

## DC Interagency Data Team

OFFICE OF THE CHIEF TECHNOLOGY OFFICER  
200 I STREET SE, WASHINGTON, DC 20003  
FIRST FLOOR  
CONFERENCE ROOMS 1001A & 1001B  
WEDNESDAY, NOVEMBER 2, 2016  
1:00pm – 3:00pm

- Opening remarks and Introductions, Archana Vemulapalli, Chief Technology Officer, OCTO
- Introduce Ms. Natalie Evans Harris, Archana Vemulapalli
- Keynote, Natalie Evans Harris Senior Policy Advisor, Data-Driven Government at White House Office of Science and Technology Policy
- DC Interagency Data Team Mission & Governance, Barney Krucoff, Chief Data Officer, OCTO
- Draft DC Data Policy, Barney Krucoff, OCTO
  - Preview IT Assessment & Data Inventory Tool, Mario Field, OCTO
- *11/17/2016 – Note: with a full agenda, this is the point where the meeting adjourned.*
- OCA's Data Team and the Lab in DC, Jenny Reed, Chief Performance Officer, OCA and David Yokum, Director, The Lab @ DC
- ~~OCTO Data Team, Goals and Structure, Barney Krucoff~~
- BI Tool Consolidation, Michael Bentivegna, OCTO
- Open Discussion and Next Steps



Opening remarks and Introductions Archana  
Vemulapalli,  
Chief Technology Officer



Natalie Evans Harris

Senior Policy Advisor, Data-Driven Government at  
White House Office of Science and Technology Policy



# Data Driven Culture

## Mission

*Responsibly unleash the power of data to benefit all Americans*

## Challenge

*To meet our nation's greatest economic and social challenges, the federal government must leverage their data as a mechanism for engaging with other agencies and the public.*

*Innovative Agencies use data to:*

- Drive Decisions
- Inform and Engage the Public
- Increase Capabilities
- Foster Collaboration

## EOP

*Empower agencies with the people, processes and practices necessary to build their data capacity.*

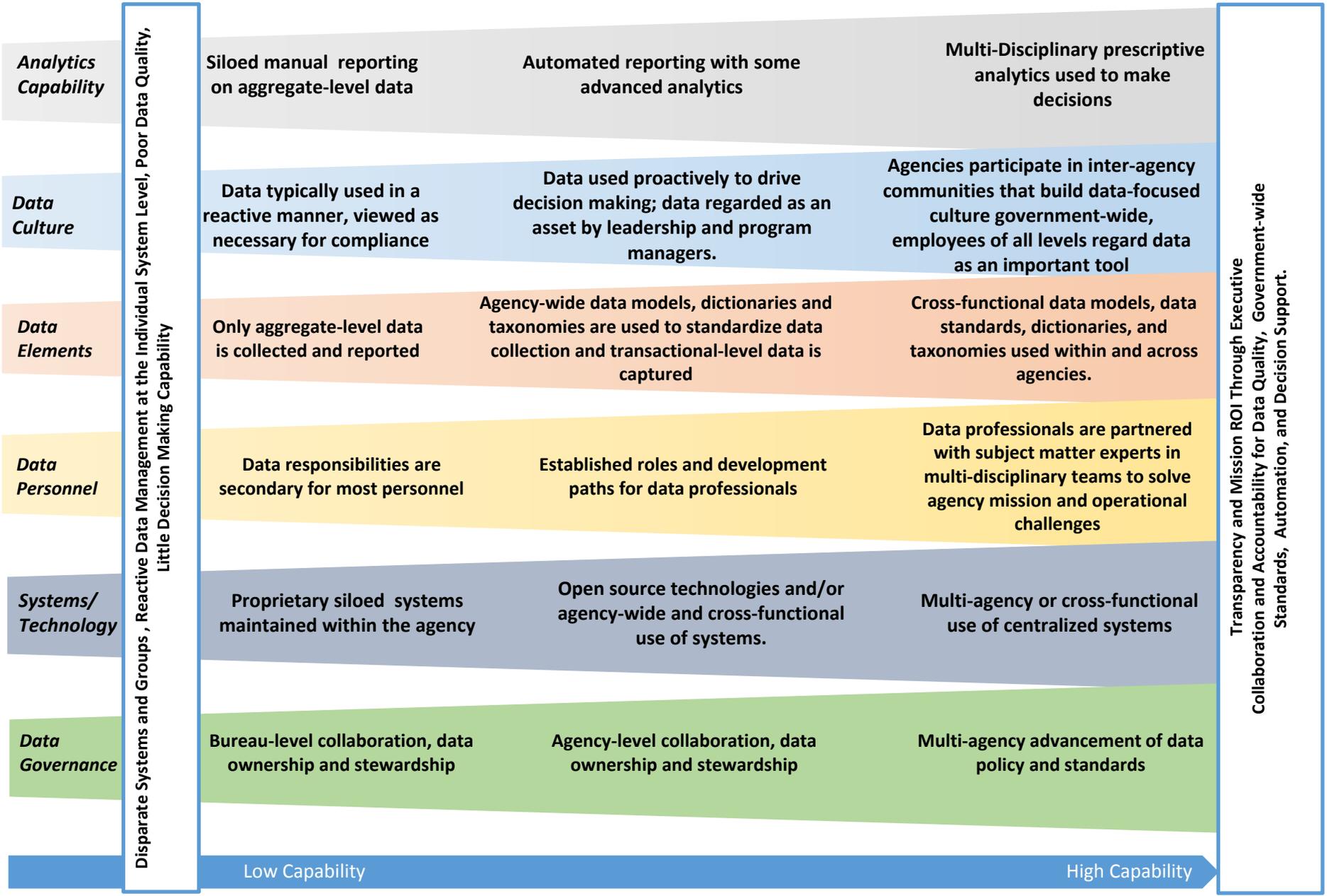
- **Data Cabinet** – Community of Practice to advance adoption of data science best practices (i.e., maturity model)
- **NSTC Data Science subcommittee** – Senior data leaders prioritize the critical science and technology areas of investment for building data capacity.

