

Chief Data Officer's Annual Report

March 12, 2019



Acknowledgments

Since last year's report, I was privileged to spend most of my work time serving as the Interim Chief Technology Officer of the District of Columbia, as well as Chief Data Officer. OCTO is a unique and special agency, with a breadth of responsibility that runs from complex statewide applications for sophisticated customers like the Department of Motor Vehicles to hyper-local hands-on tech support in the public schools. That acknowledged, I'd like to thank Mayor Muriel Bowser for trusting me to run the agency that has shaped much of my career, and thanks also to our newly appointed CTO, Lindsey Parker, for diving right into data and open government efforts and supporting our team.

Thanks also to Agency Data Officers and Agency General Counsels for their hard work gathering and preparing the Enterprise Dataset Inventory for publication and to the Open Government Advisory Group for their advice and helping us stay on task.

Finally, special thanks to the staff and contractors of the OCTO Data Team, who really stepped up and filled in while I was Interim CTO.

Barney Krucoff

Chief Data Officer
District of Columbia
Office of the Chief Technology Officer

Office of the Chief Technology Officer

Enterprise Dataset Inventory, Data Curation and Analysis

Michael Bentivegna
Program Manager, Data Visualization and Analysis

Eva Reid
Senior Analyst and Enterprise Data Inventory Lead

Mario Field
Program Manager, Data Curation

Peter Casey
Senior Data Scientist

Open Data Portal Initiatives Page

Alexandre Santos
GIS Program Analyst

Introduction

In June 2018, Mayor Bowser issued Mayor’s Order 2018-050,¹ aligning the publication of the District’s Enterprise Dataset Inventory and this annual report with the national observance of Sunshine Week. Sunshine Week is a national initiative spearheaded by the American Society of News Editors to educate the public about the importance of open government. As Chief Data Officer, it is my job to help the District of Columbia government realize the greatest value from the data it collects and manages. Per Mayor’s Order 2017-115, the District of Columbia Data Policy, it firmly remains “the policy of the District Government that the greatest value of data assets is realized when freely shared to the extent consistent with the protection of safety, privacy, and security.”²

While the policy balances openness and security, it prioritizes transparency. The policy states that “enterprise datasets shall be **open by default**, meaning that their existence will be publicly acknowledged, and further, if enterprise datasets are not shared, an explanation for restricting access will be publicly provided.” In other words, “open by default” means that the District will publicly acknowledge all enterprise datasets, which is accomplished by publishing the annual Enterprise Dataset Inventory (EDI), although some may be labeled as confidential.³

¹ <https://www.dcregs.dc.gov/Common/NoticeDetail.aspx?noticeId=N0071235>

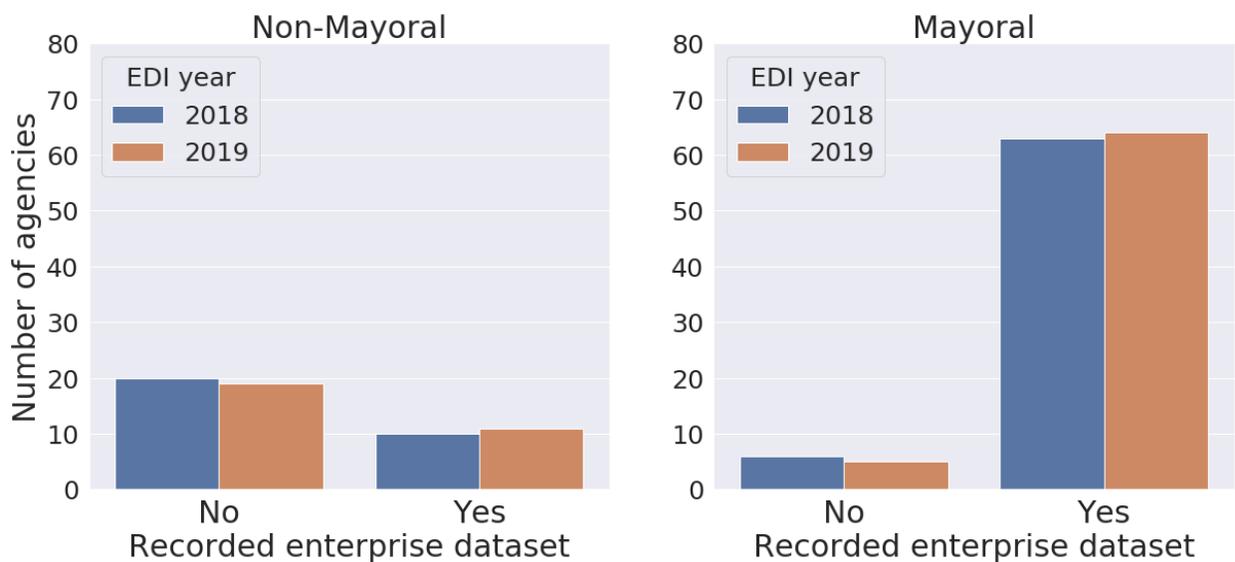
² <https://www.dcregs.dc.gov/Common/NoticeDetail.aspx?noticeId=N0063620>

³ <https://opendata.dc.gov/datasets/enterprise-dataset-inventory/data>

Enterprise Dataset Inventory (EDI)

The Data Policy mandates that public bodies in the District government create and maintain an EDI following the leadership of the Office of the Chief Technology Officer. The inventory requires agencies under the direct authority of the Mayor to record any “enterprise dataset,” which is “a dataset that directly supports the mission of one or more public bodies.” The Data Policy also requests that independent District government agencies, not under the Mayor’s authority, participate in the EDI, though their participation is not required. What follows is an analysis of the metadata generated by the EDI. This metadata is available through the city’s Open Data Portal.

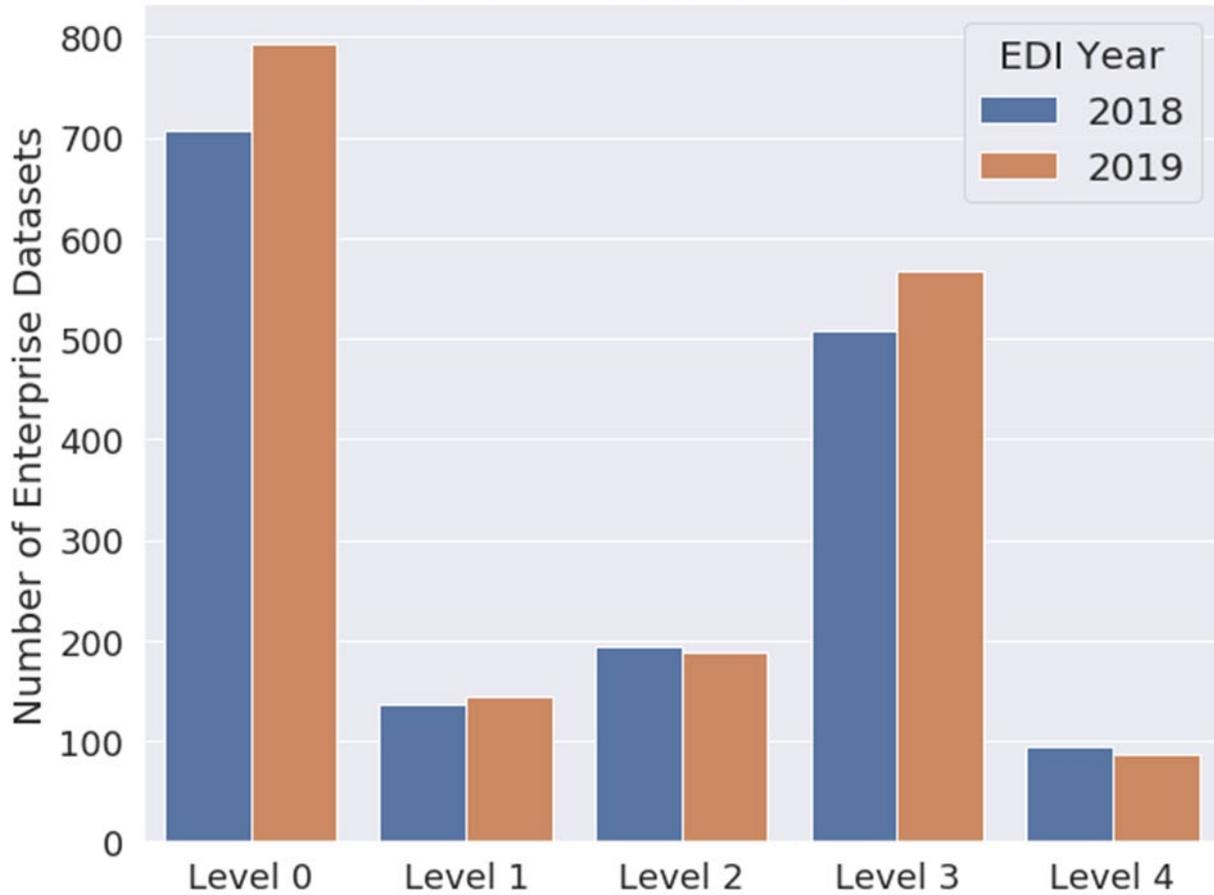
Figure 1
How many agencies recorded enterprise datasets?



