

# Baltimore City Open Data Report 2019



To: Members of Baltimore City Council  
 From: Michael Wisniewski, BCIT CDO  
 Date: 2019-July-10  
 Subj: Yearly Report on Open Baltimore Site (Final Report Pursuant to CCB—16-0615)

1	Overview .....	1
2	Assessment & Prescriptive Analysis.....	2
2.1	Progress Toward Goals of the Open Data Program .....	2
2.2	Assessment of Agency Compliance.....	3
2.3	List of Datasets on the Site .....	4
2.4	Long-Term & Shorter-Term Ongoing Improvement .....	4
2.5	Staff Recommendation – Adjusted from 2018 Request .....	5
3	Appendix A – Miscellaneous Views.....	6
4	Appendix B – List of Datasets & Maps .....	8

## 1 Overview

The historic background for the open data initiative and the Open Baltimore Site (Site) is noted in Appendix A. We are in our 8<sup>th</sup> year with the Site. Each year, this report is provided to Council at June end.

For this Baltimore City Open Data Report – 2019, there are 4 specified reporting topics (Appendix A notes these topics). Beyond this, we also review some aspects of how the Site currently functions, what’s working, what’s not – and how some key foundational initiatives should help the Site slowly, methodically improve in the next year.

Note that we are finalizing our Civic Analytics Strategic Plan (Plan) - Baltimore’s first ever civic analytics strategic plan (due July-end). This Plan addresses the broader, all-City view of analytics and data. The Plan is effectively operationalizing a part of Baltimore City IT’s (BCIT) broader Plan (Inclusive Digital Transformation Strategic Plan – City of Baltimore 2018-2023)<sup>1</sup>. Open data and the Site are also addressed within this overall analytics plan.

From this parent Plan, the data team’s objectives and workstreams for the Site are framed under 3 aims:

<sup>1</sup> <https://technology.baltimorecity.gov/sites/default/files/BCIT%20Strategic%20Plan%20Final%207.10.18.pdf>

1. Refine
2. Systematize
3. Creatively leverage

Example activities in these three buckets include:

**Refine:** fix existing site - existing content/data/meta-data; Ongoing – refine our business as usual (BAU) approach to be sustainable/feasible with limited team resources

**Systematize:** develop and institute data governance, automated data refreshes, expansion of datasets and genuine relational datasets, automated data dictionary, ...

**Creatively leverage:** expand the site’s usability to far broader community of users; expand beyond mere data posting – supporting innovations, proof-of-concepts, smart cities initiatives, ...

For 2019-Q3 thru 2020-Q1, emphasis is more with (1) and (2), and then we slowly evolve to a steady state BAU blend of 1,2, and 3.

## 2 Assessment & Prescriptive Analysis

Reporting on the 4 topics is as follows.

### 2.1 Progress Toward Goals of the Open Data Program

The prior status reports for 2017, 2018, as well as interviews with stakeholders paints a picture of good effort at publishing datasets but significant difficulty with sustainably keeping data up to date and useable on the Site. This is caused by 3 key factors:

1. Insufficient resources: We have limited resources, partial time of one FTE, to support the Site
2. Traditional data problems: Data and informatics for the City overall, in useable value-added form, is in bad shape, and the Site is merely reflects this.
3. Approach to date is labor intensive – rather than naturally aligning with internal needs

---

#### 1 - Insufficient resources:

We have a Chief Data Officer, as requested from the 2017 status report, albeit with turnover (3 CDO’s thus far) and periods of vacancy. A data-team of 4 contract workers comprises the team. The primary responsibility of the data team is to support the databases and data flows underlying software applications used by City staff (e.g. call-center software application supporting the 311-service). Secondly, the team helps agencies get access to data for various purposes and instate new application databases. The data team struggles to meet this primary task of critical database support. To tactically affect the Site, e.g. fix a dataset issue, team members take turns – stopping mission critical work to make such fixes.

Open Data Coordinators (ODCs) were identified, to help find and gather data. However, this operational model appears too labor intensive and prone to stall out. For example, four of the initial six ODCs from 2017 have either left City employment or moved to new, unrelated roles. And, the data-team lacks capacity even to simply co-ordinate recurring meetings or police data gathering

efforts. More critically, it is difficult or infeasible for Open Data Coordinators to really know data well and to prioritize such data work over their day to day primary roles.

