship), cannot currently combine datasets

- Provide functionality to allow for matching points with boundaries in the future
- Added through API geoJSON built out for developers to query data, can do boundary queries with this data also

Question: How do you access old visualization?

- Once made generally available, determine by data owner to publish a "data lens" page or user experience can still go directly to the table/traditional view

Question: Is there a common description box?

- This functionality is coming
- Currently just pulls from metadata

- Can add comments to datasets is already available functionality and the data owners will be notified when a comment has been submitted
- Can also comment on a specific point within a dataset

Open Discussion

- DoIT recommended that agencies take advantage of the existing data portals to build a website with graphs or maps, and not to contract for a separate system, try to leverage this available system first
- All hosted in the cloud, already paid for and there is capacity on systems today

Question: Will there be Webex Meetings in the future?

- No, we only have four meetings annually and the meeting is too large (size of room) for Webex to accommodate

Comment: Consider “the agency” of the public

- Want to capture data coming during an event, tweets, how do we accommodate bringing in this data, which might prove valuable
- A lot of agencies capture this kind of public feedback already
 - For Example: MEMA captures this data when EOC is activated
- Have to be careful as government entities when this type of data is captured, must have a validation process
- Opening to unsolicited data could create unreliable data, which would not pass auditable standards and generate a culture of inaccurate and unreliable data
- There are processes that exist to capture public data
- There is also a legal issue, because we do not own Twitter and Facebook data
 - Do not know how we could avoid lawsuits because the data does not belong to the state or the data providers
- There is data out there that could be captured and then build around the instance of the data
- Existing contract using [Hootsuite](#), can contact web representative, go through and look at social media trends (emergency and criminal activity already being tracked)
 - Contact your agencies’ DoIT Web Team representative to learn more
- Difference between collecting and analyzing data streams
 - Collect and analyze at various agencies for various purposes
 - Not redistributing from state site or state resource
- Could capture feedback from the public through a hashtag

Future Meetings

Date: June 12, 2015 @ 10AM - 12PM

Location: Maryland Department of Transportation, Harry Hughes Conference Center

Date: September 3, 2015 @ 10AM - 12PM

Location: Maryland Department of Transportation, Harry Hughes Conference Center

Date: November 5, 2015 @ 1 - 3PM

Location: Maryland Department of Transportation, Harry Hughes Conference Center