According to figure 1, Seventy-five (75) agencies recorded 1,779 enterprise datasets in 2019, up from 1,640 datasets recorded by 73 agencies in 2018.⁴ Of the 69 mayoral agencies, 64 recorded an enterprise dataset, up from 63 in 2018. Two more independent agencies also contributed datasets, increasing the number of participating independent agencies to 11 from 10 the previous year.

⁴ In the 2018 report, an agency “participated” if they responded to the inventory and either recorded an enterprise dataset or reported they did not have an enterprise dataset record, as was the case for some small offices. In 2019, we are only reporting agencies that recorded an enterprise dataset. In 2018, FEMA was incorrectly counted among participating agencies. FEMA data is provided in the Open Data Portal and is properly attributed to OCTO.

Figure 3
How were datasets classified?



The DC Data Policy also lays out a **dataset classification** system to help determine which enterprise datasets should be open to the public and which should not be proactively released. Figure 3 shows how agencies classified their datasets.

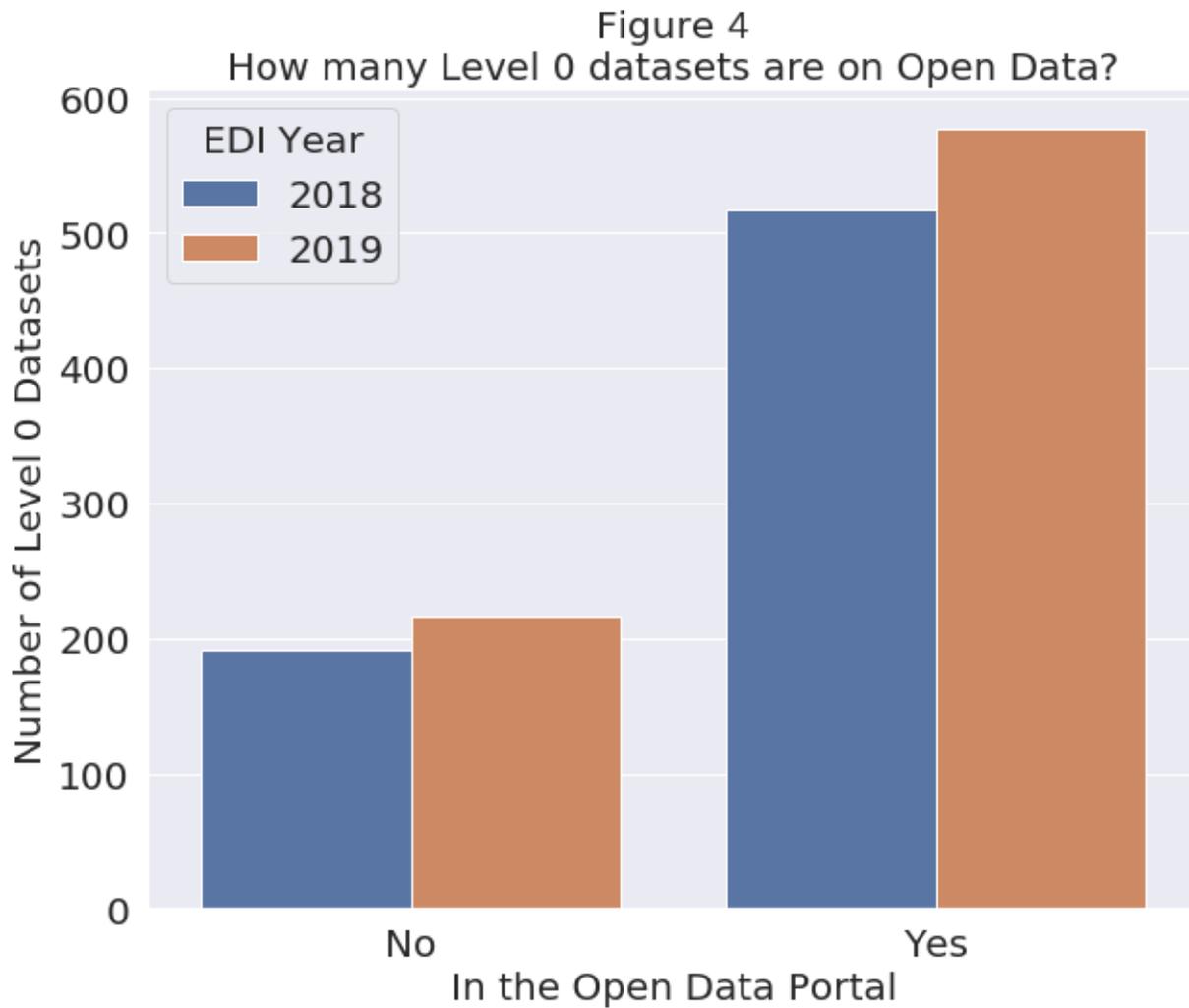
Level 0 is Open. Level 0 data is any data that is open to the public and should be proactively released. This is the default classification for the EDI and applies to any dataset that agencies do not determine to have a higher security level. In 2019, 793 enterprise datasets were classified as Open, up 85 from 708 the previous year. Open is the most common classification in the EDI.

Level 1 is Public but Not Proactively Released. Level 1 data is not protected from public disclosure but is not proactively published because of concerns over safety, privacy, security, or legal concerns. The number of datasets classified Level 1 increased slightly in 2019 to 144 from 137 the previous year.

Level 2 data is For District Government Use. Level 2 data is “subject to one or more FOIA exemptions, [but] is not highly sensitive and may be distributed within the District government.” The number of datasets classified Level 2 decreased to 188 in 2019 from 193 the previous year.

Level 3 data is **Confidential**. This includes data that is “protected from disclosure by law” and that is either highly sensitive or legally restricted from disclosure to other public bodies. In 2019, 567 datasets were classified Confidential, up 59 from 508 the previous year. Confidential is the second most common classification in the EDI.

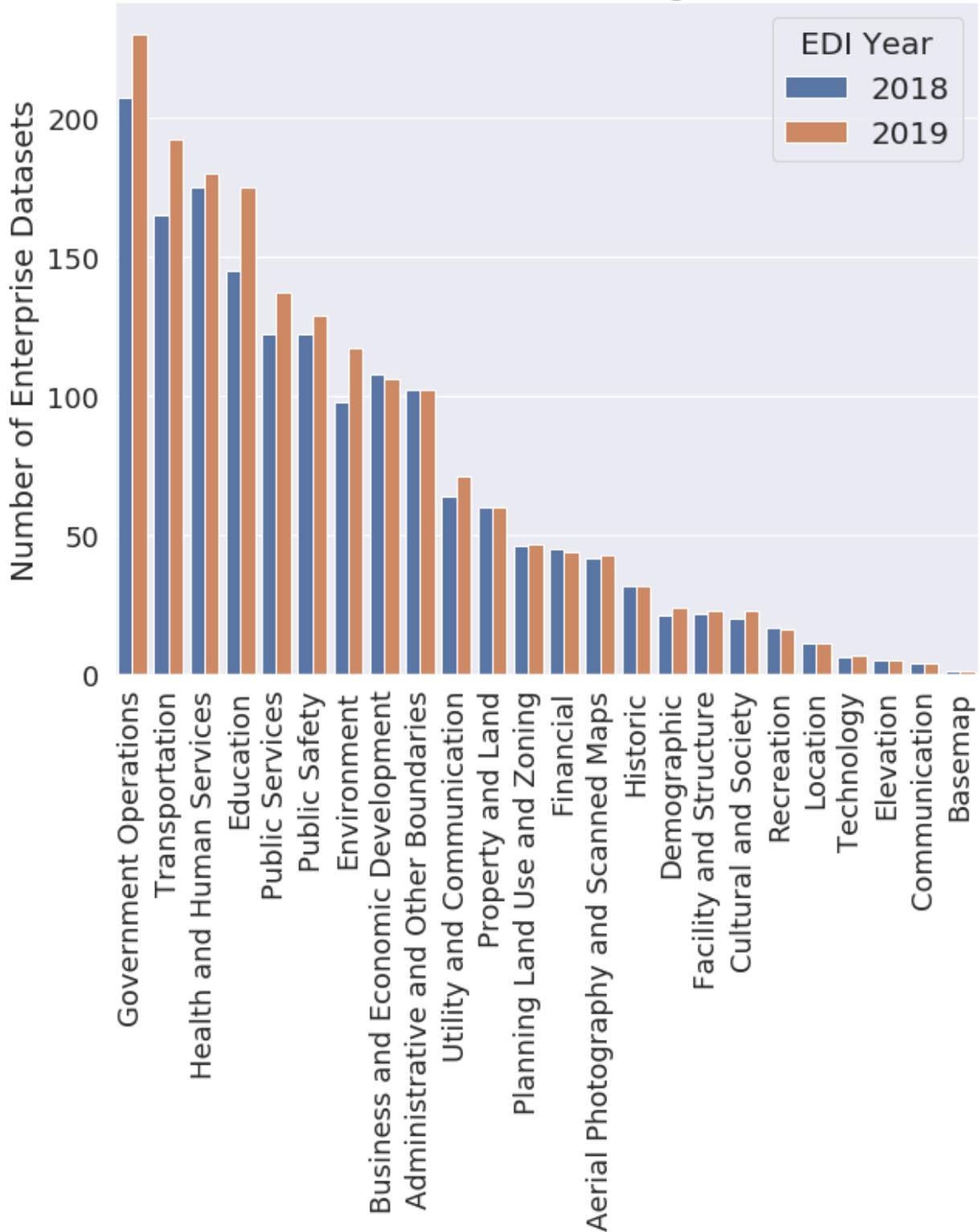
The rarest classification in the EDI is Level 4, **Restricted Confidential**. This refers to datasets for which “unauthorized disclosure could potentially cause major damage or injury, including death ... or otherwise significantly impair the ability of the agency to perform its statutory functions.” The number of datasets classified Restricted Confidential decreased to 87 in 2019 from 94 the previous year.



