GOVERNMENT OF THE DISTRICT OF COLUMBIA
Office of the Chief Technology Officer

Annual Chief Data Officer Report

Barney Krucoff

Interim Chief Technology Officer

Sunday, March 11, 2018
200 I Street SE
Washington, DC 20003

Introduction

Mayor Bowser is “making our local government one of the most accessible systems in the country.” To that end, the Mayor issued Executive Order 2017-115, District of Columbia Data Policy, on April 27, 2017, with the stated goal of leading the District of Columbia government toward more open and efficient use and sharing of government data.

The policy established these principles acknowledging the value of data to the District and the inherent need to balance openness with other concerns:

- Data are valuable assets independent of the information systems in which the data reside.
- The greatest value from those assets is realized when freely shared to the extent consistent with the protection of safety, privacy, and security.

Because the District deals with open data and protected data in the same policy, it is a “data policy,” not an “open data policy.” The balance between open and closed was potentially controversial but open government advocacy groups including the District’s own Open Government Advisory Group[1] and the Sunlight Foundation have recognized its importance:[2]

> DC has incorporated security and privacy into its data policy in a way that should advance the international dialogue around how municipal governments collect, structure and disclose public information.… DC’s transparency regarding its classification system and efforts to publicly balance valid security and privacy concerns with the public access expectations that are standard in a democratic state is admirable, as was the decision to put the draft policy up for public comment.

Although the policy balances openness and security, it prioritizes transparency. The policy states that “enterprise datasets shall be **open by default**, meaning 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 although the District has many datasets that it cannot share, we should not have any enterprise datasets that we will not publicly acknowledge, which is accomplished by publishing the annual Enterprise Dataset Inventory.

Enterprise Dataset Inventory

The Data Policy mandates that public bodies in the District government create and maintain an Enterprise Dataset Inventory (EDI) following the leadership of the Office of the Chief Technology Officer. The inventory requires agencies under the direct authority of the Mayor to

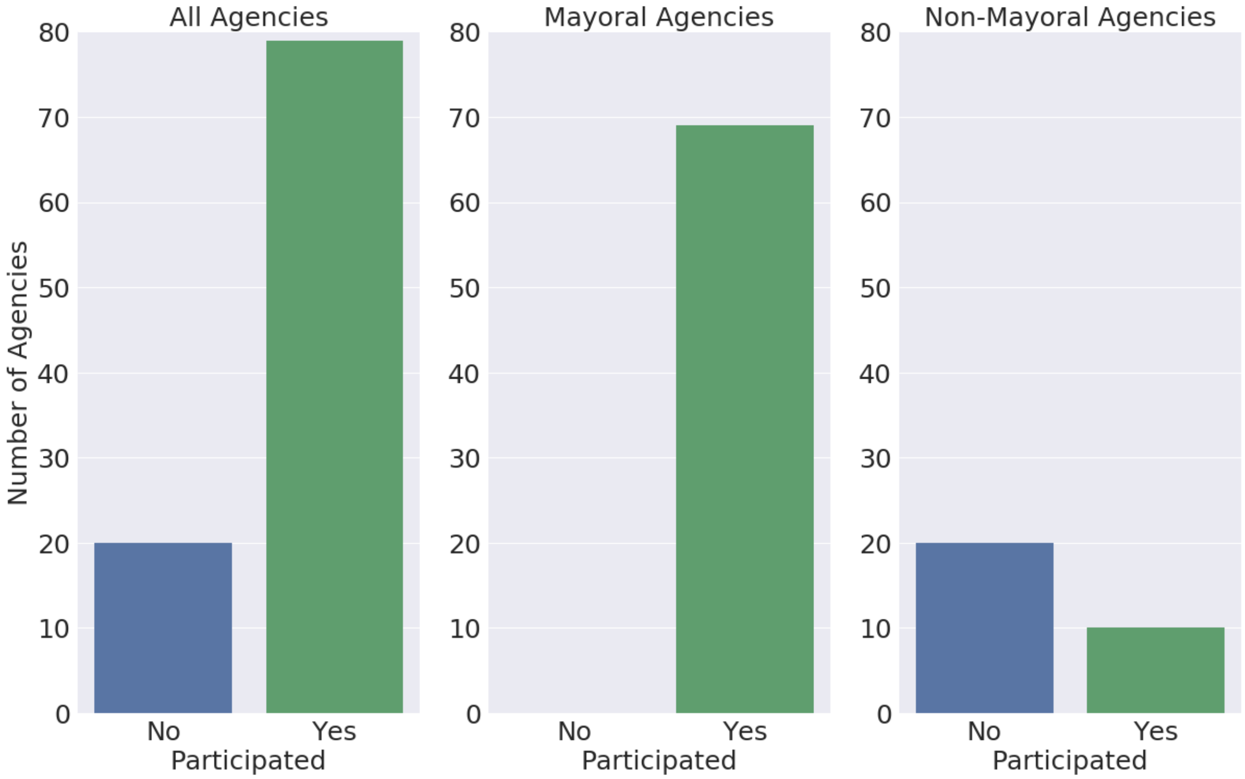
[1] https://ogag.dc.gov. Accessed 3/11/2018.

[2] https://sunlightfoundation.com/2017/05/04/dc-data-policy-balances-privacy-security-and-openness/. Accessed 3/11/2018.

record any "enterprise dataset," which is "a dataset that directly supports the mission of one or more public bodies."[3] 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.[4]

How Many Agencies Participated?

