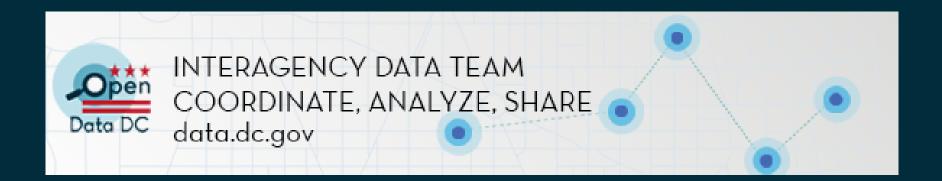


# Interagency Data Team November 6, 2018







# WELCOME, AGENDA DATA TEAM NEWS

Michael Bentivegna Data Visualization and Analysis Program Manager Office of the Chief Technology Officer

#### :octo

#### CUSTOMER SERVICE. ACCOUNTABILITY. EFFICIENCY. SECURITY.

### AGENDA

#### Welcome, News & Updates

Michael Bentivegna, Data Visualization and Analysis Program Manager Office of the Chief Technology Officer

#### Data Science Working Group

Peter Casey, Senior Data Scientist Office of the Chief Technology Officer, Lab @ DC

- FY19 Agency Data Submission Process for Level O Datasets Mario Field, Data Curation Program Manager Office of the Chief Technology Officer
- Enterprise Data Inventory A Preview of New System in Testing Michael Bentivegna, Data Visualization and Analysis Program Manager Office of the Chief Technology Officer





#### CUSTOMER SERVICE. ACCOUNTABILITY. EFFICIENCY. SECURITY. BUSINESS INTELLIGENCE SOFTWARE UPDATES

#### Tableau Production version 2018.1.1

- Google sheets as a data source in production
- Major upgrade to move to 2018.2
  - Hold off on Desktop upgrades
- Architecture review session
- Tableau Public w/automated ETL updates
  - Hyper not currently supported by Informatica Tableau extension at this time
  - Box plug-in will be tested as a public data source





#### CUSTOMER SERVICE. ACCOUNTABILITY. EFFICIENCY. SECURITY. BUSINESS INTELLIGENCE SOFTWARE UPDATES

- MicroStrategy Production Version 10.11
  - Moving MicroStrategy Web servers from Linux to Windows (WebLogic to IIS)
  - Creating public facing MicroStrategy web servers
  - Developing MicroStrategy integration guidelines for imbedding analytics in Agency websites and SharePoint.
  - New physical servers with more CPU power and memory is being deployed as the MicroStrategy Production servers



#### : OCto

### CUSTOMER SERVICE. ACCOUNTABILITY. EFFICIENCY. SECURITY. GIS SOFTWARE DC AGOL – ARCGIS ONLINE/PORTAL

### ArcGIS Server Upgraded to version 10.6

 Some WMS map service format issues reported in Tableau

### ArcGIS Portal version at 10.6.1

- Reach users by sharing maps imbedded in web pages and through web applications
- Easily Add DC GIS Data
- Leverage ready to use apps, Story Maps, Survey tools and data collection applications.
- Use custom templates
- Easy access to Mobile
- Create Map/BI Galleries
- Data Initiatives Pages (http://opendata.dc.gov/)
- Hybrid of Cloud/DC GIS Infrastructure
- Included with citywide Esri ELA

http://dcgis.maps.arcgis.com https://dcgis.dc.gov/



#### :octo

#### CUSTOMER SERVICE. ACCOUNTABILITY. EFFICIENCY. SECURITY.

# **BI TRAINING**

#### • **Tableau Intro & Intermediate** (Eventbrite – search Tableau)

- Introductory training
  - December 5 and 6 (FULL)
- Intermediate training
  - December 12 and 13 (two seats left)
- MicroStrategy Intro Classes (Eventbrite search MicroStrategy)
  - November 27 28 (FULL)
  - December 18 19 (eleven seats left)



#### **Octo**

CUSTOMER SERVICE. ACCOUNTABILITY. EFFICIENCY. SECURITY.