Of the 794 Level 0 datasets recorded in the 2019 EDI, 577 are publicly-available in the Open Data Portal,⁵ up from 517 the previous year. The EDI also identified 216 Level 0 datasets that are not in the Open Data Portal, an increase of 21 from 195 the previous year. OCTO will work with agencies, the Open Government Advisory Group, and the community to prioritize posting these remaining Level 0 datasets identified on the Open Data Portal.

⁵ <https://opendata.dc.gov/>

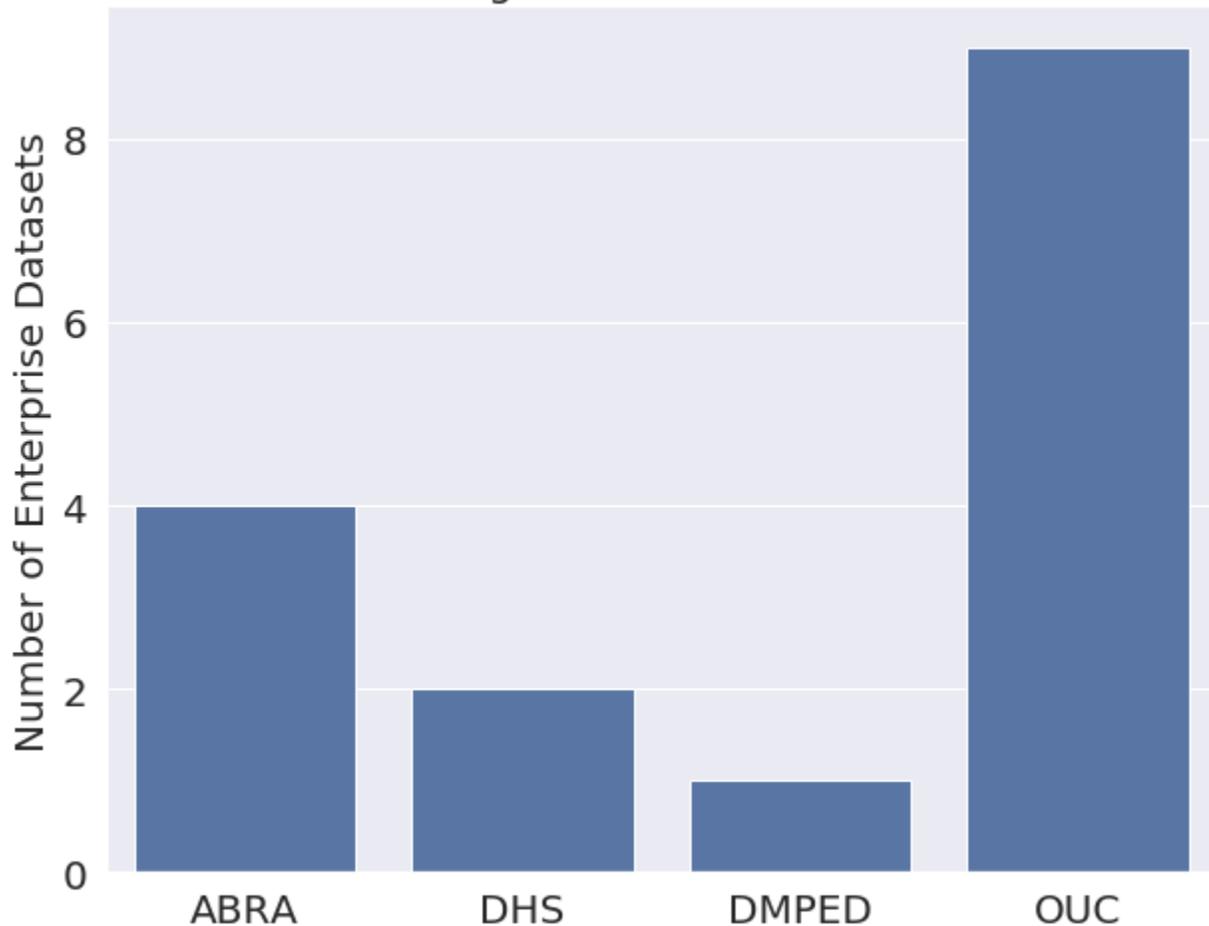
Figure 5
How were datasets categorized?



Agencies were also asked to categorize their datasets according to their contents and purpose. Figure 5 shows how agencies categorized their datasets. The number of enterprise datasets increased in a

number of categories, especially Education (+30), Transportation (+27), Government Operations (+23), Environment (+19), and Public Services (+15). The number of datasets only decreased slightly in a few categories, including Recreation (-1), Financial (-1), and Business and Economic Development (-2).

Figure 6
Which agencies retired datasets?



The 2019 EDI was the DC government’s second Enterprise Dataset Inventory and the first opportunity agencies had to retire datasets. Datasets are retired if they are no longer in use for agencies’ operations, duplicated, or otherwise improperly recorded. In 2019, 16 enterprise datasets were retired from the previous universe.⁶ The Office of Unified Communications (OUC) retired 9 datasets and made a net contribution of 11 datasets to the 2019 EDI. Similarly, the Department of Human Services retired 2 datasets and made a net contribution of 8 datasets. ABRA also retired 4 datasets and DMPED retired 1 dataset.

⁶ <https://opendata.dc.gov/datasets/enterprise-dataset-inventory-retired-datasets/data>

Figure 7
How were retired datasets classified?