Figure 1
How many agencies participated?



Over the course of the data inventory, 79 agencies recorded 1,640 enterprise datasets, nearly 80% of the 99 agencies that were asked to participate (see Figure 1).[5] All 69 mayoral agencies, which were required to participate, worked with the Office of the Chief Technology Officer (OCTO) to identify and record their enterprise datasets. In contrast, only 10 of the city's 30 independent agencies participated in the data inventory. Unlike mayoral agencies, independent agencies were not required to participate in the inventory.

[3] For more on how we define enterprise datasets, see Section III of the DC Open Data Policy https://octo.dc.gov/page/district-columbia-data-policy. Accessed 3/11/2018.

[4] http://opendata.dc.gov/datasets/enterprise-dataset-inventory. Accessed 3/11/2018.

[5] An agency participated if they had an enterprise dataset recorded in the inventory, or attempted to provide an enterprise dataset in the inventory. In some cases, very small offices had no enterprise datasets, and therefore participated, despite having no dataset in the inventory.

How Many Datasets Did Agencies Inventory?

Figure 2
How many datasets did agencies inventory?

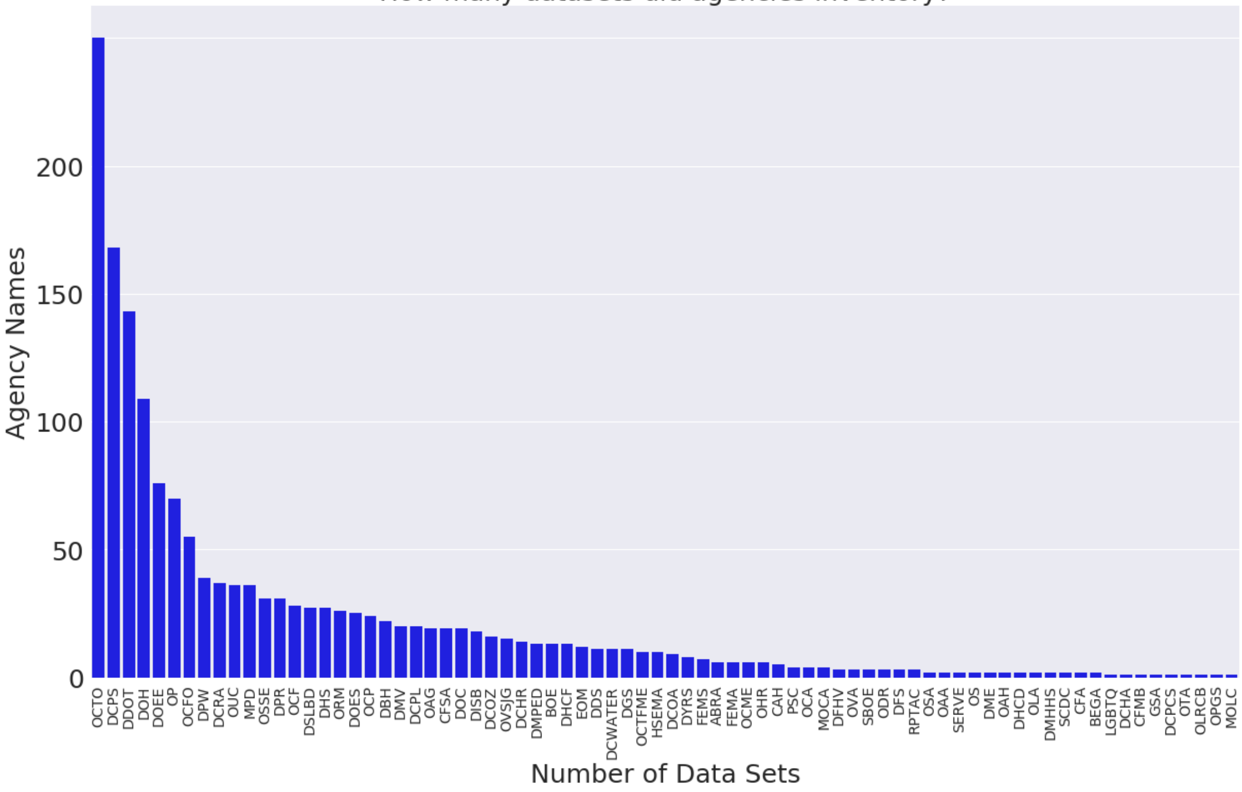


Figure 2 shows the number of datasets recorded by each agency that participated in the EDI. Half of all participating agencies recorded 10 or more datasets to the inventory. The Office of the Chief Technology Officer (OCTO) submitted the most datasets to the inventory with 250 enterprise datasets. Other large contributors included DC Public Schools (DCPS) with 168, followed by the Department of Transportation (DDOT) with 143, and the Department of Health (DOH) with 109.

How Were Datasets Classified?