# GIS TRAINING

- GIS (PeopleSoft request training Search by name "GIS" or course number)
  - 230 Overview DC GIS Services February 6
  - 233 ArcGIS Online January 16





# DATA SCIENCE WORKING GROUP

Peter Casey, Senior Data Scientist Office of the Chief Technology Officer Lab @ DC





# DATA SCIENCE @ DC

- A monthly meeting of DC government employees interested in data science hosted by OCTO and The Lab @ DC
- Topics:
  - State of Data Science in DC Gov and as an industry
  - Issues in Data Science, such as ethics and fairness
  - Challenges to Data Science in DC Gov
  - Projects across DC Gov
- Next Meeting: Wednesday, November 14
- Interested? Peter.Casey@dc.gov





# OPEN DATA THAT IS NOT ON THE OPEN DATA PORTAL...YET!!!

#### Mario Field, Data Curation PM Office of the Chief Technology Officer





### COUNT BY AGENCY (SAMPLE)

Agency	Document	Open Data	Num	•	1
DDOT		2	19		
DMV			13		
DPR			13		
OVSJG			8		
OCTO		13	8		
DOH		1	7		
DCPS			7		
DCRA		1	7		
DGS			5	•	F
DBH	3	1	4		C
OCP	1	2	4		(
DPW		4	4	•	(
DCOZ		2	4		
DCPL			4		

Numbers TOTAL - 189 DOCUMENT - 15 DUPLICATE - 32 ELIGIBLE - 142

Records marked as Documents and Duplicates are cursory OCTO review.

**Covers 40 Agencies and Offices** 



#### **i octo**

#### CUSTOMER SERVICE. ACCOUNTABILITY. EFFICIENCY. SECURITY.

# BY AGENCY (TOTAL)

Agency Open Datasets									
DDOT 21	DBH DCRA 8 8				DPW 8		01	OVSJG 8	
осто 21	DCPS 7 OCP 7		ABF 4 CF		DCPL 4	DMI 2	DMPED 4	EOM 2	OCF 4
DMV 13	,		oc			2		2	
	DCOZ 6		3		0A# 2	A	SERVE 2		
DPR 13			0	P 3	OLA 2		LGBTQ		
	DGS 5		BE	GA	SBO 2	E	MOLC OCA	PSC	SCDC





### BY CATEGORY THEME

#### Dataset Category

Government Operations38Transportation34Health19Business and Economic D14Public Safety12Education12Public Services10Planning Land Use and Zo9Aerial Photography and S9Location7Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Environment2Historic1	Dataset Category	
Health19Business and Economic D14Public Safety12Education12Public Services10Planning Land Use and Zo9Aerial Photography and S9Location7Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Government Operations	38
Business and Economic D14Public Safety12Education12Public Services10Planning Land Use and Zo9Aerial Photography and S9Location7Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Transportation	34
Public Safety12Education12Public Services10Planning Land Use and Zo9Aerial Photography and S9Location7Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Health	19
Education12Public Services10Planning Land Use and Zo9Aerial Photography and S9Location7Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Business and Economic D	14
Public Services10Planning Land Use and Zo9Aerial Photography and S9Location7Administrative and Other7Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Public Safety	12
Planning Land Use and Zo9Aerial Photography and S9Location7Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Education	12
Aerial Photography and S9Location7Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Public Services	10
Location7Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Planning Land Use and Zo	9
Administrative and Other7Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Aerial Photography and S	9
Recreation5Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Location	7
Utility and Communication4Property and Land2Financial2Facility and Structure2Environment2	Administrative and Other	7
Property and Land       2         Financial       2         Facility and Structure       2         Environment       2	Recreation	5
Financial2Facility and Structure2Environment2	Utility and Communication	4
Facility and Structure2Environment2	Property and Land	2
Environment 2	Financial	2
	Facility and Structure	2
Historic 1	Environment	2
	Historic	1