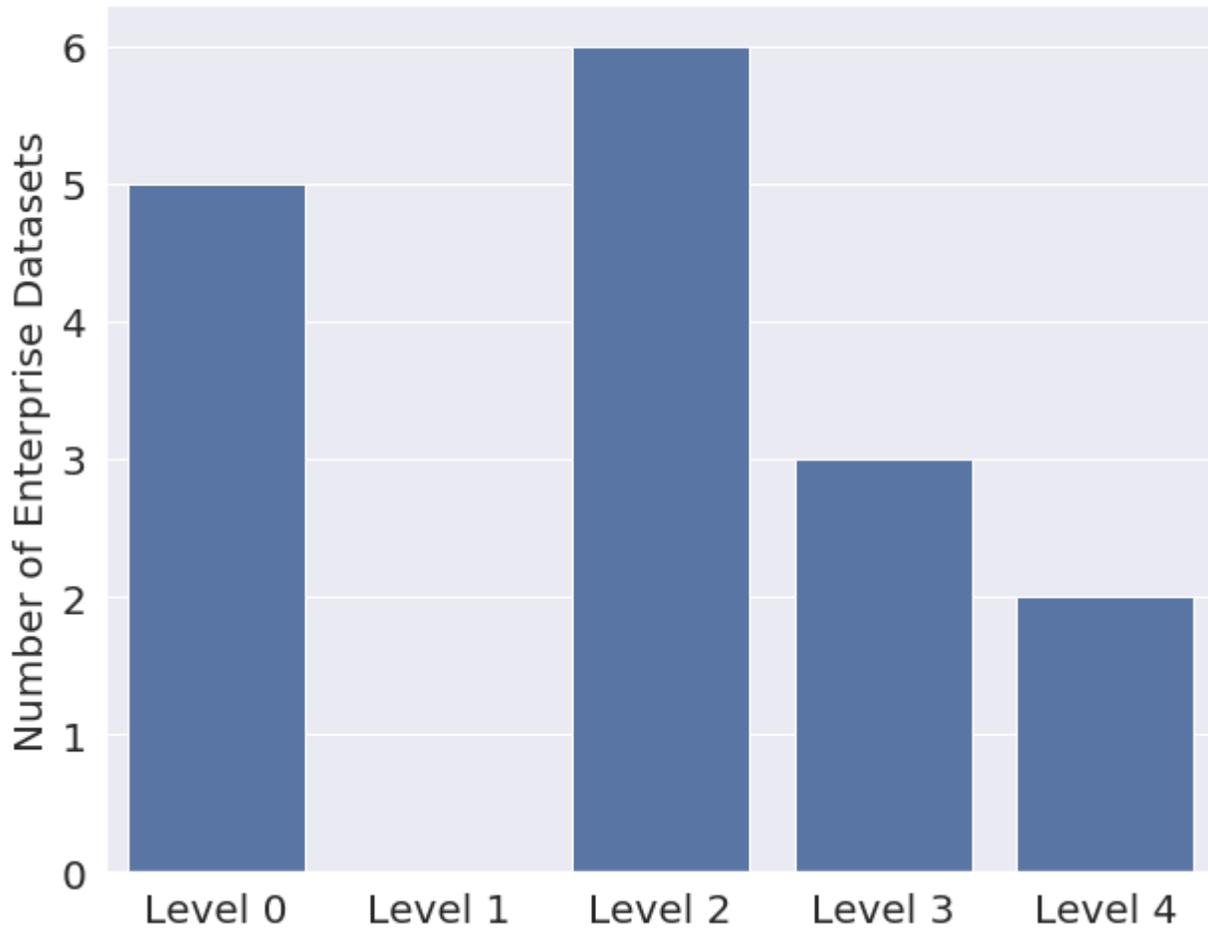


Figure 7 shows the classifications of the 16 datasets that were retired from the EDI in 2019. Five (5) of the datasets were classified Level 0, Open, only 1 of which was on the Open Data Portal. Another 6 datasets were categorized Level 2, For Government Use Only. Three (3) of the datasets were categorized Level 3, Confidential, and 2 were categorized Level 4, Restricted Confidential. None of the retired datasets were classified Level 1, Public Not Proactively Shared.

Establishing and Analyzing a FOIA-Open Data Feedback Loop

The Data Policy envisioned a “nexus between FOIA and level zero open datasets, where FOIA and Open Data are distinct but complementary practices.” The data policy holds that:

- FOIA request-tracking data should inform public bodies about public demand for open data. In fact, the Data Policy mandates use of the system for mayoral agencies
- Open data publication of FOIA request-tracking data can help residents hold public bodies accountable for the timely and consistent processing of requests
- Successful appeals for datasets previously denied under FOIA exemptions can inform public bodies about potential errors in dataset classification.

The above benchmarks are challenging, unless District agencies track FOIA requests in a way that is reasonably complete and uniform. Fortunately, the District has invested in an enterprise system to do just that. The system is known by its commercial brand name, FOIAXpress. The FOIAXpress system aids FOIA Officers in tracking, processing, and reporting on their work. Moreover, DC’s version of FOIAXpress includes a “Public Access Link” to facilitate online, rather than postal-based, interaction with requestors. I don’t know of any other state-level government that has invested in such a government-wide system. In almost every state and many localities, such public information law tracking systems, if they exist at all, are in service at the agency level, not the government-wide level.

The system enables managers to know what exemptions are being cited. Is the District meeting the prescribed time periods for response described in the law? Particularly for reporting purposes, the system must not only be used but also be used consistently and by all agencies. In calendar year 2018, OCTO put forth a serious effort to drive utilization of the system up. OCTO used to charge agencies that needed new FOIAXpress accounts. That pushed agencies out of the system. OCTO negotiated a better deal with the vendor and no longer charges agencies. We have also provided FOIAXpress training to over 100 FOIA Officers and related staff. The results, as shown in Table 1, have been tangible.

Table 1

FOIAXpress utilization measures by calendar year	Calendar year 2017	Calendar year 2018
# of requests and appeals tracked in FOIAXpress	7,346	8,120
# of agencies enrolled in FOIAXpress	53	55
# of agencies processing at least one request in system	51	52
# of current active DC government users	112	141

Despite the success described above, the number of FOIA requests is increasing even faster than use of the FOIAXpress system. As a result, the share of requests processed outside FOIAXpress increased in 2018, see Table 2. Obviously, more work is required to increase the percentage of requests tracked inside the system.

Table 2

FOIA Requests	FY 2017	FY 2018	Increase
Total reported by EOM	8,274	10,450	26%
Tracked in FOIAXpress System	6,762	7,943	17%
% in FOIAXpress	82%	76%	

As CDO, I read through many FY2018 FOIA requests where “structured data” was specifically requested and the data granted in whole or in part. I also reviewed the annual FOIA Appeals Report. I was looking for instances where a requester was granted access to structured data not otherwise available on opendata.dc.gov. Several such datasets were identified by the civic hackers of Code for DC at their “FOIA Party” on March 11, 2018. The datasets below should be reviewed for publication on opendata.dc.gov. Many of these are available publicly in some form, but not on opendata.dc.gov.

- Arrests (similar level of detail to what was published in 2016),⁷ MPD

⁷ <https://opendata.dc.gov/datasets/felony-arrest-charges-in-2016>

- Condom distribution statistics, DCPS, DOH
- Dockless bikes and scooters, DDOT
- Facility surveys, PCSB, but likely DGS and DCPS as well
- Health inspections, DOH
- Health professional licenses, DOH
- Moving violations by DC-owned vehicles, DMV
- Property tax credits and penalties, OCFO
- Seizure data, MPD
- Vacant property determinations, DCRA

Often the opposite problem is also true, meaning the public persists in filing FOIA requests for data that are available on opendata.dc.gov. These datasets seem to be good candidates for better education on how to use the open data or improved user interfaces that point to open data.

- Building permits, DCRA
- Business licenses, DCRA
- Underground storage tanks and similar, DDOE

Beyond the District government’s utilization of the FOIAXpress system, there are frustrations for citizens and businesses that attempt to use the “Public Access Link.” Both the public users’ experience and the government’s utilization calls for further investigation and action are addressed in the goals section of this report.

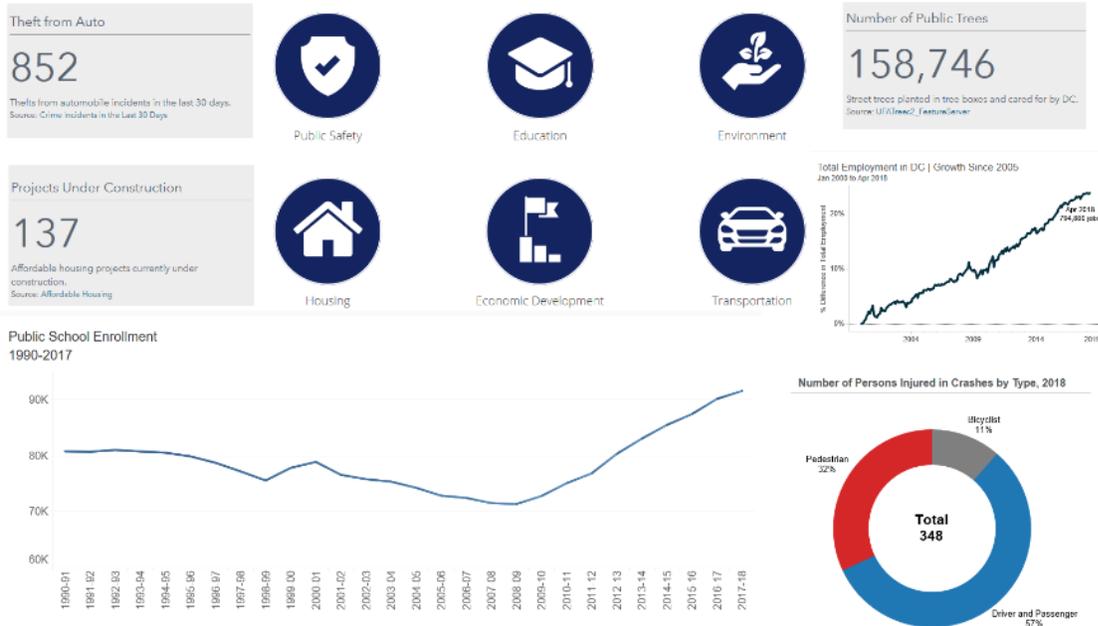
OCTO Data Team and Notable Agency Accomplishments

Here are some of the accomplishments of the OCTO Data Team and some of the agencies we work with.