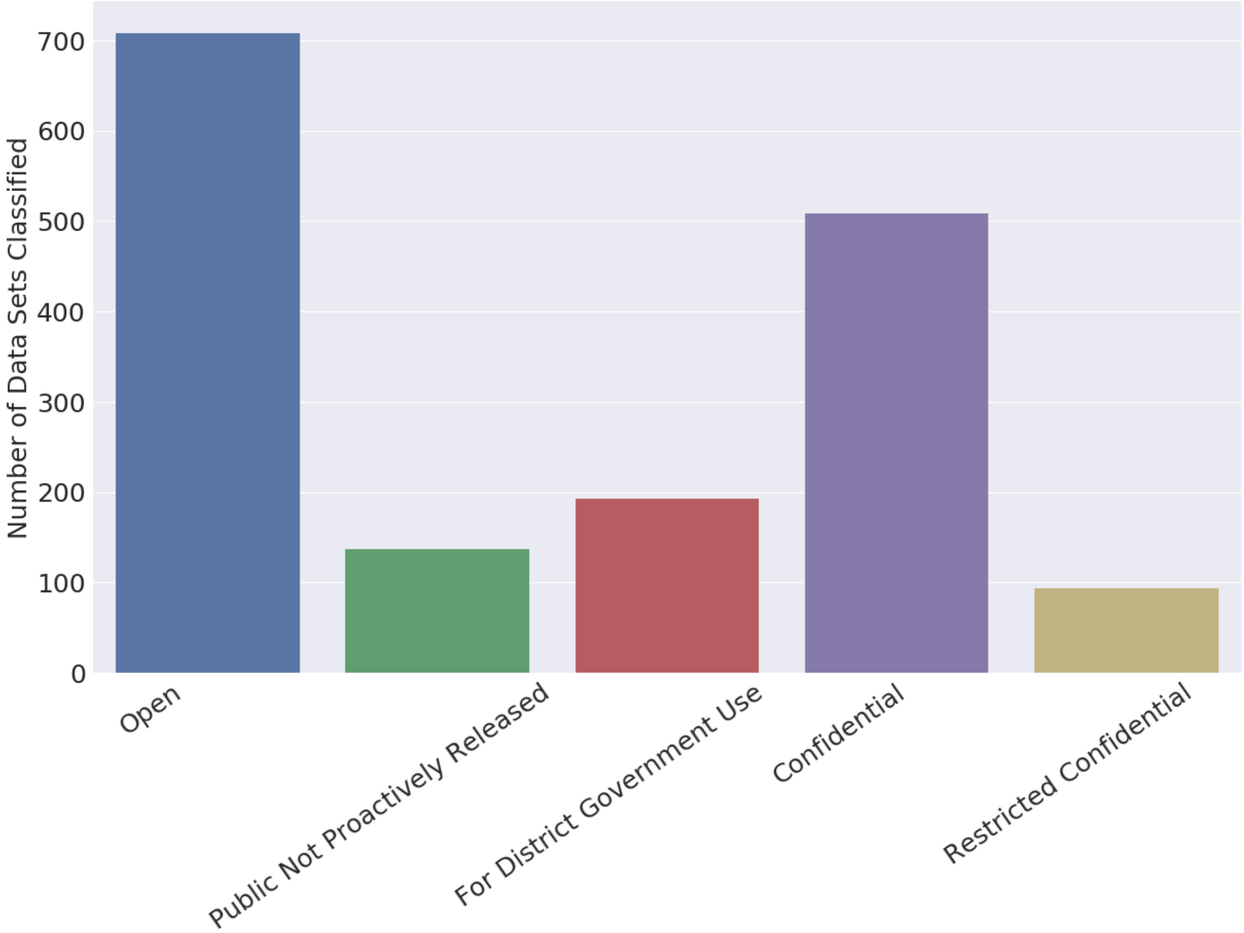
The DC Data Policy also lays out a **dataset classification** system to help determine which enterprise datasets should be open to the public and which should not be proactively released.[6] Figure 3 shows how agencies classified their datasets.

Level 0 is **Open** data. 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

[6] For more information on dataset classifications, see the District of Columbia Data Policy, Section III https://octo.dc.gov/page/district-columbia-data-policy. Accessed 3/11/2018.

agencies do not determine to have a higher security level. Of the 1,640 datasets in the inventory, 708 (43.2%) were classified Open, making it the most common classification.

Figure 3
How were datasets classified?



Level 1 data 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.[7] Only 137 (8.4%) of datasets in the inventory were classified Public Not Proactively Released, making it one of the least common classifications.

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.”[8] 193 (11.8%) of datasets in the inventory were classified For District Government Use.

