



## Data policy history in D.C.

The District recently established a comprehensive data policy, but it took many years to get there. We hardly think of governments as entities at the forefront of technological innovation, but government data policies can have a big impact by changing perceptions and expectations, invigorating (or stifling) innovation, collaboration and most importantly inclusion.

The District established a new data policy<sup>1</sup> on April 27, 2017 with the stated goal of pushing the D.C. Government towards more efficient usage and sharing of government data. This most recent step follows many others taken in the last decade under four different Mayors, including an earlier policy from 2014 and two previous memoranda from the Executive.

### 2006-2011: The first steps towards a comprehensive data policy

The District's first policy, via a memorandum, on the use of government data was issued in 2006 by then-City Administrator Robert Bobb under Mayor Williams.<sup>2</sup> The guidance contained in this memorandum, while very rigid and inflexible by today's standards, were still ahead of their time. The policy attempted to create a culture of data usage and sharing in and out of the government, with the end goal of giving non-governmental entities the same level access to data that government had via the internal intranet. It would do so through DCStat, a program instituted by the Office of the Chief Technology Officer (OCTO) to help facilitate internal data collection, sharing, and usage. This policy went on to describe a dataset release schedule and a few guidelines such as including release of metadata—descriptions of the data itself—so outsiders can “use [the data] to analyze operations.” While commonplace today, the requirement to release metadata was forward-thinking in the sense that it recognized that the public will need context about how data was created in order to use the data correctly.

After the 2006 policy, however, data policy in D.C. stood still until Mayor Gray's office issued a policy in 2011 with new guidance.<sup>3</sup> The 2011 memorandum went beyond the 2006 proto-policy in its focus on encouraging transparency and public engagement, but it did not provide specific details about how this openness could be implemented.

### 2014: Defining “open data” (and many other terms)

In the years since these first two data policies, the possibilities of open data have been expanded by the development of cheaper data processing tools and there has been an increased focus on data-driven policymaking at all levels of government. In 2014, D.C. issued its first data policy explicitly focused on open government data. This data policy came in the form of an Executive Directive from Mayor Gray and had an effective date of July 21, 2014.<sup>4</sup>

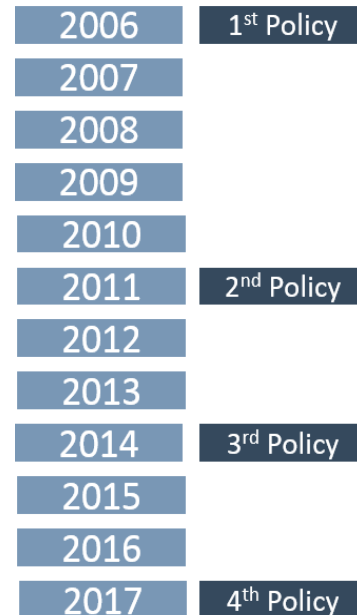


Figure 1. Timeline of signed data policies

This 2014 open data policy identified and attempted to establish some important positions to facilitate a new culture in the government, such as a Chief Data Officer (CDO) within OCTO and Open Government Coordinators in each agency to serve as the CDO's point of contact throughout the government. Among other forward-thinking actions, it suggested the use of open data formats and more fully defined D.C.'s Open Data and FOIA online portals, promoting not just transparency but also public participation. While limited and slow to take root—the CDO role would not be filled for two more years—this was a notable first step in the open data policy world, possibly even the first of its kind<sup>5</sup> among state and local governments in the nation.