## Agencies

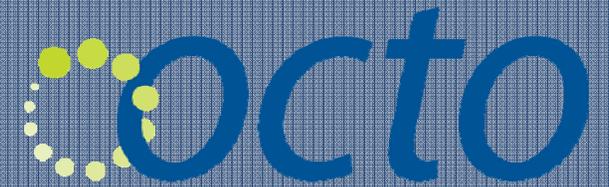
*Maximizing the value of data and ensuring it is responsibly accessed in the most meaningful ways.*

- **Mission imperatives** – Identify data needs against mission priorities (i.e., maturity model)
- **Build data capacity** – Adopt data science best practices to drive innovative culture (i.e., CDO role, data resource management)

# Building Federal Data Capacity



Transparency and Mission ROI Through Executive Collaboration and Accountability for Data Quality, Government-wide Standards, Automation, and Decision Support.



DC Interagency Data Team Mission & Governance,  
Barney Krucoff, Chief Data Officer



## Mission



My go-to mission statement is (same as DC GIS almost),

“Improve the quality and lower the cost of District services, through collective investment in and effective application of ~~geospatial~~ data and systems; and reach beyond District government by making data freely and publicly available to the fullest extent possible in consideration of **safety**, privacy and security.”

Unit	Data Curation	Data APIs & Application Development	Data Visualization & Analysis (Customer Service)	Enterprise Mapping Data
<b>Manager</b>	Mario Field	Julie Kanzler	Michael Bentivegna	Matthew Crossett
<b>Role</b>	Collect, review, document, store, and archive datasets. Design and administer database of record. Manage data flows from agency systems including automated extract, transform, and load functions and leverage OCTO's Service Oriented Architecture. Manage the data.dc.gov and data.in.dc.gov catalogs.	Develop and maintain APIs that allow the District to manage data and systems as distinct assets. Develop reusable user interfaces for maps and dashboards. Take the overall lead on system architecture and operations.	Serve as the primary contact point for agency customers including project and data intake. Manage most external communications with agencies and the public. Lead the GIS steering committee and create a similar interagency user group for BI and urban modeling.	Builds and maintains core dataset. These datasets often serve as the foundation on which other high quality dataset are built.
<b>Primary liaison to</b>	<ul style="list-style-type: none"> <li>• IT-to-IT data transfers and flows</li> <li>• OCTO Enterprise SOA team</li> <li>• OCTO Enterprise DBAs</li> </ul>	<ul style="list-style-type: none"> <li>• Civic hackers</li> <li>• DC Gov developers</li> <li>• External developers</li> </ul>	<ul style="list-style-type: none"> <li>• All agencies not covered by the Foundational Data Development Team.</li> <li>• OCA performance management</li> <li>• Agency data analysts (GIS &amp; BI)</li> </ul>	<ul style="list-style-type: none"> <li>• Part of OCFO (Real Property)</li> <li>• Part of DCRA (Office of the Surveyor, Permitting)</li> <li>• DDOT</li> <li>• OUC (911 and 311)</li> </ul>