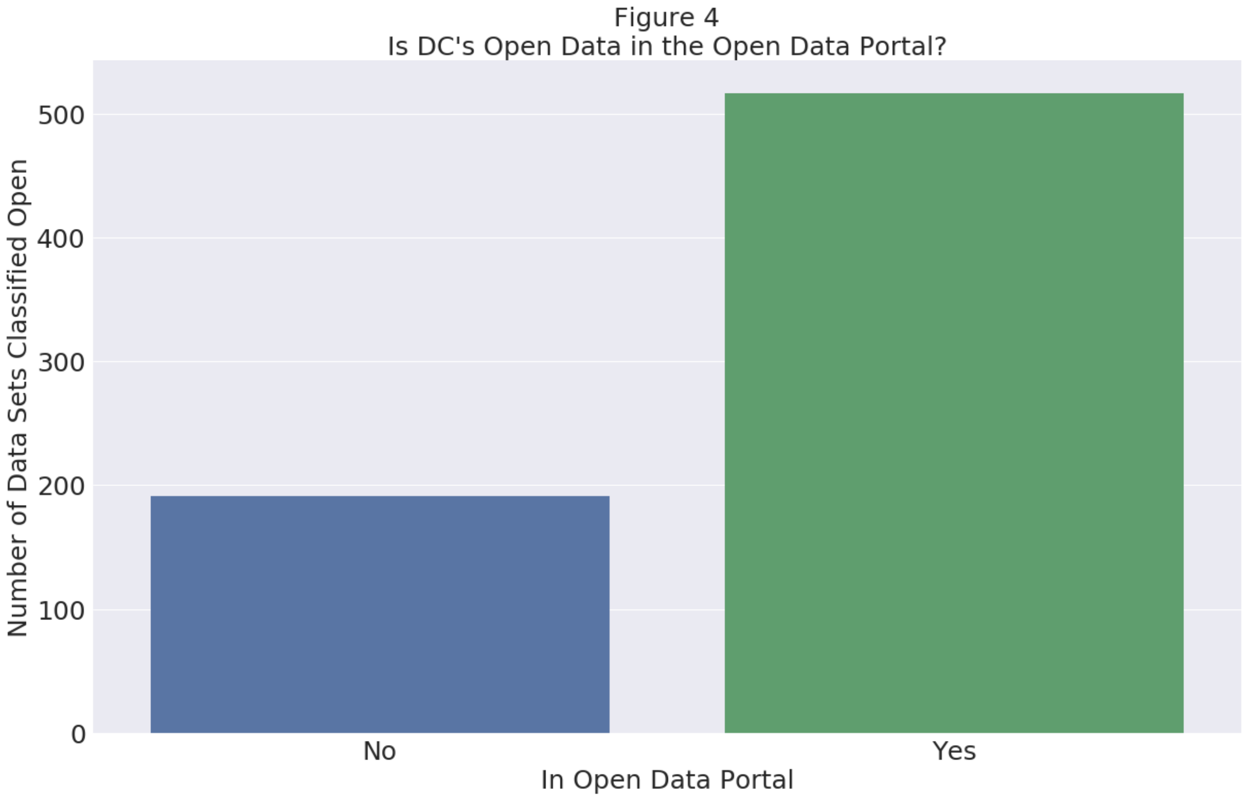
[7] Ibid.

[8] Ibid.

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.[9] 508 (31.0%) of the datasets in the inventory were classified Confidential, making it the second most common classification. This is unsurprising since the District government collects a lot of data it cannot legally share with the public, including health data (that is protected by HIPAA), student data (that is protected by FERPA), and various kinds of personally identifying information (PII).

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."[10] Only about 94 (5.7%) of the datasets recorded in the inventory were classified Restricted Confidential, making it the least common classification in the inventory.

How Many Open Datasets Are on the Open Data Portal?



Of the 708 datasets classified Open in the EDI, 517 (73.0%) are on the city's Open Data Portal (Figure 4). OCTO will work with agencies, the Open Government Advisory Group, and the

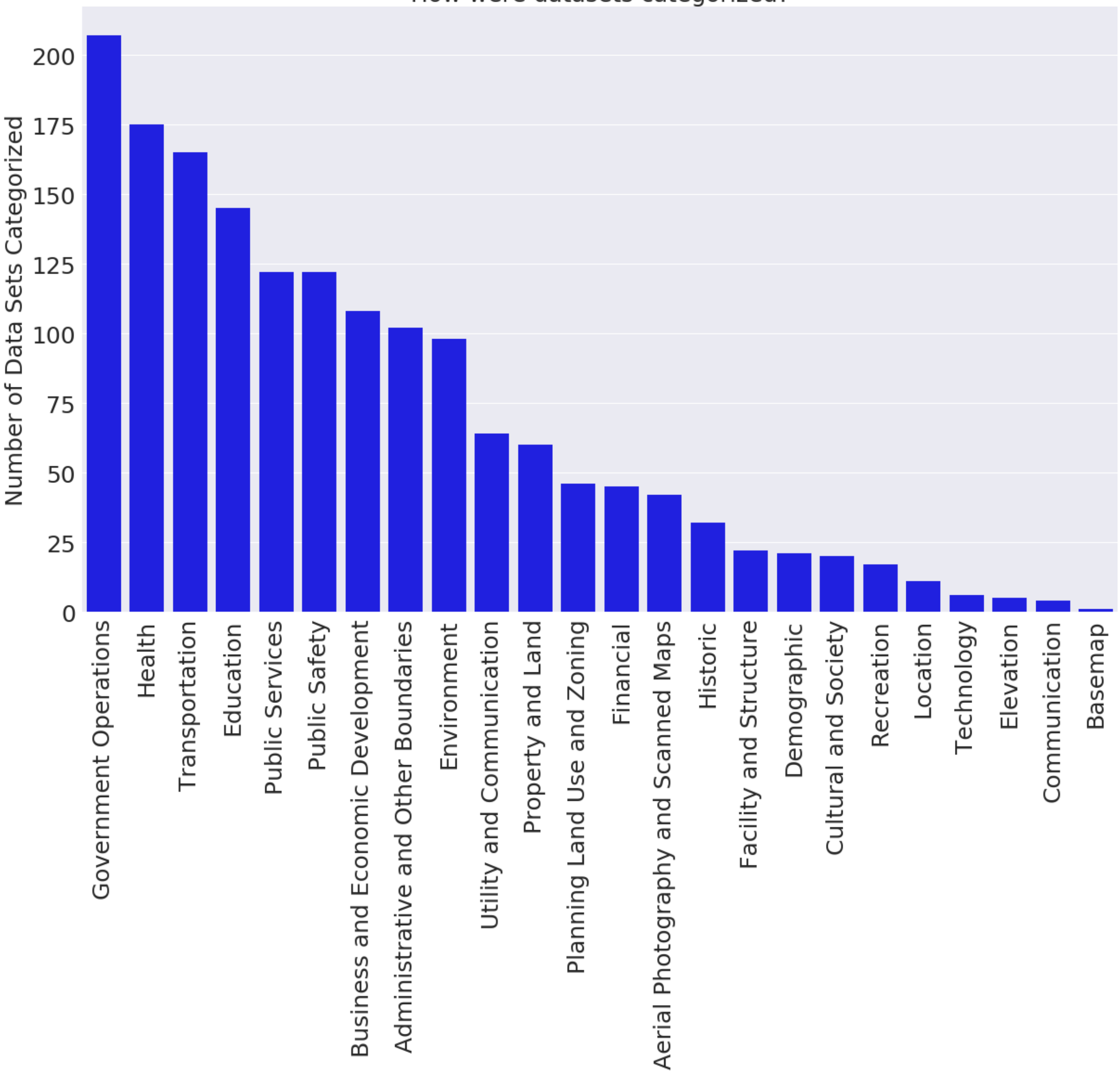
[9] Ibid.