## 2014-2017: The road to the second D.C. open data policy

The period between D.C.'s first and second data policies (the latter was signed a month ago) was a period of increasing knowledge within the broader open data community of how data could be collected and used. The next step was to create a new data policy for the District. This new data policy would be wider reaching under the Mayor's office and have a new-found push towards opening up the government and helping the Council pass similar legislation to the policy. OCTO's experiences implementing the previous policies taught the agency that it needed stronger definitions for data classification in order to promote releasing data that should be available, a more expansive list of what data should be released along with the new classifications, and new guidelines on how individuals would be able to access this data.<sup>6,7</sup>

In January 2016, Mayor Bowser and the Chief Technology Officer (CTO) Archana Vemulapalli announced<sup>8</sup> a draft Data Policy as part of a "renewed focus" on open data within the D.C. government. The draft was posted<sup>9</sup> on [drafts.dc.gov](https://drafts.dc.gov), a new platform which allowed public markups. While a press release about the draft boasted that it had around 1,500 visitors and received over 250 "comments and annotations" in its month-long comment period, this still represented the activity of just 35 people or organizations.<sup>10</sup>

In the Summer of 2016, the District hired Barney Kurcuff, an OCTO veteran, as the first Chief Data Officer—the position created back in 2014.<sup>11,12</sup> OCTO then posted an expanded draft of the data policy on [drafts.dc.gov](https://drafts.dc.gov) that included discussion of information sharing and security practices.<sup>13</sup> Individuals and organizations commented on the new draft policy; the Sunlight Foundation, which had tracks data policies around the nation, was invited to OCTO to discuss their concerns in depth.<sup>14</sup> There were also comments from the Mayor's Open Government Advisory Group, a committee with members from both the public and private sector.<sup>15</sup> However, beyond this high level of engagement from a few stakeholders, overall participation from the general D.C. data community was low through the [drafts.dc.gov](https://drafts.dc.gov) page. On the posting for the new data policy, there were just 13 participants (including the CDO), with only 7 comments and 195 notes.<sup>16</sup> Comments ranged from trivial (e.g., noting a broken link) to more complex (e.g., asking how Freedom of Information Act requests and the newly described open data levels would interact). Other interactions with the public came from several meetings with the Code For DC group and some small hackathons to the policy.<sup>17</sup> All of this back and forth went on for several months until public comment period ended in December and the data policy went back to OCTO for the finishing touches.

## 2017: The Mayor signs the second D.C. data policy

On April 27<sup>th</sup>, 2017 Mayor Bowser signed the new D.C. data policy just at the eve of the District's annual technology-focused InnoMAYtion event. The newly signed data policy, with its twelve sections, is more comprehensive and much better developed than the previous data policy and executive memorandums that preceded it. Specifically, this new policy dives deep into how data should be classified, handled, distributed, and continuously assessed.

## Looking at the data policy through the lens of the user

Some organizations have put together checklists<sup>18,19,20</sup> of principles that an data policy should have. Most of these are comprehensive and very thoughtful. As a data user, when I look at data policies, I look for a few simple things:

1. **Data policy should enforce release of data in accessible format.** There is an ever-changing terrain of data formats but some are proprietary, some are not user friendly (PDF), and some are just outdated. Pretty graphs are not open data; a data policy should enforce release of data in formats accessible by those with the slimmest computing capabilities (such as comma delimited files).
2. **Data policy should clearly define roles and accountability.** The user should know who in the government is responsible for implementing the data policy and how they implement the policy required.
3. **Data policy should be clear in classifications.** The user should know what type of data they should always expect to have access to—and if they don't have access to other data, why that is.
4. **Data policy should require full documentation.** Users should always be able to see the data's metadata and background information so that they can fully understand the context of a data set before using it.
5. **Data policy should enforce timely releases and provide easy discovery.** Users should not have to wait unreasonable amounts of time before data is released. It should be clear to users how to search for data that they know about, and they should be able to explore and discover new data sets easily.

There are obviously other aspects of data policies that these five points do not touch on, but this provides a useful starting point. Ultimately, the user needs to feel the data is useful and worth their efforts to find, clean, analyze, and work with. If the data is not truly accessible and usable, the government lacks true data transparency—only the illusion of it. An interesting aspect to the new D.C. data policy is that it is not explicitly an open data policy. It attempts to create one comprehensive data policy that includes data across classifications. This is somewhat unique, most jurisdictions layout several policies for different classifications of data. D.C. has attempted here to create one policy that agencies in the government can follow across their data types.

