



**MARTIN O'MALLEY**  
GOVERNOR

STATE HOUSE  
100 STATE CIRCLE  
ANNAPOLIS, MARYLAND 21401-1925  
(410) 974-3901  
(TOLL FREE) 1-800-811-8336

TTY USERS CALL VIA MD RELAY

**MEMORANDUM:**

**December 18, 2014**

**To:** Governor Martin O'Malley  
Anthony Brown, Lieutenant Governor  
John Griffin, Chief of Staff  
Eric Beane, Deputy Chief of Staff  
Isabel Fitzgerald, Secretary, Department of Information Technology  
Gregory Urban, Deputy Secretary, Department of Information Technology  
Mike Powell, Chief Information Officer  
Matt Power, StateStat Director  
Eloise Foster, Secretary, Department of Budget and Management  
Elizabeth Harris, Chief Legal Counsel  
Members of the Open Data Council

**From:** Barney Krucoff, Geographic Information Officer  
Joshua Exler, StateStat

**Subject:** Open Data Council Executive Briefing Memorandum

**\*\*Please note the following dates and times for future meetings\*\***

February 25, 2015, 10:00 am – 12:00 pm, at Maryland Department of Transportation, Harry Hughes Conference Center.

For this Executive Briefing Memorandum, the following issues are highlighted:

- Draft Council Report to Legislature & Discussion (Barney Krucoff)
- Selling Data vs. Open Data. MDP Data Usage Statistics (Jim Cannistra)
- Progress on Data Inventory (Josh Exler)
- DoIT News Items (Barney Krucoff)
  - iMap 1.0 decommissioning
  - Transparency site data update
- Data Desired from Local Governments (Josh Exler)
- Open Discussion

- **Draft Council Report to Legislature & Discussion (Barney Krucoff).** As part of the Open Data Act passed and signed during the 2014 Legislative Session, on or before January 10th of each year, the Council on Open Data is required to report to the Governor and the General Assembly, in accordance with § 10-1401 through 10-1404 of the State Government Article. On December 16<sup>th</sup> DoIT submitted a draft report to the Council, requesting that the Council members and participating agencies review the draft before today's meeting. The draft report is also included as an attachment to this meeting's memo. The draft memo summarizes the current status of Open Data initiatives and websites in Maryland, including the Open Data Portal (<http://data.maryland.gov>), iMap (<http://imap.maryland.gov>), DBM's [Maryland Funding Accountability & Transparency](#) site, and [MDP's website](#). The draft memo also summarizes the topics discussed to-date through Open Data Council meetings and presents recommendations for ongoing future initiatives, building off the work of the previous Council meetings.

Questions for Council consideration:

- What other priorities do the agencies and Council members see for statewide Open Data initiatives which are not yet included in the draft report?
- Does the Council approve of the draft language for modifications to the Automated Mapping GIS law?

- **Selling Data vs. Open Data. MDP Data Usage (Jim Cannistra)**

Questions for Council consideration:

- Has the surge in demand resulted in increased costs to MDP?
- To what extent does MDP believe that the surge in demand will be sustained? Could the surge be due to pent up demand or publicity, or is a long term trend?

- **Progress on Data Inventory (Josh Exler).** Using feedback received from the previous Open Data Council meeting on what items should and should not be included in a statewide data inventory, StateStat has drafted a framework for such an inventory, to be hosted on the Open Data Portal. The Council should note that the inventory framework database on the Portal will not be a public dataset until the Council has approved the items to be included for each database. The draft report puts forth tentative dates of June 30<sup>th</sup> 2015 to complete pilot data inventories in a few select agencies, then December 31<sup>st</sup> 2015 for the completion of the statewide data inventory. The current items included in the inventory framework are listed below.
  - **Unique ID.** Each dataset which is inventoried should contain a unique ID. The primary purpose of this is to save agencies time and effort when they upload new datasets. After the data inventory begins, any inventoried dataset with a unique ID will be able to, upon import of the data itself to the Open Data Portal, automatically populate the required metadata – using the Socrata Publisher and Socrata Consumer APIs. This way, agencies will not need to duplicate effort by writing the metadata in the data inventory, then again in the metadata upon data import. StateStat and DoIT plan to begin writing the necessary scripts to do this after the data inventory begins.