# A plan (It is not original)



## What DC GIS Has

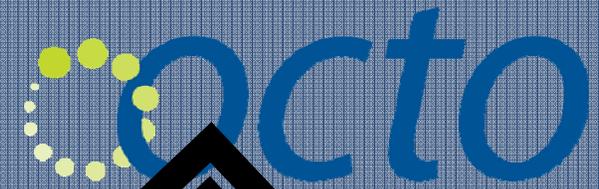
- DC GIS Steering Committee
- Standard practices for interagency data documentation, discovery and services
- A standard technology stack
  - Efficient purchasing
  - Standardized support training
  - Peer-to-peer support
  - Centralized server side management
- Centralized funding model that allows new uses to onboard quickly with no transaction costs

## Other data types

- No interagency user committee
- No dominant common data store (although CDW is closest)
- Multiple technology investments with little IT centralization
  - Some OCTO support
  - No citywide training or organized peer- to-peer interaction
- No central funding



# DC Interagency Data Team

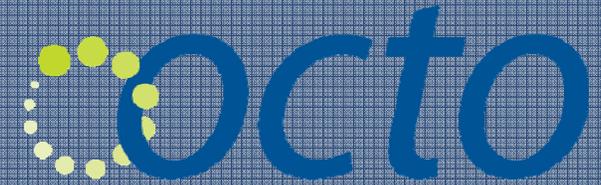


- Create a community of data analysts
- Have a say in citywide investments
- Ensure the right BI tools meet the requirements of your agency needs
- Inventory and classify datasets
- Prioritize datasets for sharing and open publication.
- Coordinate the redaction/generalization and publication and s of datasets
- Assist in the development, implementation, and sharing of data standards and related best practices
- Ensure the presence and accuracy of the metadata





## Lessons from DC GISSC



### **The DC Interagency Data Team meetings will:**

- Occur regularly (approximately every 6 weeks to start)
  - **Next meeting, here, December 14, 1:00 p.m.**
- Be announced publically and almost always allow outside participation
- Always have an agency presentation (show and tell)
- Always have a data report
- Have bylaws but won't often vote

### **Outside of meetings we will have:**

- An email group
- Collaboration tools
- A web page

The District of Columbia

[Website](#)

District of Columbia Custom Web Map Applications


Sort by
Sign In

The District of Columbia offers several custom online mapping applications highlighting data and information from various fields of [More](#)