## What the policy did well

The new policy was thought out, collaborative, and really succeeds in most of the criteria stated above. Going through each criterion it is clear that the CDO and his team tried and mostly succeeded in creating

a data policy that is as open as a government currently can be. The few and far between shortcomings are addressed later, but first it is important to point out what the policy writers did well.

## 1. Accessible format

The District's data policy begins with the stated purpose of changing its culture and practice to view data as valuable assets to the city and its residents. The new data policy states that data, as an asset, can only realize the full value by facilitating the sharing knowledge as freely as possible among the District agencies, federal and regional governments, and the public. That is a great statement, in that the openness through ease of discoverability, timely publication, and availability in common and non-proprietary machine-readable formats, really sets the stage for how data should be published. This differs from the earlier policy where machine-readable has been mentioned but timeliness and discoverability are not. By being able to not only ingest the data but find data as it comes out is essential.

## 2. Clear roles

On this count, the District's policy does a good job of defining the roles of the CDO, an Agency Data Officer (ADO, similar to the Open Government Coordinator of the previous policy), and different security officers. These positions should hopefully allow for quick and easy collaboration in the pipeline between the data collection, through the aggregation of the data, and to the security and classification officers who allow the appropriate data to be released.

The personnel and roles section further contains useful and straightforward definitions of other relevant roles. The Chief Information Security Officer (CISO) is responsible for security oversight; their agency counterparts, Agency Information Security Officers (AISO), serve as an internal officer to each agency. The CDO has a similar structure with the ADO, where the ADO is responsible for their agency compliance and the CDO has oversight.

The policy makes clear that each agency is responsible for its own data and the information is reported to the parent organization, all the way up to OCTO. This puts a lot of responsibility on the AISOs and ADOs to keep their agencies in compliance as well as inventoried. The CISO and CDO then obtain all of their information from the agency counterparts.

### 3. Clear classifications

The data policy nicely defines five data classification levels from open to restricted confidential.

Having levels defined allows for quick access to defined datasets and allows defensible and reproducible queries into various levels of data. More importantly the definition of Level 0 explicitly says “all enterprise datasets that do not fall within the definitions of level 1, 2, 3,

or 4.” This means open access is default, and higher levels of restrictions require action, not the other way around. The policy also explicitly states that any data released as Level 0 is placed in the public domain through the usage of the Creative Commons CC0 Universal License. This means there are no restrictions of the usage of the data provided. The data being open and accessible is something that allows the citizens of the District to more fully understand their government that represents them and allows the citizens to use data to inform decisions as they pertain to the city.

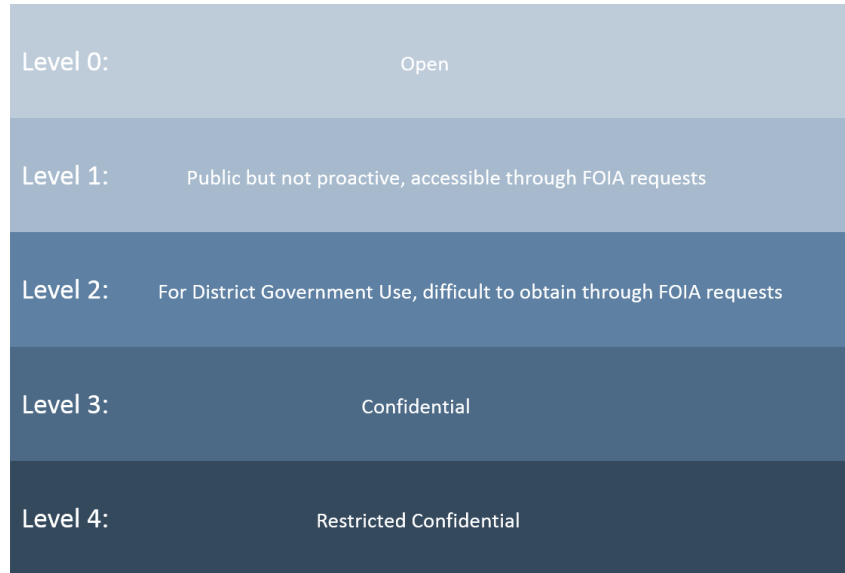


Figure 2. data classification levels

### 4. Data documentation

Another important aspect of the definition is the inclusion of “a dictionary to support the correct interpretation of data; and documentation of methodology or business rules,” in the metadata definition. Well documented datasets with useful metadata is absolutely required for anyone to correctly use the data itself. I need to trust where the data is coming from and what methodology was used.