- **Dataset Title.** This will require that the Council define a standard nomenclature for new datasets, making sure to include the agency name and time period which the data cover.
- **Can this be public, open data?** This field is the crux of the data inventory. A dataset's owners will decide, using guidelines in the process of being developed by the Council, whether a dataset can be Open Data. Any dataset containing immutable Personally Identifying Information (PII) and/or with legal restrictions will not be able to be published as Open Data. However, the inventory should still list these datasets, so that the inventory is an exhaustive survey of statewide data, with records showing why and why not the agencies have decided to make their datasets public.
- **Already published as open data? (Link)** Between the Open Data Portal and iMap, large portions of agencies' data are already published. This field will contain direct links to these data sources.
- **Already published elsewhere, e.g., PDF? (Link)** State agencies are collectively publishing even more data in the form of PDFs, Excel spreadsheets, and other forms of static, less publicly accessible reports – than they are publishing Open Data. This is a consequence of state transparency initiatives predating Open Data initiatives. Owners of these data and reports should strive to put them in Open Data formats, but in the mean time, the existing resources should be linked to within the inventory.
- **Category (Dropdown).** Dataset category is one of the pre-existing metadata fields within the Open Data Portal and is also a metadata field on iMap. These categories are show in the table below. This information will be used to automatically assign a dataset a category upon upload to the Open Data Portal.

Data Categories on the Open Data Portal and iMap					
Category	Open Data Portal?	iMap?	Category	Open Data Portal?	iMap?
Administrative	√		Historic		√
Agriculture	√	√	Housing	√	
Biota		√	Hydrology		√
Boundaries		√	Imagery		√
Budget	√		Location		√
Business/Economy	√	√	Planning	√	√
Demographics	√	√	Public Safety	√	√
Education	√		Society		√
Elevation		√	Structure		√
Environment/Energy	√	√	Transportation	√	√
Geoscientific		√	Utility		√
Health/Human Services	√	√	Weather		√

- **State Agency (Dropdown).** All agencies participating in the Open Data Council and/or with presences on the Open Data Portal and/or on iMap will be available as dropdown options.
- **Contact Name(s).** Each dataset should have at least one named primary point of contact.

- **Contact Email Address(es)**. These points of contact should have contact information available in the form of email addresses. This field as well as the latter field could be either public or private, depending on the concerns of state agencies and Council members.
- **Aggregate Data or Individual Records? (Dropdown)** At the previous Open Data Council meeting, agencies expressed the desire for an inventory to contain this field, to show whether available data have individual records (e.g., single crimes, single permits) available, or if the information was available only at the aggregate level. The three current dropdown options are:
  - Dataset contains aggregate data
  - Dataset contains individual records
  - Individual records available, but only aggregate data can be published
- **Contains PII? (Dropdown)**. Personally identifying information should not be published in any datasets. If PII cannot be scrubbed from a data source, then the data should not be published. The three dropdown options are Yes, No, and Unsure.
- **Has Legal Restrictions? (Dropdown)**. Similarly, data with legal restrictions should not be published. The three dropdown options are Yes, No, and Unsure.
- **Summary of Caveats and Barriers to Making Data Public**. This is a text field which should summarize any PII and legal concerns, as well as any other barriers to making the data public. Limitations on data quality and completeness should be noted here as well, e.g., any missing or corrupted data.
- **Source Database (Private)**. For the internal agency use as well as reference by DoIT and StateStat, the name of the source database and any additional relevant information about the location of the source data should be listed. This information does not have public utility and barring objections from the Council, this information will remain private.
- **Time Period of Content**. When did data tracking begin? Do data continue through present day? This is a metadata field in Open Data Portal datasets, and will be automatically transferred from the inventory to a dataset's metadata upon dataset upload. This field and the next field are both text fields.
- **Are there any gaps in the time period of content?** If the data source is not continuous, it should be noted in this field.
- **Update Frequency (Dropdown)**. An additional metadata field which will be transferred automatically upon dataset upload. The dropdown options are Continually, Daily, Weekly, Monthly, Quarterly, and Annually, mirroring the options in the Open Data Portal metadata. Agencies will be expected to update their data in the Portal with the same frequency that the data are updated in a source database.
- **Subjective Data Quality (Private)**. Finally, using a scale of one starts to five stars, agencies will be asked to rate the completeness and quality of the source data. If the Council believes that this information would be better captured through a qualitative summary in a text field, the Subjective Data Quality field can be retooled.

Questions for Council consideration:

- What additional fields are required in a dataset inventory?
- An agency volunteering as a pilot agency for the dataset inventory would receive significant assistance from DoIT in completing the inventory, compared to agencies joining after the pilot. Do any agencies want to volunteer as pilot inventory participants?
- What concerns do agencies and Council members have with this draft inventory framework?
- **DoIT News Item: iMap 1.0 Decommissioning (Barney Krucoff).** On December 31<sup>st</sup>, the original iMap site, “iMap 1.0,” will be permanently decommissioned, leaving iMap 2.0 in its place. Any agencies with any mapping resources still hosted at *mdimap.us* will need to instead point to resources hosted at *geodata.md.gov*. The latter domain is for the iMap 2.0 site.