[10] Ibid.

community to prioritize the remaining 191 Open enterprise datasets identified in the inventory for inclusion on Open Data Portal.[11]

How Were Datasets Categorized?

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. There are over a hundred datasets in

[11] For a list of the 191 Open enterprise datasets that were classified Level 0, Open, but are not on the Open Data Portal at the time of this report, see http://opendata.dc.gov/datasets/enterprise-dataset-inventory-level-0-not-on-open-data-dc. Accessed 3/11/2018.

categories including Government Operations (207), Health (175), Transportation (165), Education (145), Public Services (122), and Public Safety (122). Other well-represented categories include Business and Economic Development (108) and Administrative Boundaries (102). The EDI also includes many datasets containing information about Environment (98), Utility and Communication (64), and Property and Land (60).

OCTO Data Team Accomplishments

In addition to establishing the DC Data Policy and conducting the Enterprise Dataset Inventory, the OCTO Data Team had many other accomplishments over the past year.

Interagency Data Team

The OCTO Data Team established the Interagency Data Team (IDT)[12], a group composed of Agency Data Officers (ADOs).[13] The group meets regularly to work on sharing data, tools, and techniques. The goal is to empower cross-cutting analytics by District agencies.[14]

Community Events and Hackathons

The OCTO Data Team supported multiple community events and hackathons including DCFemTech, SaferSmarterStrongerDC, innoMAYtion, and Health and Human Services hackathon.

Expanding the District Government's Analytic Capability

The OCTO Data Team hired its first two Data Scientists. These positions are currently supporting the Office of Unified Communications and The Lab @ DC in the Office of the City Administrator.

Aerial Imagery Services

The OCTO Data Team was awarded a five-year aerial imagery services contract. Services include aerial photography, LiDAR, and an updated DC's base mapping layers. New leaf-off aerial photography was collected in Spring 2017.

[12] https://octo.dc.gov/page/interagency-data-team. Accessed 3/11/2018.

[13] Mayor's Order 2017-115, Section IV, Part C (7)

[14] For more about the Interagency Data Team, visit https://octo.dc.gov/page/interagency-data-team. Accessed 3/11/2018.

Centralizing Capital Asset Data

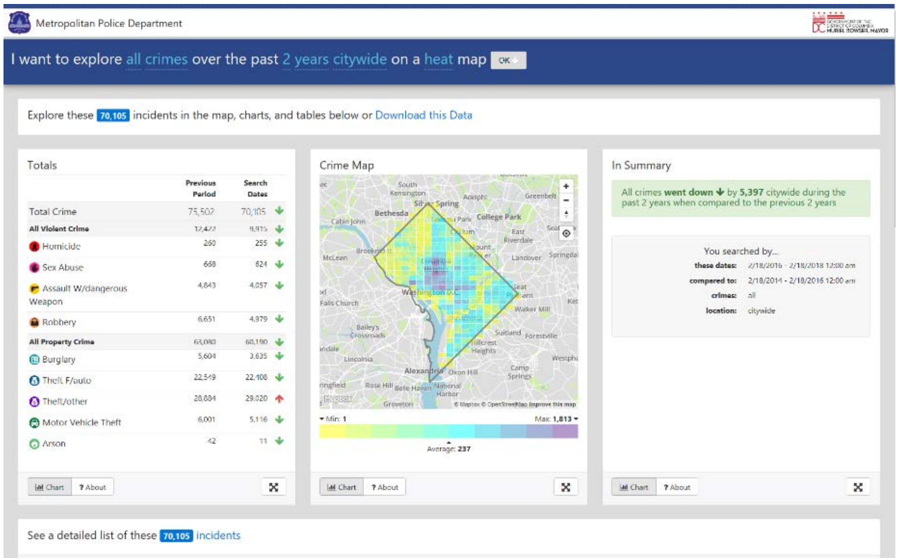
The OCTO Data Team worked with five major agencies[15] to extract, transform, and load capital asset data into a central database. These data were consumed by an application called Capital Asset Replacement Schedule System (CARSS). CARSS is used by the Office of the Chief Financial Officer (OCFO) and the Office of Budget and Planning (OBP) to model asset management and infrastructure planning, including creation of capital budget scenarios.