The personnel and roles section contains no surprises. The Chief Information Security Officer (CISO) is responsible for security oversight and their agency counterparts, Agency Information Security Officers (AISO) is an internal officer to each agency. The CDO has a similar structure with the ADO, where the ADO is responsible for their agency compliance and the CDO has oversight. This roles section also plays out a little in the next, Enterprise Dataset Inventory, Classification, and Prioritization section. A major takeaway is that each agency is responsible for its own data and the information gets pushed up the chain. This puts a lot of responsibility on the AISO and ADO to keep their agency in compliance as well as inventoried. The CISO and CDO then obtain all of their information from the agency counterparts.

## 5. Timely releases and easy data discovery

Moving forward, the section on Data Catalogs assigns the responsibility of a searchable, timely, and useful catalog squarely on OCTO. This is very nice to see, this cannot be put together piece by piece, there needs to be top-down oversight. It also serves as a statement to the agencies and other public bodies that they are responsible for the publication of datasets even if OCTO is required to maintain the catalog.

Datasets that are accessed through the Freedom of Information Act (FOIA) requests and data sets designated Level 0 hold an interesting relationship. Level 0 and released data through FOIA requests are both public and have their limited restrictions but one is posted regularly and the other must be requested. To request a dataset, you need to first have at least a clue that the dataset exists. This is why it is nice to see there will be a dataset that catalogs FOIA requests and that data is designated to be Level 0. To be able to see what datasets have been requested, granted, denied, or denied and then granted allows the public to have greater insight into the daily business of their government. This section also calls for the development of a tool for easier FOIA requests, something that can be difficult to execute currently.

## Now to move forward

Overall, the District's new data policy is thorough and addresses most of the requirements laid out by open data organizations. It lays the groundwork for an open data ecosystem that is easy to use, timely, accessible, and well-maintained. It has received praise from organizations and individuals in the open data world, such as Sunlight foundation and Josh Tauberer.<sup>2122</sup> In general it seems like the feedback is positive with comments for where future policies can improve.

There are, however, a few points I would have liked to see in the policy and should be considered for whatever comes next:

### Data lifecycle management

To begin with, I believe that Level 0 and FOIA data is interconnected and the policy generally agrees. However, there was no discussion of frequently FOIA accessed data. There are many situations in which certain data could be released through a specific FOIA request, but still not released regularly at the Level 0 threshold—but there should be a stronger process in place for making OCTO review whether it should be. Constant evaluation would show a culture of thoughtfulness regarding the lifecycles of different datasets. This assessment cycle would provide assurances that the government is constantly thinking about their Level 1+ datasets, not just letting them sit.

### Enforcement rather than suggestions

Second, I also wanted to see stronger language that the ADO and ASIO are held responsible for their inventory, listing, and metadata as reported up to the CDO and CISO. If the CDO and CISO are making decisions through reported information, it is necessary to have language that more explicitly states that the agency counterparts are acting proactively for the system. In other words, there is no enforcement mechanism right now to ensure that agency counterparts carry through with their responsibilities; the

CDO can only trust that individual agencies will report their data through the proper channels. This may come in future, more far-reaching legislation, as called for in the policy, but for chains of information to work there needs to be the assurance that everyone is working together.

## Open legislative discussion

The data policy proactively calls for a proposed legislation and an accompanying report be sent to the Council by November 1, 2017. Something I suggest is that for the CDO, CIO, and OCTO to create the proposed legislation through a similarly open process to the data policy. Allow citizens and organizations to comment on and discuss the proposed legislation before it is sent to the Council. OCTO did a great job interacting with foundations and the public during the development of the data policy; why would the Council not want the same for proposed legislation?