Mayor.dc.gov

Last year in this report we touted a significant upgrade to open data portal including “initiative pages,” which make it easier for agencies to use data to tell their stories. What is really exciting this year is which agencies have been taking advantage of that capability. Mayor Bowser has long had a commitment to data-driven government, and now she is using the capabilities of opendata.dc.gov to help support her Fair Shot agenda. The Mayor has set an example for agencies to follow, and it should result in more datasets becoming open and more opportunities for the residents to monitor progress and hold their government accountable.

Mayor.dc.gov By The Numbers



Open data stories on <https://mayor.dc.gov>

Publication of the DC GIS Archives

The DC GIS program and our progenitor program, WGIS, led by the National Capital Planning Commission, has been collecting photogrammetric data of the District since 1995, beginning with aerial orthoimagery.⁸ The first vector (line work) basemap was created from 1999 aerial photographs.⁹

This year our Data Curation Team cataloged and published the DC GIS archives as open data. This allows the public to visualize and analyze the District of Columbia over time and see the state of DC to the present day.



DC GIS building footprints in Columbia Heights in 1999 (left) and 2017 (right). New buildings annotated with year they were built.

In the mid-1990s, like many urban areas at the time, the District was struggling compared with today. The WGIS and DC GIS Programs helped inform the development that has taken place, and we've mapped it all.

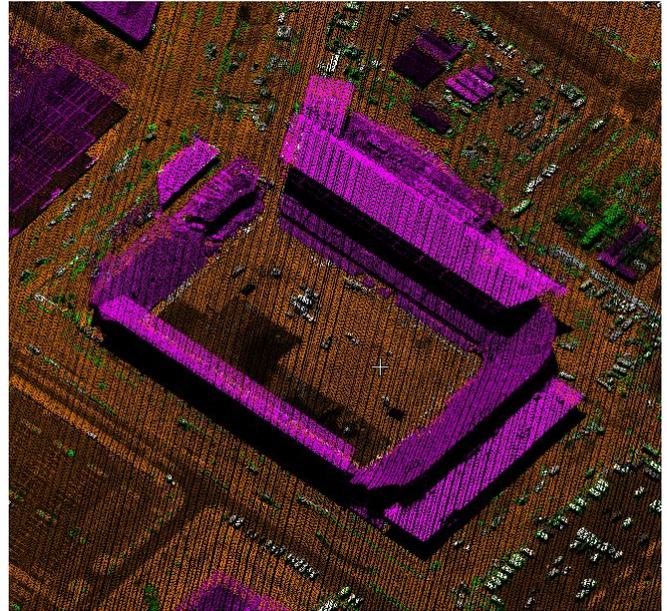
⁸ <https://opendata.dc.gov/search?q=orthophoto>

⁹ <https://opendata.dc.gov/datasets?q=1999>

LiDAR 2018

DC GIS has also adapted its aerial data collection to include new sensors and capabilities. LiDAR (Light Detection and Ranging) is a remote sensing method that uses light in the form of a pulsed laser to measure variable distances to the District's surface from airborne survey. The data collected is known as a point cloud dataset, which includes x, y, and z coordinates with additional details that can help classify landscape features.

The District of Columbia published new LiDAR captured on April 5, 2018. The previous LiDAR effort was done in 2015. The 2018 data differs from 2015's data because, for the first time, it includes vegetation classifications, building classifications, and mass elevation points. Other published products from this successful project include a Digital Surface Model (DSM), Digital Terrain Model (DTM), and Normalized Digital Surface Model (nDSM), all available as 1-meter resolution along with 0.6 meter topography. DC GIS plans to continue its partnership with Amazon Web Services to make the LiDAR Data publicly available. The data will be published as soon as some legalities are in place.



Audi Field Intensity of LiDAR point cloud

Data Analysis and Visualization Training for DC Agencies and Employees

OCTO's goal is to remove as many barriers as possible for DC agencies and employees that want to be data driven in perusing their missions. Through our Data Visualization and Analysis Team, OCTO has software licensing and training in place to make it easy for DC agencies to use modern data analyses and visualization tools. DC GIS has long offered training in the Esri ArcGIS line of mapping software. In 2018, OCTO expanded our training offerings to include the MicroStrategy and Tableau lines of Business Intelligence software. Our classes for DC employees are free of charge; the registration process can be found on the DC intranet.¹⁰

¹⁰ [ArcGIS Training](#), [Tableau Training](#), [MicroStrategy Training](#), (links only work from inside DC Network)

Publicly Available Business Intelligence Dashboards

Last year at this time, the OCTO Data Team was busy improving the internal server infrastructure for Business Intelligence. Now both Tableau and MicroStrategy have public-facing servers as well. Agencies like the Office of Planning (OP) can share their data analysis with the public more easily and graphically. For example, the adjacent population pyramid hosted on OP's DC State Data Center Visualization Portal¹¹ website shows the predominance of millennials in Ward 6.



The Office of Planning's <https://dcdataviz.dc.gov> allows the public to explore the District's demographics.



Julie Kanzler, Eva Reid, and Joy Whitt with their awards 4/23/2018.

The Top Women in Data

The District prides itself on being “The Capital of Inclusive Innovation.” In fact, in 2018, Washington, DC, was the top city for women in technology, according to SmartAsset.¹² As you would therefore expect, the nominees for the 2018 FemTech Award constituted a very competitive field.

The 2018 FemTech Award winners for Data were Julie Kanzler, OCTO; Eva Reid, OCTO; and Joy Whitt, Deputy Mayor for Planning and Economic Development (DMPED) on detail to OCTO.¹³

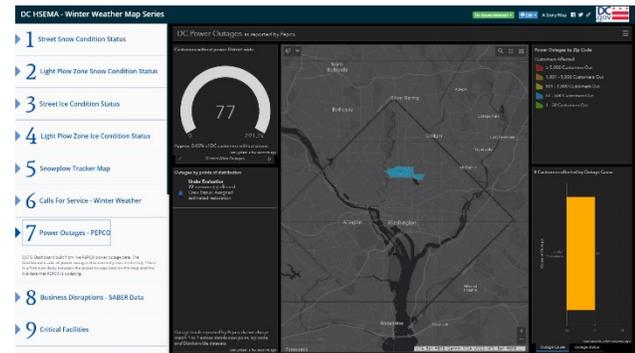
¹¹ <https://dcdataviz.dc.gov/>

¹² <https://smartasset.com/mortgage/best-cities-for-women-in-tech-2018>.

¹³ <http://dcfemtech.com/awards-2018-media>

Public Safety Portal

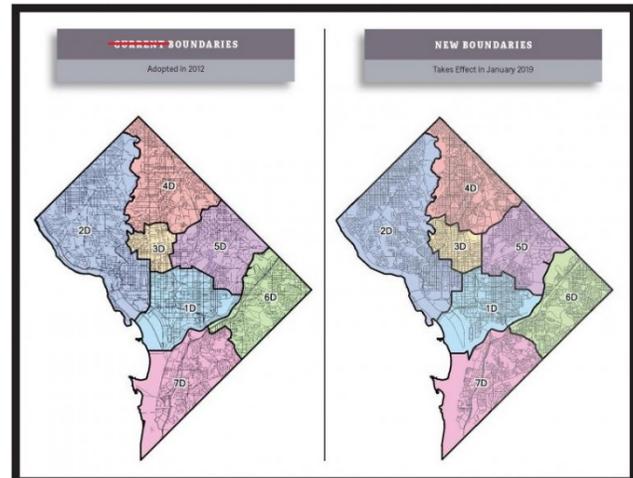
In 2018, OCTO's Data Team successfully deployed the District of Columbia Public Safety Portal for ArcGIS. This application enables DC public safety agencies to quickly create and share maps and data in a secure environment. It continues to modernize the "map making" and "dashboarding" process to meet increasing needs for sharing secure geospatial content internally throughout the District. Since its launch, District public safety agencies have deployed numerous applications via this portal. A few notable applications include a Winter Weather Map Series, a mapping application to monitor power outages, Hurricane Florence planning maps, and evacuation and emergency service maps.