# Data Submission Guide – Draft -Briefing

NOVEMBER 6, 2018

#### OCTO Data Division – Curation Program



# BACKGROUND

1. Provides transparency and accountability for DC's data program

2. Allows agencies and the public to understand the overall data curation process.

3. Documents the requirements and guidelines needed for submission to ensure the integrity and consistency of data

4. Balances openness and transparency with privacy and security protections

# Research

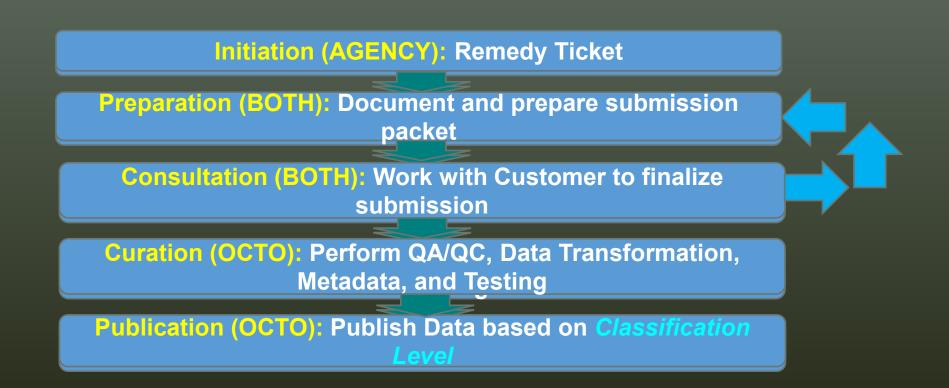
**Review other jurisdictions** 

- New York State
- New York City
- Chicago
- •Seattle
- San Francisco
- Maryland

Findings

- New York State has the most comprehensive documentation (roles, responsibilities, processes, dataset rules, etc.)
- Others provided more overall/general processes and steps.

# OVERVIEW



# INITIATION

Request For	Mario Field				
Phone	(202) 727-1761				
Email	mario.field@dc.gov				
Requested By	Mario Field				
GIS/BI Data and Visualization - Web Map	and Agency Data Publishing				
Requester Information					
Is the phone number listed above correct?	None				
Please select your Agency/Department/Location (use search to find location or address)	S -				
Request Details					
Publish new, refresh or agency data	None 🔻				
Please click the Submit button to create	your service request				
Add Attachments and Notes:					

Ticket is closed once it is assigned to a data analyst

# PREPARATION

#### Data Submission Packet (DSP)

- Remedy Ticket
- Data Submission Form
- Data Dictionary
- Data Mapping Document
- Approval Form
- Data Review Summary Report
- Additional Documentation

# DATA SUBMISSION FORM

- Stored in Quickbase
- Information Headers
  - Dataset
  - Classification
  - Transformation
  - Metadata
  - Contact
  - Miscellaneous
  - Internal
- Integrated with EDI