Along with resource concerns, a skill concern also exists with ODCs. ODCs are not necessarily data experts. Data cleansing and transformation of data to useful form is an involved task, even for the data-team with formal training and expertise in this area. The ODC's cannot easily gauge if the data is correct, accurate, and useable or not. It is more of an IT specialization to understand systems, applications, ongoing changes to business processes and consequent changes to data, etc.

## **2 - Traditional data challenges:**

Aside from being under-resourced, our problems are exacerbated by the fact that the City's data and systems are not easily accessible or well governed. Separate from our open data focus, departmental staff are often hard pressed to get data and utilize it to support existing internal operations and decisioning. That is, they struggle to get data even for their own basic needs.

## **3 – Approach is labor intensive:**

Some datasets are manually refreshed, which is labor intensive. Meta-data management is also manual.

---

From these above factors, our progress toward open data goals needs improvement. The footprint of what we are attempting to keep up is too broad. Symptoms of the team having a footprint of responsibility exceeding available resources and reliance on manual tasks include:

- There is little informative narrative for Users on the site, and that which exists can easily become outdated.
- Lack of growth in dataset count, or breadth/depth of data; lack of directly readable, insightful content that a layperson/resident (non-data person) can use.
- Meta-data is lacking (Meta-data includes time-stamping of when a dataset is to be refreshed, was most recently refreshed; field definitions, explanations on how to use/interpret information, etc.
- Feedback mechanisms (e.g. feedback/Tweet feed – see Appendix screenshots) are not addressed. Why mislead Users to think these tools are a point for interaction, Q&A and such, when we lack resources to do so. Perhaps it is better to turn them off, rather than frustrate Users, and construct interaction mechanisms that are albeit less immediate but at least reliable.

The below section 2.4 addresses some workstreams to help improve overall management of open data.

## **2.2 Assessment of Agency Compliance**

Compliance is the same as in 2018. The main datasets are all automatically updating with pulls from operational databases. Manual (human intervention) uploads are inconsistent. Also, it is important to

note that fixed reference tables (e.g. list of council districts), posted once as if permanently fixed, need updating. No data is ever completely static.

Aside from minimizing manual upload work, we will instate an automated email notification to data owners to help with compliance. Fixed reference tables will also be updated.

In the bigger picture, data governance and compliance are a key part of the strategic Plan and will include the data management on the Site.

### 2.3 List of Datasets on the Site

Appendix B lists the official datasets on the Site – basically unchanged between 2018-Jun and 2019-Jun. The 2020 focus is on shoring up data quality and usability. This may involve some rationalization of datasets, possibly deleting bad data or merging datasets, so the count of datasets at 2020 may be fairly unchanged. The planned areas for new datasets are just (1) support tables for 311-data and (2) geographic information (GIS) reference tabling to make GIS consistent across datasets. All changes will be documented and published in a change log visible to users. All other activity, the vast majority of planned effort, falls into (1) foundational improvement, and (2) an RFP for the Site, as explained below.

### 2.4 Long-Term & Shorter-Term Ongoing Improvement

#### **Long Term:**

The long term, fundamental resolution to the state of the City's data is to build-out our Enterprise Data Warehouse (EDW) and Analytic Data Mart (ADM) capability as outlined in the Civic Analytics Plan. This is a key component in the forthcoming Plan. In a nutshell, data from operational systems is not structured for human use, but for the applications to exchange data. The EDW process is automated – taking data from operational databases, transforming it to useful form, and cumulating a rich history of data for longitudinal/trending analysis, statistical inquiry, and automated business intelligence (BI) reporting. The key is that data is contextualized with a data dictionary and controlled for quality/usability. Also, data promotion is fully automated so that ODC's can function more as data 'spotters' and advocates but not necessarily data munging experts.

Building out the EDW happens to provide a needed alignment of incentives and a net reduction of efforts:

- agencies gain from offloading non-core IT functions to BCIT
- agencies will want their data in the EDW for automated reporting, BI and analytics purposes, and cross-agency analysis
- BCIT can establish the unified, holistic data and IT governance so desperately needed by the City of Baltimore – and datasets across departments can be structured so they are relational

BCIT and the City's legal department can more easily manage the data privacy aspects of governance – ultimately increasing the breadth and rate at which data can be promoted.

We will still retain a plan for adding data that is separate from the EDW workstream. While the EDW buildout is indeed the correct long-term approach for the City, the effort will take time. And, it may not

be perfectly congruent with open data efforts. For example, data tabling prioritized by departments to run or improve operations can be incongruent with interests for data by external stakeholders – residents, taxpayers, corporations, commuters, etc.