 <b>Sex Offender Registry</b> <small>Web Mapping Application</small>	 <b>Election Candidate Contributio...</b> <small>Web Mapping Application</small>	 <b>goDCgo</b> <small>Web Mapping Application</small>	 <b>Transportation Online Permitti...</b> <small>Web Mapping Application</small>	 <b>Seasonal Leaf Collection</b> <small>Web Mapping Application</small>	 <b>Permit Status Map Viewer (DC...</b> <small>Web Mapping Application</small>
 <b>Property Information Verificati...</b> <small>Web Mapping Application</small>	 <b>DC Circulator Route Map</b> <small>Web Mapping Application</small>	 <b>Dept of Environment Waterse...</b> <small>Web Mapping Application</small>	 <b>DCPS School Profiles</b> <small>Web Mapping Application</small>	 <b>Business Incentive Map</b> <small>Web Mapping Application</small>	 <b>Connect DC Tech Locator</b> <small>Web Mapping Application</small>
 <b>Police Service Area Finder</b> <small>Web Mapping Application</small>	 <b>DC Real Property</b> <small>Web Mapping Application</small>	 <b>Retail Site Opportunities in DC</b> <small>Web Mapping Application</small>	 <b>Public Space Permit Locator</b> <small>Web Mapping Application</small>	 <b>DC Atlas Plus</b> <small>Web Mapping Application</small>	 <b>Fire and EMS Locations in DC</b> <small>Web Mapping Application</small>
 <b>DC Crime Map</b> <small>Web Mapping Application</small>	 <b>DC Open Data</b> <small>Web Mapping Application</small>	 <b>District Environmental Maps</b> <small>Web Mapping Application</small>	 <b>Trash and Recycling Pick Up D...</b> <small>Web Mapping Application</small>	 <b>Snow Response Reporting Sys...</b> <small>Web Mapping Application</small>	 <b>DCPS Boundary Information S...</b> <small>Web Mapping Application</small>
 <b>DC Location Search</b> <small>Web Mapping Application</small>	 <b>Oblique Aerial Images For DC</b> <small>Web Mapping Application</small>	 <b>Property Quest</b> <small>Web Mapping Application</small>			

Maryland.gov Phone Directory State Agencies Online Services

## MD iMAP

MARYLAND'S MAPPING & GIS DATA PORTAL

[HOME](#) [DASHBOARDS](#) [MAPS](#) [DATA](#) [INITIATIVES](#) [INNOVATE](#) [SUPPORT](#) [TRAINING](#) [PARTNERS](#)

### Dashboard Gallery

- [Overview](#)
- [FAQ](#)

### Dashboard Gallery

Dashboards combine graphs, charts, tables and maps to help users explore and understand complex data sets. In cooperation with our partners [Salisbury University's Eastern Shore Regional GIS Cooperative \(ESRGC\)](#) and [Towson University's Center for Geographic Information Systems \(CGIS\)](#), the [Maryland Department of Information Technology \(DoIT\)](#) will help Maryland state agencies develop and host dashboards. The primary requirement is the underlying data sets are shared as open data through [Maryland's Open Data Portal](#) or the [GIS Data Catalog](#).



**Give Maryland Dashboard**

This dashboard displays support related to the Community Investment Tax Credits (CITC), a program that offers 501(c)(3) nonprofit organizations with state tax credits to attract contributions. Use the dashboard to find active nonprofit organizations, learn more about each organization's goals and donate to support efforts in the community. Developed by [ESRGC](#) for the [Maryland Department of Housing and Community Development \(DHCD\)](#) in conjunction with the [Maryland Department of Information Technology](#).

---



**Maryland Smart Energy Investment Dashboard**

This dashboard tracks the Maryland Energy Administration's (MEA) investments and illustrates MEA's contributions to the growth of affordable, reliable, renewable energy and energy efficiency industries in the State of Maryland. Data from this dashboard can be viewed and downloaded from [Maryland's Open Data Portal](#). Developed by [ESRGC](#) for the [Maryland Energy Administration \(MEA\)](#) in conjunction with the [Maryland Department of Information Technology](#).

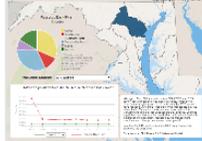
---



**OSPREY Dashboard Public**

This dashboard provides a snapshot of what's going on by Maryland region. The status of power outages, weather, traffic, hospitals and shelters are displayed color coded in an easy-to-digest table. The automated alerts include power, weather, traffic, hospitals and shelters. A mouse click or tap on an alert box will provide details about the alert(s) and relevant links. This is a public facing application, no password is required. Created by [CGIS](#) for the [Maryland Emergency Management Agency \(MEMA\)](#).

---



**BayStat - Causes Map**  
**BayStat - Solutions Map**

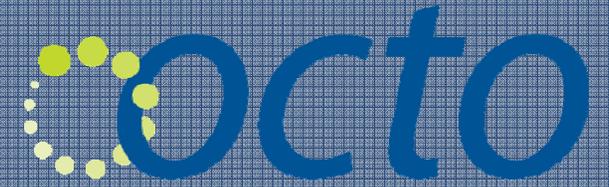
Explore the causes of and solutions to Chesapeake Bay pollution. The Causes Map provides insight into an array of assaults the Chesapeake Bay endures from our actions in the air, water and land. The Solutions Map measures actions and associated goals for reducing pollution, restoring habitats and fostering greener growth and living in Maryland. Developed by [ESRGC](#) and provided by the [BayStat Program](#), a multi-agency initiative.