Winter Map Series in the New Public Safety GIS Portal.

Metropolitan Police Department (MPD) Redistricting

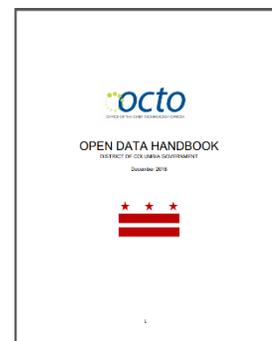
On January 10, 2019, the MPD, with assistance from DC GIS, launched new police boundaries. This was a significant data-driven effort that helps the MPD adjust its deployments to the District's growing population. The new boundaries are intended to make the delivery of police services more efficient. Behind-the-scenes data work includes matching historic crime data to the new boundaries. The new boundaries for Police Service Areas, Sectors, and Districts are available to the public on opendata.dc.gov.¹⁴



Old and new 2019 MPD District Boundaries.

Publication of the District of Columbia Open Data Handbook

Late in 2017, when the District became one of the first cities to be certified as a Data-Driven Local Government by What Works Cities, our data program scored



¹⁴ <https://opendata.dc.gov/search?q=police%20area>

very highly.¹⁵ However, we had one glaring shortcoming—the lack of a published data submission guide. We had been one of the first cities with an open data program and hadn’t written down the rules. Thanks to our Data Curation Team, we now have a published *Open Data Handbook*.¹⁶

The open data handbook contains three parts: a data submission process, set of data rules, and DC’s metadata standard. While this handbook applies to open data directly, it can also be applied any other dataset including ones with security requirements.

DC Data Lake Implementation

While we are very proud of the District’s open data program, sharing more sensitive datasets within the District government has proven more difficult. Data security, data discovery, data volume, and agency trust are among the challenges. Last year, the OCTO Data Team committed to develop a cutting-edge data platform to support analytics. As a result, we have developed the DC Data Lake, a secure big data system that serves as a central data-sharing hub for agencies and the Lab @ DC. The Data Lake is coming soon to city staff at data.in.dc.gov.

The Data Lake is built on a mix of commercial, open source, and custom software platforms including: Cloudera Hadoop, Elasticsearch, Jupyter, Apache Kafka, and Spark. The system also includes connections for the District’s Tableau, MicroStrategy, and ArcGIS so DC employees can continue to use their favorite data visualization tools. The DC Data Lake brings these capabilities to the DC government:

- **Ability to handle big data**, making high volume data accessible and analyzable to government data scientists and analysts. For example, Internet-of-Things sensor data .
- **Sandboxes for combining data sources** from different agencies, so that we can ask and answer broad policy questions, for example, “Did accidents increase when dockless bikes were introduced?”
- **Fast analytics on real-time data**, so that we can capture and monitor the many and growing streams of data already slipping through our fingers, for example, an alert pushed to emergency personnel when anomalies are detected in calls-for-service.
- **Flexible and scalable**, so that the system can be inexpensively expanded to handle the inevitable growth in data volume and usage. For example, we will need to increase data storage to keep pace as DC continues its investment in “Smart Cities” devices like smart trashcans and building and environmental sensors.
- **Tolerant to failure**, so that the system continues to operate in the event of a problem.

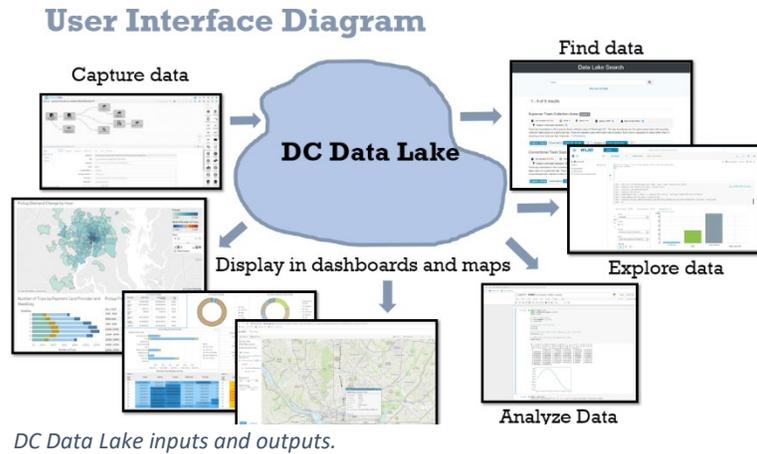
Discoverability 	Search Engine for City Staff at data.in.dc.gov
Usability 	Methods and tools to bring the data into City’s current data toolbox: MicroStrategy, Tableau, ArcGIS, Excel (csv), Jupyter notebooks, Spark, Kafka
Security 	Governance by Data Policy, encryption and role-based authentication at all phases 
Flexibility 	Unstructured and Structured data in a single platform for Advanced Analytics and AI.

DC Data Lake Characteristics. Icons from the Noun Project <https://thenounproject.com>

¹⁵<https://medium.com/what-works-cities-certification/washington-dc-excelling-in-evaluations-for-better-outcomes-53a4c8f301b4>.

¹⁶ <https://opendata.dc.gov/pages/handbook>

- **Improved data privacy and security**, so that agencies can be comfortable sharing Level 1 data and higher with other agencies.
- New ways to analyze data, so that DC government data scientists and analysts can perform analysis on unstructured data and mine large volumes of data for insights instead of relying on data summaries and subsets.
- **More discoverable**, so that DC staff can find, understand, and use available data.



Data Science and GIS User Groups

OCTO continues to chair and facilitate the Interagency Data Team¹⁷ and the DC GIS Steering Committee,¹⁸ both of which tend to draw mid-level managers interested in geospatial or business data. These groups, however, deal in technical policy, and many end users are not included. This year we started two new user groups for specialty end users: a GIS User Group for those interested in geospatial technology and a Data Science User Group for the District’s growing number of data scientists. These groups share tools, ideas, and best practices, and have been very well received. Both groups meet monthly. Data Science @ DC has discussed topics such as machine learning and ethics in algorithms. The GIS group has included talks from users and subject matter experts at MPD, DDOT, OCTO, the Baltimore City Department of Health, and the Chesapeake Conservancy.

¹⁷ <https://octo.dc.gov/page/interagency-data-team>.

¹⁸ <https://octo.dc.gov/page/dc-gis-steering-committee>.

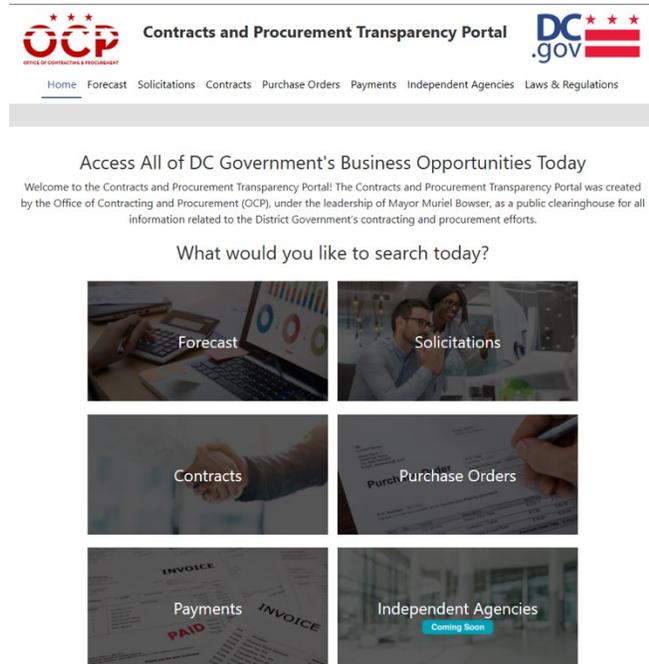
Improving Procurement Transparency

The Office of Contracts and Procurement (OCP) and the Office of the Chief Financial Officer (OCFO) have made significant strides in publishing procurement and payment data. With support from OCTO, OCP launched the Contracts and Procurement Transparency Portal¹⁹ and OCFO launched the DC Vendor Portal.²⁰ These coordinated efforts provide information to the vendor community and the public, including the following:

- Contracting opportunities
- Open, pending, and closed solicitations
- Information on past and active contracts
- Information on past and active purchase orders

For those who prefer to bypass the portal interfaces, these efforts resulted in new and improved open datasets as well:

- [Contracts](#)
- [Payments](#)
- [Forecast](#)
- [Purchase Orders](#)
- [Solicitations](#)