Establishing Enterprise Standards for Business Intelligence

The OCTO Data Team established MicroStrategy and Tableau as the enterprise standards for business intelligence tools in the District government and completed major upgrades to those systems. Progress includes

- quadrupling the District's server-side capacity to use the Tableau data visualization platform.
- creating a free Tableau class for District employees.
- reworking our MicroStrategy development environment so that it fully matches our production environment.
- publishing a library of best practices and end user support documents for the MicroStrategy Environment.

Crime Cards



Metropolitan Police Department

I want to explore all crimes over the past 2 years citywide on a heat map

Explore these 70,105 incidents in the map, charts, and tables below or Download this Data

Totals

	Previous Period	Search Dates	
Total Crime	75,502	70,105	↓
All Violent Crime	12,422	9,915	↓
Homicide	260	255	↓
Sex Abuse	668	624	↓
Assault W/dangerous Weapon	4,843	4,057	↓
Robbery	6,651	4,979	↓
All Property Crime	63,080	60,190	↓
Burglary	5,604	3,635	↓
Theft F/auto	22,549	22,108	↓
Theft/other	28,884	29,020	↑
Motor Vehicle Theft	6,001	5,116	↓
Arson	42	11	↓

In Summary

All crimes **went down** ↓ by **5,397** citywide during the past 2 years when compared to the previous 2 years

You searched by...

these dates:	2/18/2016 - 2/18/2018 12:00 am
compared to:	2/18/2014 - 2/18/2016 12:00 am
crimes:	all
location:	citywide

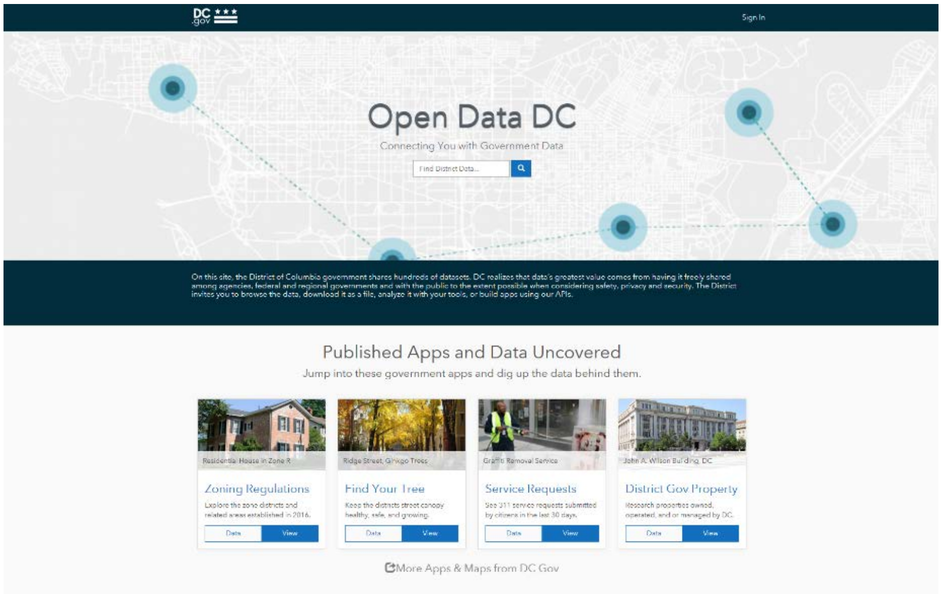
See a detailed list of these 70,105 incidents

CrimeCards (http://dcatlas.dcgis.dc.gov/crimecards)

[15] Department of Parks and Recreation (DPR), District Department of Transportation (DDOT), Department of General Services (DGS), Fire and Emergency Services (FEMS), and the Office of the State Superintendent for Education (OSSE).

The OCTO Data Team Completed Crime Cards,[16] an application that will replace the well-known crimemap.dc.gov. The new application is mobile friendly, uses a search engine to simultaneously present eight-times the amount of data in the old application, and still allows for faster analysis. The application is built on an all open source software stack including software from the District-based MapBox. The highlight of Crime Cards is an innovative sentence-based query interface that allows an ordinary non-technical person to create over 174,000 complex queries by writing a standard English sentence.

Updated Open Data Portal


DC .gov
Sign In
Open Data DC
Connecting You with Government Data
Find District Data...
Published Apps and Data Uncovered
Residential House in Zone R
Zoning Regulations
Data
View
Ridge Street, Ginkgo Trees
Find Your Tree
Data
View
Graffiti Removal Service
Service Requests
Data
View
John A. Wilson Building, DC
District Gov Property
Data
View
More Apps & Maps from DC Gov

On this site, the District of Columbia government shares hundreds of datasets. DC realizes that data's greatest value comes from having it freely shared among agencies, federal and regional governments and with the public to the extent possible when considering safety, privacy and security. The District invites you to browse the data, download it as a file, analyze it with your tools, or build apps using our APIs.

Jump into these government apps and dig up the data behind them.

Explore the zone districts and related areas established in 2016.