Field Name	Data Type	Damain	Description
5414/710			Unique Data Identifier
DATASET ED ID	integer Integer		Unique Data Identifier Enterprise Data Inventory ID
	- and the		and provide an electric state
	Rocken		Indicate A Dataset Exists in EDI
NATASET EDI NAME	String		Dataset Name
DATASET EDI CLASSIFICATION	String	EDI_CLASSFICATION	Dataset Classification Level
DATASET EDI SQURCE FORMAT	String	EDU_SOURCE_FORMAT	Dataset Source Format
DATASET EDI_DESCRIPTION	String		Dataset Description
DATASET SECURITY CHICK	incian .		
DATASIT SECURITY CHECK			Indicate Security Check Requirement
	String	DATASET_TASK_THPS	Task Types
DATASET TASK DETAILS	String		Task Detail Notes
DATASET APPLICATION DEPENDENCY	String		Dataset Application Dependency/Note
MTASET APPLICATION DEPENDENCY MTASET PRIMARY KEY	String		Dataset Application Dependency Note Dataset Primary Key Field
SUINELPHINKI MI	a strong		
	Inceper	EDI LEVEL	Dataset Security Classification Level as Scored in EDI
LASSIFICATION EDI REASON	String		Dataset Classification Reason
LASSIFICATION DERVITIVE	Regiven		Indicate Dataset Is Derivative
CLASSIFICATION METHODOLOGY	String		If Yes to Derivative, Notes on How to Process Dataset
CLASSIFICATION SECURITY	String		Additional Security Notes
SEGEPATINE PROJECTION	String	PROJECTION	Geospatial Projections
SEOSPATIAL SEOCODE	Rodiran		indicate if Geocoding is Required
SECSPATHE OUTSIDE OC	Rodivan		Indicate Locations Outside DC Skint
GEOSPATIAL GEOMETRY TYPE	Spring	GEOMETRY TYPE	Geometry Type
SEOSPATIAL TOPOLOSY	Bodiean		Indicator If Topology Check is Required
TL MITHODOLOGY	String	FTL_METHODOLOGY	\$TL Methodology for Satracting the Data from Source
TTL UPDATE FREQUENCY	String	UPDATE_FREQUENCY	ETL Update Frequency
TTL DECONSISTION DEMANS	String		ETL Database Name
TL DECONNECTION SCHEMMANNE	String		ETL Ostabase Schema Name
TL DECONNECTION TABLEMANE	String		ETL Catabase Table Name
ITL DECONNECTION USERNAME	String		ETL Database Username
T1 881 WED SERVICE 101	Spring		CTL addresses Cervice Ltd
TL FLESHARS LOCATION	Spring		ETL File Share Location
ETL EMML ATTACHMENT SUBJECT	String		671: Small Subject
IT INVERTIGATION STREET	sound		NIC NICH SUBJECT
	Soring		ETL Email Attachment Filename
CTI DANII ATTACAMENT DANIIADARISS	String		ETL Email Address
EL OTHER	String		Other FTL Type
			ALL IN 1997
	Spring	AGENCY NAME	Americy That Submits the Cata
METADATA EN OWNER	Sting		Agency that Dwiss the Data
METADATA GRI REVARSEDS	String		Keywords (Or Phrases) For Searching Purposes
METROATA CREDIT	Spring		A Recognition of Who Created or Contributed to The Dataset
METROATA EDI DATASET CATEGORY	Soring	DATAGET_CATEGORY	General Category Within Which the Dataset Should Be included in
METADATA CESATION DATE	String		Date the Data Was Created
METADATA REVISED DATE	String		Date the Data Was Revised
METADATA PUBLICATION DATE	String		Date the Data Was Published
METRINATA SYLIDES	String		Data Guerre
	String		Sartiest Date of Available Record
ALTADATA EDI FARUEST RECORDE			
			Recent Date of Available Record
	String		
MITADATA KAI BECKNT BECORDE			
MITANATA IDI ASCATI FACORIS	Soing		The URL To Source Link or Dataset Program Web Page
MITADITA EDI SECRIT RECORDE MITADITA EDI DATATIT URI MITADITA EDI UPARTI FURINA	Soing	UPDATE_FIEQUINCY	Data Update Frequency
MINANA ISI KKAT KKORM MINANA ISI AMATI UK MINANA ISI UPAN MINAN MINANA ISI UPAN INTINA KINA	Soing Soing Soing	uParts_FROUDINCY FURIC_INTERST	Data Update Frequency Public Interest in Dataset
MINANA NI AKINY AKINY AKINY MINANA NI AKINY MI MINANA NI AKINY MINANA MINANA NI AKINY MINA	Soing Soing Soing Soing	uparte frequincy Fuelic, intrest	Data Update Frequency Public Interest in Dataset Dataset Web Address
MICHARN BE ADDY AND	Soing Soing Soing Soing Soing	PUBLIC_INTEREST	Data Update Frequency Public Interest In Dataset Dataset Web Address Contact Name
Mittanta Britani Mittanta Britani Mittanta Britani Mittanta Britani Mittanta Britani Mittanta Britani Ostanta Mari Ostanta Mari	Soing Soing Soing Soing	UPDATE FREQUENCY PUBLIC_INTERST AGINC_INVASE	Data Update Frequency Public Interest in Dataset Dataset Web Address
Ministra de Anador Ministra     Ministra de Anador Ministra      Ministra     Miniitra     Ministra     Ministra	Soing Soing Soing Soing Soing	PUBLIC_INTEREST	Data Update Frequency Public Interest In Dataset Dataset Web Address Contact Name
WINANA AR KINY KIONY WINANA DI ANYA KIA WINANA DI ANYA MARKA WINANA DI ANANA WINANA DI ANANA DI OMAYA MMI COMAYA MMI	Soing Soing Soing Soing Soing Soing Soing	PUBLIC_INTEREST	Data Updana Fingunoy Public Intervet in Dataset Dissanst Win Address Contast Name Contast Name
WINNER AF VERM KOME WINNER AF VERM AN WINNER AF VERM AN WINNER AF VERM AND WINNER AF VERM AND WINNER AF VERM AND OWNER AND OWNER AND OWNER AND OWNER AND OWNER AND	Song Song Song Song Song Song Song Song	PUBLIC_INTEREST	Data Updato Responsey Adala Interest in Datatet Dataset Web Address Cartas Name Cartas Rama Pinca Nucubar Enal Address
Alchert, & Hort Hondo Alchert, & Hort Hondo Herner, H. Hondo He	Soing Soing Soing Soing Soing Soing Soing	PUBLIC_INTEREST	Bata Lipidate Frequency Public Interact in Dataset Contant Web Advense Contant Agency Ponon Number