### **Short Term:**

The primary task for CY 2019-Q3 thru 2020-Q1 is the continuation of our RFP for the Site software vendor. The existing vendor, Socrata, was selected as an initial proof-of-concept provider, but a formal RFP process is required to be in compliance with City standards. It is a stretch goal for the team to handle this RFP process within 2019.

### **Refine tasks – 2019-Q3/Q4:**

- Cleanup: Remove or archive datasets that are not maintained or are ‘fragmented’ snapshots
- Meta-data: make dataset sizes and update activity visible
- Generate automated email reminders to ODC’s doing manual data uploads
- Notice users that Twitter feed is only reviewed monthly for feedback
- Remove open Baltimore progress report webpage; Replace with status webpage noting tasks/work done that relates to open Baltimore and near term planned tasks.
- Clearly denote fixed reference tables; place such tables on yearly review/update schedule.
- Record changes in a site change log

### **Refine tasks – 2019-Q3 thru 2020-Q2:**

- Geographic information system (GIS) upgrading: WIP: There are internal needs to update GIS capabilities as part of ‘next generation 911’ work. These are federally mandated 911 upgrades that will also refine our addressing database functionality and accuracy. We are currently formulating GIS improvement plans that handle this 911 need, while also refining our base layer functionality, with subsequent impact on Site geo-data. This is a key 2020 focus to affect the Site’s datasets too. The exact plans for this GIS work will be posted to the Site.
- Example data dashboards: As noted, our 2020 focus is on fundamental fixing of processes to enable easier scaling of open data efforts. That said, we hope to allocate some capacity to exposing out some prototype dashboards to help highlight our future state with refined, systematic data management processes. The expected focus will be 311 data.

## **2.5 Staff Recommendation – Adjusted from 2018 Request**

The 2018 report suggested 2 resources were needed – (1) an Open Data Program Manager and (2) Open Data Services Engineer. Going forward, the suggestion for resources is subsumed in the Civic Analytics Plan. Essentially, 1-2 SQL BI-developers (business intelligence software developers) would have dramatic impact both on developing the City’s EDW and the Site.

### 3 Appendix A – Miscellaneous Views

#### History / Background of Baltimore City's open data initiative (same content as 2017, 2018 reports):

Baltimore is one of the first cities in the country to develop a sustained and comprehensive open data program that has been recognized nationally for its efforts. The City's Open Baltimore Portal ([data.baltimorecity.gov](http://data.baltimorecity.gov)) went live in its beta version in 2011, using the Socrata platform to publish and share data with online visualization tools. Open Baltimore has grown to become a critical resource for citizens and other stakeholders to gain access to municipal data. As the demand for open data continues to grow and policies governing data become more complex, the methodologies used to manage data must evolve to mitigate bottlenecks and promote data democracy without sacrificing data protections and privacy.

In 2016, Baltimore City codified the Open Data Program in City Council Bill #16-0615 that established the need for governance and implementation of city data to be published to Open Baltimore. As required by the legislation, this report serves as the City's 2017-2018 Annual Report and aims to satisfy the requirements of Section 9.9

#### Screenshot of Site's 'Discuss' function:

The screenshot shows the Open Baltimore website interface. At the top, there is a navigation bar with 'Home', 'Office of the Mayor', 'About', and 'Content'. Below the navigation bar is a search bar and a 'Sign In' button. The main content area displays a table with columns: #, Building#, Address, Neighborhood, PoliceDistrict, CouncilDist, and Location. The table contains 16 rows of data. To the right of the table is a 'Discuss' sidebar with a title 'Discuss' and a message: 'Looking for additional activity on this dataset? Click on More Views tab.' Below the message is an 'Add Comment' form and a list of comments. The first comment is from 'harika' and asks about the notice date for vacant buildings. The second comment is from 'dewjohnson8' and asks about vacant-to-values webpage data.