Keep the districts street canopy healthy, safe, and growing.

See 311 service requests submitted by citizens in the last 30 days.

Research properties owned, operated, and or managed by DC.

District of Columbia Open Data Catalog (http://opendata.dc.gov)

Released a significant update to the District's Open Data Catalog. The catalog now features recently published web applications built by the OCTO Data Team that directly link to the data. It also takes visitors to a wealth of educational content for using web services, APIs, and even an application starter kit for developers on GitHub.[17] The Data Team also maintained and documented hundreds of datasets for opendata.dc.gov. Many new or extensively upgraded datasets were released in the past year.[18] Highlights include the following:

- Agency List, Office of the Chief Technology Officer[19]

[16] https://dcatlas.dcgis.dc.gov/crimecards/. Accessed 3/11/2018.

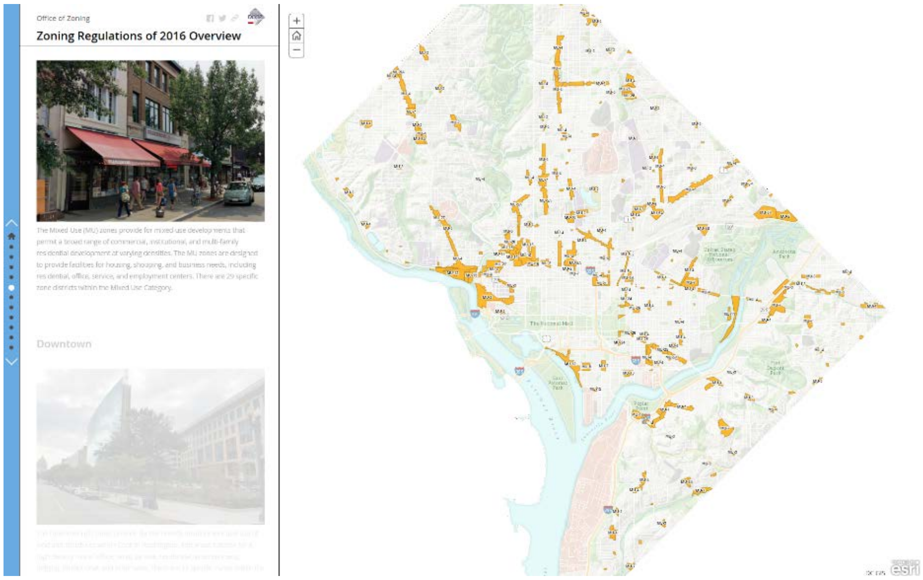
[17] https://github.com/DCgov/opendatadc-starterkit. Accessed 3/11/2018.

[18] For a complete list of datasets newly added in FY17, see http://opendata.dc.gov/datasets/enterprise-dataset-inventory-new-in-fy17, and updated in FY17, see http://opendata.dc.gov/datasets/enterprise-dataset-inventory-updated-in-fy17. For a complete list of datasets newly added in FY18, see http://opendata.dc.gov/datasets/enterprise-dataset-inventory-new-in-fy18, and updated in FY18 see http://opendata.dc.gov/datasets/enterprise-dataset-inventory-updated-in-fy18. Accessed 3/11/2018.

[19] http://opendata.dc.gov/datasets/district-government-agencies. Accessed 3/11/2018.

- Base Maps, Office of the Chief Technology Officer[20]
- Campaign Finance, Office of Campaign Finance[21]
- FOIA request tracking, Office of the Chief Technology Officer[22]
- Historic Buildings, Office of Planning[23]
- Homeless Shelters and Service Facilities, Department of Human Services[24]
- Impervious Surface, Office of the Chief Technology Officer, in cooperation with DC Water and Department of Energy and Environment[25]
- Incarceration Counts[26] and Short Term Sentenced Felons,[27] Department of Corrections
- LiDAR Point Cloud, Office of the Chief Technology Officer in cooperation with Amazon Web Services[28]
- Taxi Trips, Department of For Hire Vehicles[29]
- Vehicle Crashes, Department of Transportation and Metropolitan Police Department[30]
- Zoning Regulations 2016, Office of Zoning (reflecting new code)[31]
- 311 Service Requests, Office of Unified Communications (greatly expanded offering)[32]

Open Data Portal Initiative Pages



Office of Zoning

Zoning Regulations of 2016 Overview

The Mixed Use (MU) zones provide for mixed use developments that permit a broad range of commercial, institutional, and multi-family residential development at varying densities. The MU zones are designed to provide facilities for housing, shopping, and business needs, including residential, office, service, and employment centers. There are 29 specific zone districts within the Mixed Use Category.

Downtown

[20] http://opendata.dc.gov/datasets?q=basemaps. Accessed 3/11/2018.

[21] http://opendata.dc.gov/datasets?q=campaign%20finance. Accessed 3/11/2018.

[22] http://opendata.dc.gov/datasets/foia-requests. Accessed 3/11/2018.

[23] http://opendata.dc.gov/datasets/historic-data-on-dc-buildings. Accessed 3/11/2018.

[24] http://opendata.dc.gov/datasets?q=homeless%20shelter. Accessed 3/11/2018.

[25] http://opendata.dc.gov/datasets?q=impervious%20surface. Accessed 3/11/2018.

[26] http://opendata.dc.gov/datasets/incarceration-daily-counts-from-fy-2011-to-june-fy-2016. Accessed 3/11/2018.