OCP launched a new Contracts and Procurement Transparency Portal (<https://contracts.ocp.dc.gov>)

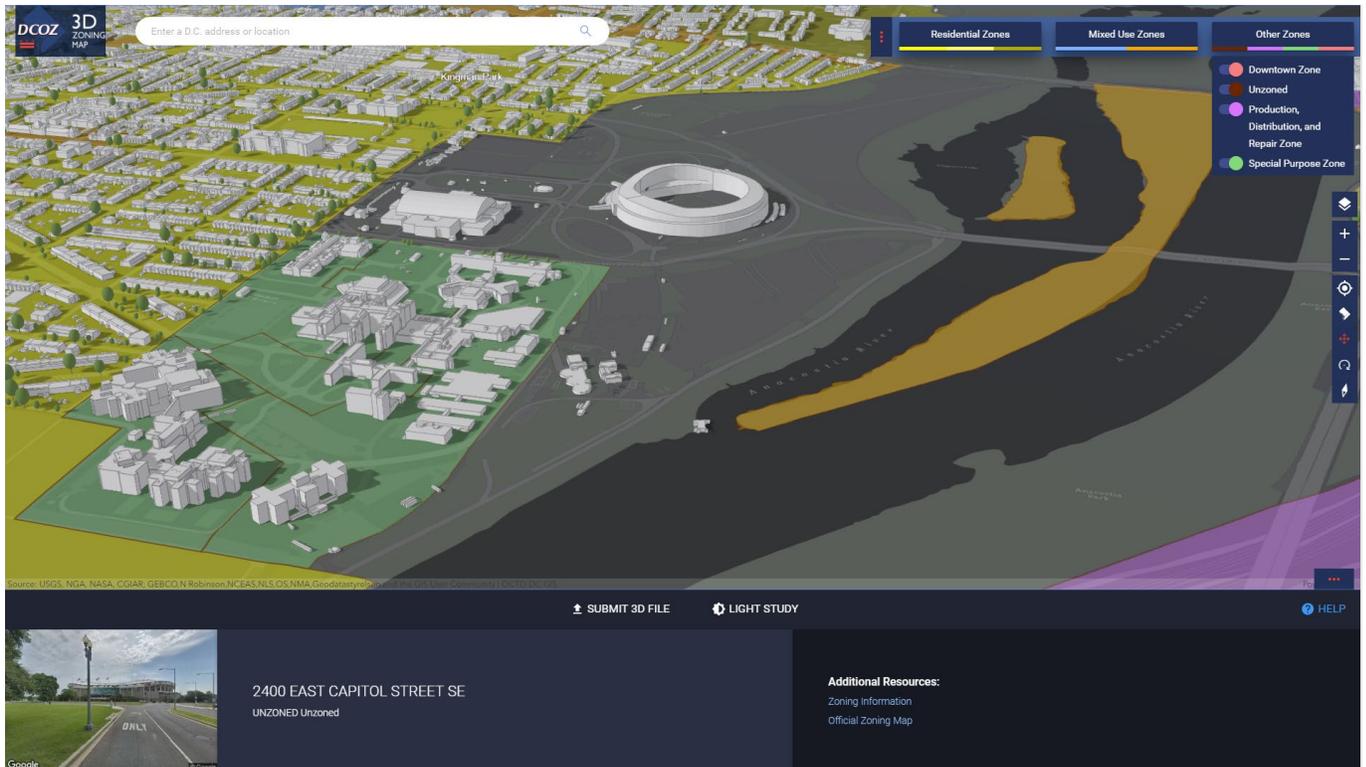
Zoning 3D Map, Last but Not Least

In October, the DC Office of Zoning released the 3D Zoning Map. The 3D Zoning Map was developed to enhance District residents' understanding, knowledge, and participation in zoning matters, and help increase transparency in the zoning process. The application builds off existing DC open datasets provided by OCTO and new zoning data to visualize the District in 3D, providing greater context for proposed development projects and helping enhance Board of Zoning Adjustment and Zoning Commission decisions throughout the District.²¹ Users of the new map can visualize proposed projects by uploading building models into the map, create before and after building scenarios, and perform shadow and light studies.

¹⁹ <https://contracts.ocp.dc.gov>.

²⁰ <https://vendorportal.dc.gov>.

²¹ <http://maps.dcoz.dc.gov/3d/>.



DCOZ 3D Zoning Map, <http://maps.dcoz.dc.gov/3d/>, shows the vicinity of DC General Hospital and RFK Stadium. Tools like this will assist with planning the future of these and other critical sites throughout the District.

Accountability for 2018 and New 2019 Goals+

2018 Goals (March 10, 2018 to March 10, 2019)	Accountability and Revised Goals (March 10, 2019 to March 10, 2020)
<p>Move Open Datasets to the Opendata.dc.gov Portal OCTO will work with agencies, the Open Government Advisory Group, and the community to prioritize posting the remaining 195 open enterprise datasets identified in the EDI that are not yet on the Open Data Portal.</p>	<p>Partially Accomplished Of those 195 datasets, 43 have been published on opendata.dc.gov. An additional 7 datasets turned out not to be structured data or are redundant and should not have been inventoried last year. This year's inventory identified more Level-0 datasets that should be on opendata.dc.gov.</p> <p>Revised 2019 Goal Prioritize these 189 datasets for publication prior to next year's report. The revised list is online.²²</p>
<p>Develop eMOU System to Support Data Sharing Agreements Not all data can be open, and the Data Policy calls on the CDO to develop a "streamlined process for interagency data sharing." A data-sharing agreement is a document of agreement between two agencies in which the data steward agrees to share specific data with another agency subject to certain terms and</p>	<p>Not Started This goal requires the full-time efforts of a CDO working between agencies and the Mayor to negotiate data sharing practices. Unfortunately, my CTO duties prevented me from this level of participation and from delivering this product. The silver lining: using my Interim leadership vantage point, I was able to reconsider the feasibility of the proposed solution, and I believe that the</p>

²² <https://opendata.dc.gov/datasets/enterprise-dataset-inventory-level-0-not-on-open-data-dc/data>

2018 Goals (March 10, 2018 to March 10, 2019)	Accountability and Revised Goals (March 10, 2019 to March 10, 2020)
<p>limitations. Faster execution, with better tracking and enforcement of data-sharing agreements, is needed across the District government. OCTO already maintains a system known as “eMOU,” where MOU stands for Memorandum of Understanding. With modifications, such as multilateral agreements, the eMOU can be adapted to handle standardized data-sharing agreements.</p>	<p>final product will be better designed and more widely adopted.</p> <p>Revised 2019 Goal Facilitating and tracking data-sharing agreements remains a goal for 2019. Some addition considerations:</p> <ul style="list-style-type: none"> • There is considerable legal work still required to come up with acceptable agreement templates. • The tool will need to support multilateral or brokered agreements, not just agreements between two agencies. • The technology envisioned needs to be reevaluated. EMOU is a good starting point, but the application might be better received by agency legal teams if they can continue working in Microsoft Word and we take more of a document management approach.
<p>Develop and Publish a Data Submission Guide Despite having collected open data from District agencies since 2001, OCTO does not have a formal data submission guide. Open data programs in other cities, including New York and San Francisco, have documents that establish minimum standards for submission and provide helpful instructions and examples. For example, each dataset submitted for publication should include standard metadata and a data dictionary.</p>	<p>Accomplished! As described above. The data submission guide, titled <i>Open Data Handbook</i>, has been published and is available.²³</p>
<p>Improve FOIA Request Tracking This complementary relationship, however, depends on having accurate and consolidated data on FOIA requests and the disposition of those requests. The District has a FOIA request tracking system, but it is not used by all agencies due to a lack of license and training. The CDO and CTO should acquire sufficient licenses and make training on the system available to all FOIA Officers.</p>	<p>Progress, but Not Enough The FOIA system used by the District is called FOIAXpress.</p> <ul style="list-style-type: none"> • OCTO used to charge agencies that needed new FOIAXpress accounts. That pushed agencies out of the system. OCTO no longer charges agencies for the service. • OCTO also greatly increased FOIAXpress training for FOIA Officers. • In March 2018, OCTO began publishing redacted versions of FOIA requests from FOIAXpress as open data.²⁴ • These improvements closely corresponded to the calendar year, and all utilization measures are up as shown below. <p>Despite a significant increase in utilization of FOIAXpress, the total number of total FOIA requests increased faster than the number of requests processed inside the system.</p> <p>Revised 2019 Goal The District is still failing to track a significant percentage of FOIA requests inside its enterprise system. Therefore, we are resolved to:</p>

²³ <https://opendata.dc.gov/pages/handbook>.