---

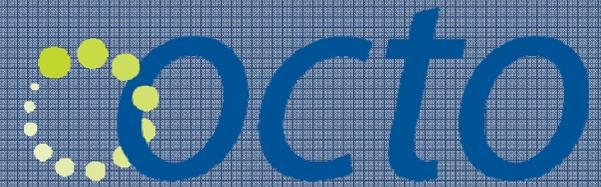


**Maryland Data Explorer**

Ever wanted to know how Maryland stands up against the competition? The Maryland Data Explorer compares Maryland with the other States and Baltimore/Washington, D.C. metro areas



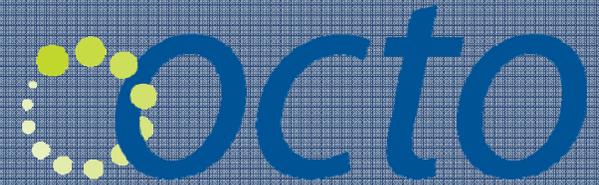
How can an the Interagency Data Team help you in your job?



Draft DC Data Policy  
Barney Krucoff, Chief Data Officer



## Section 1 Intent

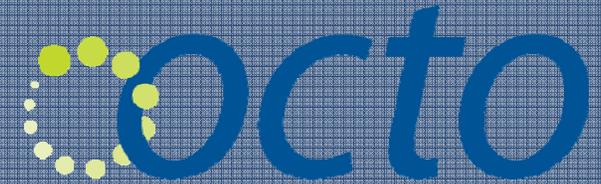


The data created and managed by the District government are valuable assets and are independent of the information systems in which the data reside. As such, the District government shall

- maintain an inventory of its enterprise datasets.
- classify enterprise datasets by level of sensitivity.
- regularly publish the inventory, including the classifications, as an open dataset.
- strategically plan and manage its investment in data.



## Section 1 Intent

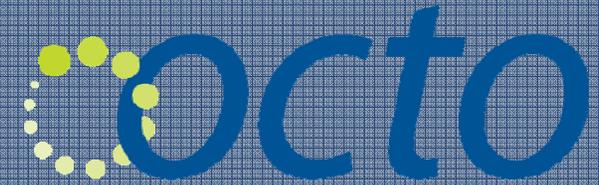


The greatest value from the District's investment in data can only be realized when the data are freely shared among District agencies, with federal and regional governments, and with the public to the fullest extent consistent with safety, privacy, and security. "Shared," means that enterprise datasets 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.
- published online and made available to all at no cost.
- discoverable and accessible.
- documented.
- as complete as can be shared.
- timely.
- unencumbered by license restrictions.
- available in common, non-proprietary, machine-readable formats that promote analysis and reuse.



## Section 1 Intent



By so sharing, the District will

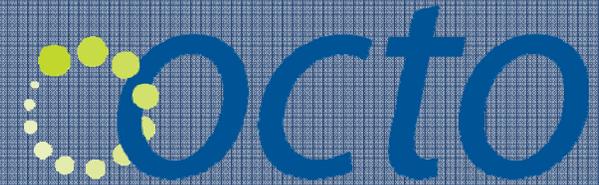
- improve the quality and lower the cost of government operations.
- make government more open, transparent, and accountable.
- enhance collaboration between public bodies, with partner organizations, and with the public.
- further economic development, social services, public safety, and education by making data available to work with and study.

Because inappropriate disclosure of personal information and misuse of data for activities such as identity theft are significant concerns, the District's data must also be managed and responsibly protected. To protect the safety, privacy, and security of residents, workforce members, clients, partners, stakeholders, visitors, and others, datasets requiring protection shall be identified and

- securely stored, transported, and otherwise technically and physically protected against unauthorized access, destruction, modification, disclosure, or loss.
- disseminated only to those persons and entities who reasonably require the information to perform their duties.
- reviewed to determine if useful derivative datasets can be created and publicly distributed by segregating sensitive portions of an enterprise dataset.
- appropriately disposed of or archived when no longer needed.