[27] http://opendata.dc.gov/datasets/short-term-sentences-felons-2013-to-2016. Accessed 3/11/2018.

[28] https://aws.amazon.com/public-datasets/dc-lidar/. Accessed 3/11/2018.

[29] http://opendata.dc.gov/datasets?q=taxi%20trips. Accessed 3/11/2018.

[30] http://opendata.dc.gov/datasets?q=vehicle%20crash. Accessed 3/11/2018.

[31] http://opendata.dc.gov/datasets/zoning-regulations-of-2016. Accessed 3/11/2018.

[32] http://opendata.dc.gov/datasets?q=service%20requests. Accessed 3/11/2018.

District of Columbia Open Data Catalog (http://opendata.dc.gov)

The Data Team added “Initiative Pages” to opendata.dc.gov. These pages provide agencies that release data as Level-0, Open, with an expanded opportunity to explain the data to the public.

Goals for the Coming Year

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.

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 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.

Develop and Publish 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.

Improve FOIA Request Tracking

The data policy was intended to complement FOIA in the following ways:

- Where FOIA covers all District government information, this policy shall apply to a more finite subset of information—“data”—that has been organized into “enterprise datasets.”
- Where FOIA is largely reactive—relying on requests to trigger reviews and potential release of information—this policy shall be proactive, requiring agencies to consider and classify datasets for purposes of protection and distribution in advance of any requests.
- Where FOIA permits agencies to transfer the costs of classifying, processing, and distributing information to the requestor, under this policy the costs of classifying and,

when determined to be appropriate, processing and distributing datasets via the Internet shall be borne by the government.

- Where FOIA has an appeals process, this policy shall do nothing to inhibit the rights of requestors to make appeals under FOIA.

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.

Develop Cutting-Edge Data Platform to Support Analytics

The OCTO Data Team provides data and analytics services, public information applications, APIs, and support services to OCTO internal teams, other District agencies, and the public. Our goal is to continually meet the growing needs and expectations of the stakeholders and to serve as a technology partner with a focus on providing solutions. There is a clear need to provide more relational databases, which will require a cutting-edge big data computing platform. OCTO will develop a Hadoop-based backend in 2018 capable of hosting critical data and support while lowering application development and support costs. The following are a few initial use cases being considered:

- 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.

Provide More Assistance to Agencies to Comply with FOIA

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:

- 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
- Organizational chart

Recommendations

Move EDI and CDO Report Deadline to the Monday of Sunshine Week

It is recommended that the Mayor update the Data Policy to change the annual due date for the EDI and CDO report from November 1 to the Monday 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 and the dangers of excessive and unnecessary secrecy. It was established in March 2005 with funding from the John S. and James L. Knight Foundation. Sunshine Week occurs each year in mid-March, coinciding with James Madison's birthday and National Freedom of Information Day on the 16th.” It is simply fortuitous that the initial EDI and report are being published March 11, 2018, the start of Sunshine Week 2018. Nevertheless, 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.

In addition to this, the current report date of November 1 means that the EDI will coincide with many District reports that are due at the end of the fiscal year. There are also many personnel, procurement, and financial actions due during the same time. Moving the EDI deadline to mid-March will mean that the EDI will not conflict with the other priorities facing District managers.

Finally, having just completed the initial inventory, it is too soon for agencies to revisit their work for November 1, 2018. The data inventory can be beneficial to agencies, but only if it is not too burdensome.

Adopt EDI Participation as a Key Performance Indicator for All Agencies

The Office of the City Administrator should develop Key Performance Indicators (KPIs) that encourage agency participation in the EDI:

- City-wide: total number of datasets logged in the EDI.
- City-wide total number of datasets classified as Level-0; total number of datasets posted on the online data portal.
- Each agency: number of datasets in the inventory.
- Each agency: total number of datasets classified as Level-0; total number of datasets posted on the online data portal.

Adopt Reasonable and Uniform Retention Policy for Email

Currently the District of Columbia 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:

- 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.

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.

Encourage Independent Agency Participation in the Enterprise Dataset Inventory

The current Data Policy is a Mayor's Order and therefore cannot be enforced for independent agencies. During this first EDI, all mayoral agencies participated, but only a third 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.

Legislation to Clarify FOIA-Exempt Critical Infrastructure Information

Over 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.

Office of the Chief Technology Officer

Authors

Barney Krucoff
Interim Chief Technology Officer
Chief Data Officer

Peter Casey
Senior Data Scientist, Data Visualization and Analysis

Data Management and Curation

Michael Bentivegna
Program Manager, Data Visualization and Analysis

Eva Reid
Senior Analyst, Data Visualization and Analysis

Mario Field
Program Manager, Data Curation

Open Data Portal Initiatives Page

Alexandre Santos
GIS Program Analyst, GIS Enterprise Mapping

Acknowledgments

Special thanks to the OCTO Data Team, Agency Data Officers, and Agency General Counsels for their hard work gathering and preparing the Enterprise Dataset Inventory for publication.

Special thanks also to the members of the Open Government Advisory Group for their contributions to the District of Columbia Data Policy and their advice on and attention to the data policy and inventory. Finally, thanks to Mayor Muriel Bowser and her administration, particularly the Office of the City Administrator, for their commitment to government transparency and for keeping everyone on task.