²⁴ <https://opendata.dc.gov/datasets/foia-requests>.

2018 Goals (March 10, 2018 to March 10, 2019)	Accountability and Revised Goals (March 10, 2019 to March 10, 2020)
	<ul style="list-style-type: none"> • continue regular training FOIA Officers; • identify agencies and FOIA officers not using the system, and work with them to increase system utilization; • begin publishing the District’s Annual FOIA Reports as open data in addition to PDF format so that the reports can be more easily analyzed²⁵; and • work with agencies to review publish the priority list from the FOIA section of this report. <p>Also, on behalf of FOIA requesters we will review and improve or replace the FOIAXpress Public Access Link.</p>
<p>Develop Cutting-Edge Data Platform to Support Analytics</p> <p>Develop a Hadoop-based big data backend capable of hosting critical data and support while lowering application development and support costs. The following are a few initial use cases being considered:</p> <ul style="list-style-type: none"> • Access and analyze large datasets. The crime data needs to be broken down into multiple views in the current Open Data Portal. A relational database on a big data computing platform would allow us to store and offer data within a single table. • Handle data streams from the Internet of Things (IoT) including a wide variety of Smart City sensors. • Generate multiple data visualizations (dashboards, maps, etc.) of big data with our supported enterprise tools—MicroStrategy, Tableau, and ArcGIS. • Support data analysis of large datasets by data scientists and analysts in The Lab @ DC, OCTO, and other agencies across the District. 	<p>Significant Progress</p> <p>We have developed the DC Data Lake.</p> <ul style="list-style-type: none"> • The Hadoop system is up and running. • We have demonstrated the ability to intake streams of data from IOT. • We have been able to connect the District’s enterprise COTS data analysis and visualization platforms including MicroStrategy, Tableau, and ArcGIS. • The Data Lake is now being put into production in cooperation with data scientists who work for The Lab @ DC. <p>Revised 2019 Goal</p> <p>Now is the time to push utilization of the Data Lake.</p> <ul style="list-style-type: none"> • The Lab @ DC has agreed to migrate at least one agency’s datasets into OCTO’s searchable “Data Lake” and undertake at least one multiagency Data Science Project with the system. • The OCTO Data Team will gather and catalog as much Level 1 and Level 2 data as we can to populate data.in.dc.gov.

²⁵ <https://os.dc.gov/page/annual-reports>

2018 Goals (March 10, 2018 to March 10, 2019)	Accountability and Revised Goals (March 10, 2019 to March 10, 2020)
<p>Provide More Assistance to Agencies to Comply With FOIA</p> <p>OCTO web editors should provide more assistance to agencies seeking to comply with Section 5 of the District's FOIA law. D.C. Official Code §2-536 states agency public-facing websites must contain specific information, including:</p> <ul style="list-style-type: none"> • public employee salary information; • administrative staff manuals and instructions; • statements of policy; • information dealing with the receipt or expenditure of public funds; • budget information; • minutes of public meetings; • frequently requested public records; • District-wide and agency FOIA reports; and • organizational chart. 	<p>Largely Accomplished</p> <p>OCTO Web Editors have updated all the open government pages that they manage. Some changes depend on agencies.</p> <p>Revised 2019 Goal</p> <p>With opendata.dc.gov, users can link to filtered data. The OCTO Data team trains web editors on this technique, so they can link to agency specific records from agency FOIA and open government web pages.</p>
<p>New Goal</p>	<p>Develop Application User Interface (API) Best Practices</p> <p>Part of the CDO mission is to facilitate data sharing. APIs are a common tool for fast, flexible and secure data integration. OCTO maintains minimal API standards for District agencies on its intranet site, however much more guidance is required. This is especially true where more and more, but not all data, is moving to cloud-based software and service applications.</p>
<p>New Goal</p>	<p>Train Agency Communications and Engagement Teams in Data Storytelling</p> <p>As Mayor Bowser has demonstrated with her Mayor.dc.gov website, opendata.dc.gov offers many ways to explain and track policies using open data. Familiarizing communications teams with the capabilities should drive publication and utilization of open data.</p>

In addition to goals, last year’s report included recommendations. Recommendations differ from goals in that they are clearly beyond the authority of OCTO to implement. Here are last year’s recommendations and their status.

2018 Recommendations (March 10, 2018 to March 10, 2019)	Status
<p>Move EDI and CDO Report Deadline to the Monday of Sunshine Week</p> <p>The goals of Sunshine Week and of the Data Policy closely align, and Sunshine Week activities offer significant opportunities to call attention to the District’s progress on open governance.</p>	<p>Done!</p> <p>Thank you, Mayor Bowser. The new deadline is March 10 of every year.</p>
<p>Adopt EDI Participation as a Key Performance Indicator for All Agencies</p> <p>The Office of the City Administrator should develop Key Performance Indicators (KPIs) that encourage agency participation in the EDI:</p>	<p>Done!</p> <p>Thank you, City Administrator Young. The new citywide KPI is “IT POLICY AND FOIA COMPLIANCE -- Percent of ‘open’ data sets identified by the annual Enterprise Dataset Inventory published on the Open Data Portal.”</p>
<p>Adopt Reasonable and Uniform Retention Policy for Email</p> <p>Currently the District does not have a retention schedule for email. Absent a policy, OCTO stores all email for all agencies indefinitely. Currently, OCTO stores more than 293 terabytes of email and attached documents. The oldest email in OCTO’s collection is from 1998. This results in two problems:</p> <ul style="list-style-type: none"> • Email storage must be spread among multiple servers, and searching for old emails is cumbersome. When District agencies fail to meet legislated requirements for processing FOIA requests, slow email searches often constitute a large portion of the delay. • Storing that much email is expensive, and costs continue to mount. <p>Therefore, the Mayor should adopt a reasonable email retention policy that requires email be stored for a fixed period. Ideally, the Mayor can also adopt a uniform standard that applies to all agencies and types of email content. A uniform standard can be cost effectively administered.</p>	<p>Remains a Recommendation</p> <p>From last year’s report to now, there was no significant movement toward this recommendation. Meanwhile the District’s email retention costs increased approximately 15%. Many of the emails the District currently retains have aged beyond any foreseeable statute of limitations. Moreover, the growing quantity of emails continues to slow FOIA responses, many of which include email searches. Additionally, OCTO does not know how much PII and other sensitive information the District’s vast archive of emails contains in its archive. The recommendation of moving to a reasonable and uniform retention policy remains.</p>
<p>Encourage Independent Agency Participation in the Enterprise Dataset Inventory</p> <p>The current Data Policy is a Mayor’s Order and therefore cannot be enforced for independent agencies. During this first EDI, 100% of mayoral agencies participated, including all large agencies, but only 33% of independent agencies participated. It is therefore recommended that OCTO do more to encourage participation. This should include seeking support from Deputy Mayors. If that fails, seeking legislation would be an option.</p>	<p>Some progress but More to Do</p> <p>Two more independent agencies contributed to the inventory this year. As independent agencies hold valuable datasets, OCTO will continue to encourage their participation.</p>

2018 Recommendations (March 10, 2018 to March 10, 2019)	Status
<p>Legislation to Clarify FOIA-Exempt Critical Infrastructure Information</p> <p>During the course of the EDI, it became clear that multiple agencies have enterprise datasets that they believe contain “critical infrastructure information” and should not be publicly released. The District’s exemptions to FOIA, D.C. Official Code §2-534, does not define “critical infrastructure.” It does exempt “Any critical infrastructure information or plans that contain critical infrastructure information for the critical infrastructures of companies that are regulated by the Public Service Commission of the District of Columbia.” Other DC agencies that clearly have critical infrastructure information, including DGS, DDOT, OCTO, and OUC, are not covered by the exemption. It is therefore recommended that D.C. Official Code §2-534 be amended to clearly define and exempt critical infrastructure information.</p>	<p>Remains a Recommendation</p> <p>From last year’s report to now, there was no significant movement toward this recommendation. Critical instruct remains worthy of definition and protection.</p>

Conclusion

The CDO and the OCTO Data program exist primarily to serve other District agencies. We are what DC calls an “internal service.” We strive to help agencies manage and use data, so they can be more efficient and effective. Although agencies are our direct customers, the underlying reason for everything we do is to serve the public. We want to hear about the datasets you need and the ideas you have. What did you like about this report and inventory list? What could we be doing better? What are we not doing at all that we should be doing? Please reach out by email to open.data@dc.gov or Twitter [@opendatadc](https://twitter.com/opendatadc) and let us know what you think and how we can do better.