## Section 3 Definitions



- Information
- **Data**
  - Determination
  - Measurement
  - Transactions
- **Dataset**
- **Enterprise Dataset**
- Metadata
- Freedom of Information Act (FOIA)
- Originating agency
- Dataset classification
- **Automated-anonymization-aggregation-generalization-redaction**
- Subjective redaction
- Mayor's Open Government Advisory Group
- **Dataset classification levels**
  - **Level 4, Restricted Confidential**
  - **Level 3, Confidential**
  - **Level 2, For District Government Use**
  - **Level 1, Public Not Proactively Released**
  - **Level 0, Open**
- Office of the Chief Technology Officer (OCTO)
- Chief Technology Officer (CTO)
- **Chief Information Security Officer (CISO)**
- **Chief Data Officer (CDO)**
- **Agency Information Security Officer (AISO)**
- **Agency Data Analyst (ADA)**
- District of Columbia Data Catalog
- District of Columbia Intranet Data Catalog
- **Enterprise Dataset Inventory**
- **Citywide Domain Tables**
- Creative Commons CC0 Public Domain Dedication



## Section 3 Definitions



**Data** means a subset of information that is statistical, quantitative, or qualitative. Data are regularly maintained by, created by or on behalf of, and owned or licensed by a covered governmental entity in non-narrative, alphanumeric, or geospatial formats. Data are an asset independent of the systems in which they reside.

Data shall include

- records of determinations, measurements, or transactions made by covered governmental entities and related to the mission of those entities.
- geographic image files including aerial photography.
- sensor data.

Data shall not include:

- draft versions.
- non-geographic image files, such as designs, diagrams, drawings, photos, or scans.
- documents, emails, messages, videos, recordings, or hard-copy records.



## Section 3 Definitions



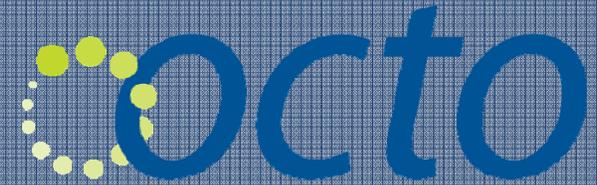
**Enterprise dataset** refers to a dataset that directly supports the mission of one or more public bodies. Typically, enterprise datasets are stored in named information technology systems. For example, the District’s general ledger is a dataset hosted in the “System of Accounts and Records (SOAR).” Typically, such named systems and the datasets they contain are accessible to multiple workforce members. Any named system may hold one or more enterprise datasets.

Enterprise datasets also exist in small systems and spreadsheets. Any dataset, even a spreadsheet, is an enterprise dataset if it currently is maintained and

- is (or has been) used in decision making, or
- documents a public body’s performance, finances, assets, or activities, and
- is not largely duplicative of data in an inventoried enterprise dataset.



## Section 3 Definitions

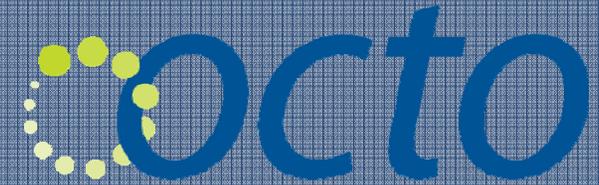


**Dataset classification levels** are defined as the following:

- **Level 4, Restricted Confidential** refers to datasets for which the originating agency has determined that unauthorized disclosure could potentially cause major damage or injury, including death, to residents, agency workforce members, clients, partners, stakeholders, or others identified in the information, or otherwise cause major harm to the agency. Includes any dataset classified by a federal agency or cooperating District agency as “Confidential,” “Secret,” or above.
- **Level 3, Confidential** refers to information for which the originating agency has determined that unauthorized disclosure could adversely affect residents, agency workforce members, clients, partners, stakeholders, or others identified in the information. Includes datasets designated by a federal agency or cooperating District agency as “Attorney-Client Privileged,” “Sensitive but Unclassified,” “Critical Program Information,” “Criminal Justice Information,” “Health Insurance Portability and Accountability Act (HIPAA) Protected Health Information Definition,” “Law Enforcement Sensitive,” “Payment Card Information,” “Family Educational Rights and Privacy Act (FERPA), “For Official Use Only” (FOUO), or “Personally Identifiable Information” (PII).