#	Building#	Address	Neighborhood	PoliceDistrict	CouncilDist	Location
8	1612 WESTWOOD	08/06/2017	Sandtown-Wil.	Western	7	(39.30542566, -76.644872)
0	1720 N CARROLL	08/06/2017	Sandtown-Wil.	Western	7	(39.30832648, -76.644235)
9	1630 APPLETON	08/06/2017	Easttown	Western	7	(39.30725397, -76.64913)
5	1710 PRESIDENTAL	08/06/2017	Sandtown-Wil.	Western	7	(39.30533547, -76.645132)
7	834 N GILMAN ST	08/06/2017	Sandtown-Wil.	Western	9	(39.29997047, -76.64234)
5	1827 RAYNER AVE	08/06/2017	Harlem Park	Western	9	(39.29682937, -76.64893)
8	927 S BENTLON	08/06/2017	Carmonton Hl.	Southwestern	9	(39.28982097, -76.651949)
6	907 S BENTLON	08/06/2017	Carmonton Hl.	Southwestern	9	(39.28131308, -76.65219)
8	1728 N CARROLL	08/06/2017	Oliver	Eastern	12	(39.31331297, -76.98863)
3	1113 GREENMOUNT	08/06/2017	Johnston Sq.	Eastern	12	(39.30322017, -76.66792)
1	2513 E CHASE ST	08/06/2017	Biddle Street	Eastern	13	(39.30326697, -76.58174)
1	607 N DECKER A.	08/06/2017	Ethwood Park	Southwestern	13	(39.28883497, -76.57427)
0	519 N BELMONT	08/06/2017	McElderry Park	Southwestern	13	(39.29750187, -76.57849)
7	2882 DULANY ST	08/06/2017	MtRill	Southwestern	9	(39.27450911, -76.57679)
4	306 SIMMONS ST	08/06/2017	Shiplay Hill	Southwestern	9	(39.28277301, -76.58891)
0	2840 HARLEM AVE	08/06/2017	Moshier	Western	9	(39.29501047, -76.64649)
1	2843 WALBROOK	08/06/2017	Walbrook	Southwestern	7	(39.30947867, -76.64702)

#### Screenshot – Tweet and Feedback on front page:

The screenshot shows the front page of the Open Baltimore website. On the left, there is a 'Feedback' form with fields for 'Name', 'Email Address', and 'Comments and Suggestions'. Below the form is a checkbox labeled 'Do you wish to be contacted?' and a 'Submit' button. On the right, there is a 'Tweets by @OpenBaltimore' section. The first tweet is from @OpenBaltimore and says: 'Interested in seeing recent 911 calls by location? Check out our new 911 map [gis.baltimorecity.gov/911/](http://gis.baltimorecity.gov/911/) based on 911 calls on Open Baltimore. This is just the start so let us know what else you'd like to see!' The second tweet is from @OpenBaltimore and says: 'Our 2018 Open Data Report is now posted on Open Baltimore! [baltimorecity.gov/files/baltimor...](http://baltimorecity.gov/files/baltimor...)' Below the tweets is a 'BPD Consent Decree Implementation Unit' logo.

## Screenshot – Track our progress view

**Open Baltimore Progress Report**

The data team works year round to improve Open Baltimore and assist agencies to currently in the middle of a comprehensive data inventory effort that will inform agencies Open Data Policy. We are Agencies are expected to complete their initial publishing plans before the end of Q4 2018.

**2018 Open Data Report Progress**  
October 18, 2018

- 26 Agencies/Divisions with datasets on Open Baltimore
- 35 City & Quasi-agencies (Mayor's Cabinet)
- 23 Agencies/Divisions engaged
- 24 Open Data Coordinators (ODC)
- 7 Inventories Complete
- 0 Publishing Plans Complete

**Data Inventory Progress**

Category	Count
Complete	7
Incomplete	24
Partially Complete	0
No Authoritative Data	0

## 4 Appendix B – List of Datasets & Maps

### List of Datasets & Maps

<u>Agency / Area</u>	<u>2018-Jun</u>	<u>2019-Jun</u>
Aging & Retirement Education	4	4
Department of Transportation	7	7
Health Department	1	1
Housing	4	4
Office of Promotion & the Arts	3	3
Board of Elections	1	1
Combined Charity Campaign for Baltimore City (CCC)	1	1
Enoch Pratt Free Library	1	1
Enterprise Geographic Information Services	4	4
Environmental Control Board	1	1
Finance	3	4
Fire Department	1	1
General Services	2	2
Human Services	2	2
Information Technology	1	2
Liquor License Board	2	2
Minority & Women-Owned Business Development	1	1
Office of the Mayor	14	14
Planning	4	4
Police Department	9	8
Public Works	1	1
Recreation & Parks	2	2
<b><u>Maps</u></b>	<b><u>2018-Jun</u></b>	<b><u>2019-Jun</u></b>
Department of Transportation	4	4
Enterprise Geographic Information Services	32	(~ 32)
General Services	1	1
Information Technology	4	4
Mayor's Office of Emergency Management	1	1
Office of the Mayor	1	1
Planning	5	5