# CONSULTATION

Review DSP with Agency

# CURATION

- Data Processing
- QA/QC
  - Data Rules
    - Data Format (5)
    - Data Field (12)
    - Addressing (3)
    - Miscellaneous (3)

- Metadata
  - Description
  - Topic and Keyword
  - Citation
  - Resource Details
  - Resource Maintenance
  - Resource Constraint
  - Spatial Reference
  - Resource Quality
  - Resource Lineage
  - Distribution
  - Entity Attribute
  - Metadata Reference

# SECURITY

<u>Database level</u> – When data is loaded into a database, it gets placed in a secure schema with separate login credentials. Furthermore, the database can be encrypted to protect content.

<u>Map service level</u> – When the data is published in a map service, the software allows for security roles and users. The agency would define the role and provide a list of users and applications that can access the data.

<u>Data Processing level</u> – There are ways to process the data that create derivatives that mask individuality. Aggregation is the most common where data is presented in geographic areas or blocks that simplify and veil the data. Summarization is also a way to anonymize the data by combining and presenting individual records as groups or summaries.

Encryption - In select cases, encryption can be used in ETL development.

<u>Other</u> – OCTO possesses several other tools to customize and secure data. This includes Service-Oriented Architecture (SOA) and Application Programming Interface (API) Gateway

## APPROVAL FORM

- Privacy and Security Concerns
- Proposed Mitigation
- District Agency
  - ADO and General Counsel
- OCTO
  - Data Curation Program Manager and General Counsel
- Changes Applied to address concerns

Appendix C – Data Submission Approval Form	
DATASET TITLE	
POTENTIAL PRIVACY OR SECURITY RISKS OR SENSITIVITY ISSUES	
POTENTIAL PRIVACY OR SECURITY RISKS OR SENSITIVITY ISSUES	
PROPOSED MITIGATION	
District Agency or Office	
Agency Data Officer:	Date:
Agency Director (level 2):	Date:
General Counsel:	Date:
Office of the Chief Technology Officer	
Data Curation Program Manager:	Date:
Chief Technology Officer (Level 2):	Date:
General Counsel:	Date:
Recommendation:	
Approved for publication	
□ Changes necessary	
CHANGES NEEDED	
CHANGES MADE	
L	
Department Agency Data Officer:	Date:
Data Curatian Branson Manager	Dates

# PUBLICATION ADDITIONAL STEPS

- Dataset Review
- Archiving
- Error Submission

# ACTION ITEMS

- Develop a formalized, documented data curation workflow
- Finalize approval form and workflow
- Create an online data submission entry form in the same application that manages the EDI and integrate the two
- Develop training materials and a course describing this process
- Create an online mechanism for the public and agencies to submit correction requests