## Section 3 Definitions

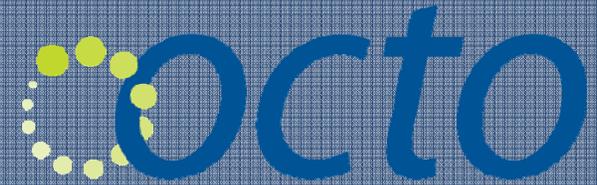


### Dataset classification levels cont..

- **Level 2, For District Government Use** refers to information that the originating agency determines is subject to one or more FOIA exemptions but is not highly sensitive and may be distributed within the government. \
- **Level 1, Public Not Proactively Released** refers to information that may not be protected from public disclosure under any other law. Nevertheless, publication on the public Internet and exposure to search engines would
  - jeopardize the safety, privacy or security of residents, agency workforce members, clients, partners, or anyone else identified in the information.
  - require subjective-redaction.
  - impose an undue financial or administrative burden on the agency.
  - expose the District to litigation or legal liability.
- **Level 0, Open** refers to all enterprise datasets not designated by an agency as being level 1 to level 4. Moreover, any dataset regularly published in machine-readable format on [opendata.dc.gov](https://opendata.dc.gov) or another [dc.gov](https://dc.gov) website prior to this order is “Level 0, Open” unless an agency makes a proactive determination to raise the classification.



## Section 4 Definitions



### **Dataset classification levels cont..**

- **Level 2, For District Government Use** refers to information that the originating agency determines is subject to one or more FOIA exemptions but is not highly sensitive and may be distributed within the government. \
- **Level 1, Public Not Proactively Released** refers to information that may not be protected from public disclosure under any other law. Nevertheless, publication on the public Internet and exposure to search engines would
  - jeopardize the safety, privacy or security of residents, agency workforce members, clients, partners, or anyone else identified in the information.
  - require subjective-redaction.
  - impose an undue financial or administrative burden on the agency.
  - expose the District to litigation or legal liability.
- **Level 0, Open** refers to all enterprise datasets not designated by an agency as being level 1 to level 4. Moreover, any dataset regularly published in machine-readable format on [opendata.dc.gov](https://opendata.dc.gov) or another [dc.gov](https://dc.gov) website prior to this order is “Level 0, Open” unless an agency makes a proactive determination to raise the classification.



## Section 4 Personnel and Roles



**Chief Information Security Officer (CISO)** means the senior official reporting to the CTO who has overall responsibility for the District's information protection strategy and practices.

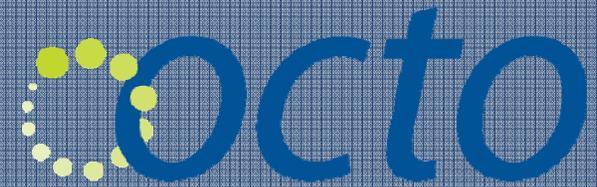
**Chief Data Officer (CDO)** means the senior official reporting to the CTO who has overall responsibility for the District's data governance processes, including the collection, creation, maintenance, documentation, dissemination, and archiving of high-quality, highly interoperable datasets.

**Agency Information Security Officer (AISO)** means an employee, designated by a District agency head, who, in coordination with the CISO, ensures that the agency is implementing this policy.

**Agency Data Officer (ADO)** means an employee, designated by an agency head, who, in coordination with the CDO, ensures that the agency is implementing this policy.



## Section 5 Enterprise Dataset Inventory, Classification, and Prioritization



1. Within 60 days of the issuance of this order, OCTO shall provide public bodies with an intranet-based data inventory tool and train ADOs and AISOs on its use.
2. Within 180 days of receiving the tool, public bodies shall inventory and classify their enterprise datasets using the online tool.
3. Where enterprise datasets are not classified as Level 0: Open, an explanation for the higher classification shall be included in the inventory.
4. Within 270 days of the issuance of this order, OCTO shall publish the first iteration of the Enterprise Dataset Inventory as Level 0, Open.
5. Public bodies and OCTO shall update the Enterprise Data Inventory continuously as new datasets are discovered, created, or archived. The enterprise inventory process will be holistically revisited annually for publication on or before November 1 of each year reflecting a snapshot of inventory as of September 30.
6. By November 1, 2017, the published Enterprise Dataset Inventory shall include prioritization by public bodies for Level 0, Open, datasets within the new fiscal year beginning October 1.