Questions for Council consideration:

- Do any agencies require any assistance migrating mapping resources to iMap 2.0?
- **DoIT News Item: Transparency Site Data Update**
- **Data Desired from Local Governments (Josh Exler).** Socrata, the vendor for the Open Data Portal, is now hosting monthly phone calls for the primary Open Data contacts within DoIT and StateStat, as well as Baltimore City, Prince George’s County, and Montgomery County, all of which have Open Data sites hosted through Socrata. This workgroup is called the Maryland User Group. During a December 16<sup>th</sup> call, StateStat and DoIT presented a draft list of county data to which the state would like access. Please note that this is a draft list and is a starting point only for data requests of these counties.

<b>Draft List: County Data Desired by the State</b>
○ Impervious surface data (geographic data)
○ Building permits
○ Certificates of occupancy
○ 311 data (Priority: emergency services for example trees down)
○ 911 data (Whatever is not legally restricted)
○ County and municipal PD crime data -- Currently the state's access to these data are through data aggregated at the end of the year by the Maryland State Police for Uniform Crime Reports (UCRs). More real-time and flexible crime data feeds would be extremely beneficial.
○ Counties' budget data
○ County jail average daily populations
○ County/local agency/department offices and service locations – The Maryland Department of Information Technology has lists and maps already, but counties' assistance would be useful keeping the lists up to date and filling in whatever gaps may exist.

In addition, Montgomery County presented a draft list for state data which it desires:

Draft List: State Data Desired by Montgomery County
○ SDAT property tax data (real property, home owner credit, personal property, new construction)
○ School data
○ Traffic and pedestrian collision/injury/fatality data
○ Grants data
○ Uniform Crime Report
○ Consolidated Transportation Plan
○ Election data

The Maryland User Group decided that as a starting point for data integration between the state and county level, budget and crime data are among the highest priorities. Prince George’s County publishes its [crime data](#) to its [Open Data website](#) using the [SpotCrime Open Data Standard](#) (SOCS). If the Maryland State Police were to publish crime data to the Open Data Portal using this standard as well, then direct data integration between county and state data sources would become possible. A list of data fields required for SOCS, copied in full from the SpotCrime website<sup>1</sup>, is show in the table below.

Questions for Council consideration:

- What additional data do Council members and state agencies desire at the county level, particularly from Prince George’s County and Montgomery County, as well as from Baltimore City?
- What data formats does MSP currently use for its internal records of individual crimes? Are these records close to matching SOCS?

SpotCrime Open Data Standard (SOCS): Copied from <a href="http://blog.spotcrime.com/2014/03/the-spotcrime-open-crime-data-standard.html">http://blog.spotcrime.com/2014/03/the-spotcrime-open-crime-data-standard.html</a>				
Data type	Required	Permitted Values	Examples	Details
Area	Yes	Text	Baltimore County, MD	Defines the city/county/jurisdiction and state
Date	Yes	ISO 8601; YYYY-MM-DD	2014-03-10	Date occurred. Allows easy identifier and separation of data. Date displayed in mm/dd/yyyy on SpotCrime.
Time	Yes	ISO 8601; UTC and 24 hour	23:11	Time occurred. Allows easy identifier and separation of data. Displayed in 12 hour format on SpotCrime.
Incident Type	Yes	Text	Aggravated assault	Allows icons to be assigned and decipher between crime types. SpotCrime dictionary categorizes each incident into 1 of 9 icons (listed below).
Location	Yes	Street number	322 York Road	Full and block addresses are accepted.

<sup>1</sup> <http://blog.spotcrime.com/2014/03/the-spotcrime-open-crime-data-standard.html>

		and street name.		Needed for geo-coding accuracy. All full addresses are masked to the block level on SpotCrime.
City	Yes	Text	Towson, MD	Needed for geo-coding accuracy.
Narrative	Yes	Text	A victim was stabbed in the abdomen by an unknown subject after a verbal altercation	Provides more description and information on incident type (commercial burglary or residential burglary). Displayed in the details section on SpotCrime.
Latitude, Longitude	Yes	Geographic coordinate system	39.399262, -76.602990	Geo-coding accuracy. Not displayed by SpotCrime.
Case Number	Yes	Text and number accepted	2014-031023111	Helps identify specific incidents. Displayed in the details section on SpotCrime.
Incident Number (if different from case number)	Optional, highly recommended	Text and number accepted		Helps identify specific incidents. Displayed in the details section on SpotCrime.
Beat/ Sector	Optional	Text and number accepted	6A	Displayed in the details section on SpotCrime.
District/ Precinct	Optional	Text and number accepted	Towson	Displayed in the details section on SpotCrime.
ZIP Code	Optional	US and international postal codes accepted	21204	Helps with geo-coding accuracy. Displayed in the details section on SpotCrime.

- **Open Discussion**