# ENTERPRISE DATA INVENTORY PREVIEW NEW SYSTEM IN TESTING

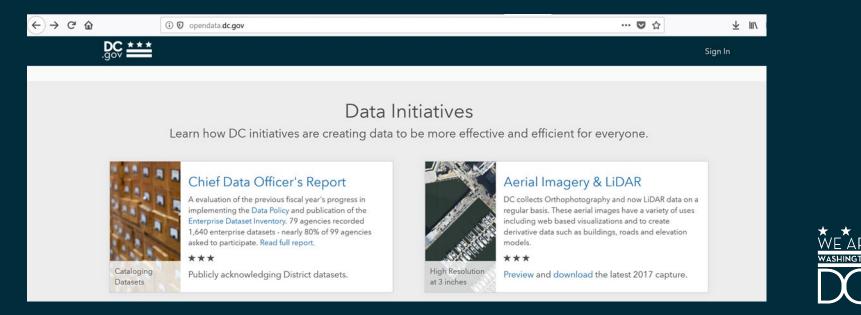
Michael Bentivegna, Data Visualization and Analysis Program Manager Office of the Chief Technology Officer



# ENTERPRISE DATA INVENTORY

Mayors Order 2017-115 : 2018-115

- Requires Agencies to annually review and update their Enterprise Data Inventory
- New application developed to review and update inventory
  - Application tested and in production: on or before Dec. 1<sup>th</sup>
- Agency inventory is due, reviewed by council <del>Dec. 11<sup>th</sup> ightarrow</del> Feb. 11<sup>th</sup>



#### :octo

#### CUSTOMER SERVICE. ACCOUNTABILITY. EFFICIENCY. SECURITY.

• • • • The Dist of Colu	ି ୦C	to			
				👤 Michael Ben	tivegna 👻
Select Agency :	Office of the Chie	ef Technology Officer	T		
Data Inventory	Data Cen	ter Applica	tion Summa	ry	
Show 10 • entri	es		Sea	irch:	
	Dataset Name 🎝	Application Name 1	Dataset Classification 1	Dataset Classification Reason	Data O
~9	theheh-	Services	Level0	theheh-	agency
<ul><li>✓</li><li>∞0</li></ul>	the dATASET	Google Analytics	Level 2: For District Government Use	KFHGHXFGHG	the gov
•0	Telecommunications Request and Approval Tracking		level4	The request include critical infrastructure information about building network infrastructure changes, names of DC Government staff who are traveling and travel dates/locations, equipment orders, etc.	OCTO-
-0	SQA Testing Data for OCA Application Life		level2	This dataset is subject to one or more FOIA exemptions.	осто-

### :Octo

#### CUSTOMER SERVICE. ACCOUNTABILITY. EFFICIENCY. SECURITY.

THE DIST OF COLU	INDIA En	Notes ×		
				Add a new note
Select Agency :	Office of the Chie			
Data Inventory	Data Cen	iter Applicat	tion	
Show 10 • entrie	es			Updated By:
	Dataset Name 🗍	Application Name	Dataset Class	Michael Bentivegna
29	theheh-	Services	Level0 L	11/6/2018 12:00:16 PM
	the dATASET	Google Analytics	Level 2: For Di	Comments : Reviwed by OCTO Liaison
20			Government U	Updated By: Priya Kaja
-	Telecommunications Request and		level4	Updated Date: 11/6/2018 11:52:13 AM
	Approval Tracking			Comments : Dataset is updated
				Updated By: Priya Kaja Updated Date: 11/6/2018 11:40:15 AM
~0	SQA Testing Data for OCA Application Life		level2	Comments : Dataset is updated
	Cycle Management			<b>Updated By:</b> Priya Kaja
•0	Avaya Experience Portal Call Statistics		level1	Updated Date: 11/6/2018 11:38:45 AM



### OPEN DISCUSSION

#### NEXT MEETING SCHEDULED FOR NEXT YEAR

THANK YOU 🙂