# Data Assessment

The screenshot shows a web browser window with the URL `localhost:36630/Assessment/DataInventory.aspx`. The page header includes the logos for 'THE DISTRICT OF COLUMBIA' and 'Office of The Chief Technology Officer', along with an 'Agency IT Dashboard' logo. The user is logged in as 'Bikkumalla, ManiKumar (OCTO-Contractor)' and the fiscal year is set to 'Fiscal Year 2017'. A navigation bar contains buttons for 'Logout' and 'Home'. Below the navigation bar is a process flow diagram with eight steps: 1. IT Spending, 2. IT Staff, 3. Infrastructure, 4. Applications, 5. Data Inventory (highlighted in blue), 6. Risks, 7. Services, and 8. Initiatives. The main content area is titled 'Data Inventory Details' and features a 'Run Report' button. The form contains the following fields:

- Dataset Name:**
- Database Name:**
- Dataset Classification:**
- Data Owner:**
- Linked Application:**
- Dataset Access:**
- Dataset Category:**
- Data Classification Reason:**

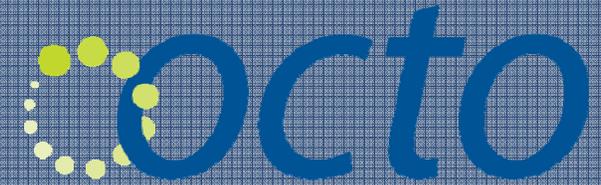
DRAFT CONTENT OF DATA ASSESSMENT YEAR 1	
ATTRIBUTE	DESCRIPTION
DATASET_ID	Unique ID
NAME	Name/Title/Brief Description
ABSTRACT	Summary of dataset
WEB_URL	Link to more information
KEYWORDS	Descriptive words
TYPE	Storage/file type
CATEGORY	Classification using Existing Categories
NOTES	Caveats/Business Rules/Notes
APPLICATION	Application(s) using data
ORIGINATOR	Source Agency
ACCESS	Access level
ACCESS_REASON	Reason for access level
CREATION_DATE	Date data was created/implemented
LAST_MODIFIED_DATE	Last modified date of data
UPDATE_FREQUENCY	How often dataset is updated

## Data Updates 6/16-9/16

- Bicycle Lane
- Bus Line
- Camera Enforcement Sites
- Campaign Finance
- Certified Business Enterprise
- Clean Team Area
- Crash
- DMPED Real Estate Portfolio
- DMPED Industrial Revenue Bond
- DMPED Affordable Housing
- Downtown Designated Street
- Downtown Designated Subarea
- Federal City (DRAFT)
- Great Street
- Hospital
- Hospital Grounds
- Impervious Surface – 2015
- Liquor License
- Metro Entrances
- Washington DC Boundary (DRAFT)
- Notary Public
- Railroad
- Recreation Facilities
- Technology Zone
- Triangle and Medians
- Water Surge
- Zoning



Remaining Sections



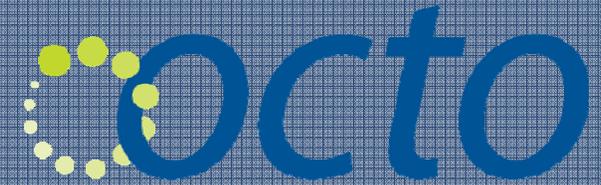
**Section 6. Minimum Data Protection Requirements**

**Section 7. Data Catalogs**

**Section 8. Intra-District Data Sharing Agreements for Datasets Classified Level 2 and Below**

**Section 9. Nexus Between FOIA and Level 0, Open Datasets**

**Section 10. Level 0, Open, Data Legal Policy and Licensing**



Open Discussion  
Complete Review Survey

Next Meeting

1:00 to 3:00

December 14, 2016

OCTO

200 I Street SE