As for the Council, I understand this policy cannot force the Council to act in the same way. We hope the Council would still adopt the policy itself. The current infrastructure that supports the legislative process (LIMS), makes research on bills easy, but it still relies on closed formats (PDFs), manual inputs, half completed systems, and some clearly just missing sources of data that aren't even collected. One of the most glaring missing sources of data is that there is no written record of Council meetings and testimony. In addition, the Council does not publish almost any data about the legislative process and notes on legislation in an open, standardized way. This includes different titles for introducers, file name formats, or even committee referral notes of legislation.

## Data usability

Another concerning exception to the openness demands of the data policy is the Office of the Chief Financial Officer (OCFO). Their data, in some cases, will naturally fall in the most restricted categories; however, the OCFO is also the source of much of the District's financial and budget data, as well as various economic indicators. While these data is extremely useful, they are often published solely in written narratives and other non-open formats, such as tables in PDFs. The OCFO also publishes budget data via the Cognos platform<sup>23</sup>, which can produce usable reports but is hard to navigate. The good news is that OCTO has been releasing data from the OCFO that was not previously available in a useful format (ITS extracts, use codes, etc.), which is extremely helpful. But sometimes there are arbitrary limits to how data is presented. For instance, [openbudget.dc.gov](http://openbudget.dc.gov) presents vital budget information as interactive graphics and web-based tables, but inexplicably fails to offer a way to download the underlying data in a machine-readable format.

## What's next?

Ultimately, the new data policy comes from the executive, and the executive does not need a law to implement its own policy. But some other agencies not under the control of the Mayor are not bound by the policy until it becomes law. This is why it is important for the Mayor's Office and the Council to work together, create actionable legislation which thinks out some of the loopholes, and implements a District wide law on how to handle open data.

As we move forward making D.C. more focused on transparency, sharing, and overall open data concepts, we need to think how the necessary policies and practices move from the executive branch to



the legislative branch to the community as a whole. Only when everyone is working together, providing inputs through comments or other forums, useful analysis about possible legislation, and even outcomes of useful products from open government data, will we know that this open data movement has succeeded.

\* This publication can be found at <https://www.dcpolicycenter.org/publications/diving-into-dcs-data-policy/>

<sup>1</sup> <https://octo.dc.gov/page/district-columbia-data-policy>

<sup>2</sup> <http://www.opendatapolicies.org/doc/washington-dc-2006-06-12/>

<sup>3</sup> <http://www.opendatapolicies.org/doc/washington-dc-2011-01-03/>

<sup>4</sup> <https://dc.gov/page/transparency-open-government-and-open-data-directive>

<sup>5</sup> <http://www.opendatapolicies.org/browse/by-date/>

<sup>6</sup> <http://datasmart.ash.harvard.edu/news/article/washington-dcs-first-ever-cdo-has-a-vision-to-equip-district-with-enhanced>

<sup>7</sup> Conversation with Barney Krucoff, April 21, 2017

<sup>8</sup> <https://mayor.dc.gov/release/mayor-bowser-announces-open-data-initiative-names-new-chief-technology-officer>

<sup>9</sup> <https://drafts.dc.gov/docs/draft-open-data-policy>

<sup>10</sup> <https://octo.dc.gov/blog/moving-dc%E2%80%99s-draft-open-data-policy-forward>

<sup>11</sup> There was an interim CDO, Dervel Reed, assigned in 2014. <https://octo.dc.gov/release/octo-launches-new-%E2%80%9Copen-data%E2%80%9D-technology-program>

<sup>12</sup> <https://octo.dc.gov/blog/meet-chief-data-officer-barney-krucoff>

<sup>13</sup> <http://statescoop.com/d-c-releases-draft-data-policy-with-new-security-safeguards-key-open-data-changes>

<sup>14</sup> <https://sunlightfoundation.com/2017/05/04/dc-data-policy-balances-privacy-security-and-openness/>

<sup>15</sup> <https://ogag.dc.gov/>

<sup>16</sup> <https://drafts.dc.gov/docs/district-of-columbia-data-policy-draft>

<sup>17</sup> <https://codefordc.org/>

<sup>18</sup> <https://opengovdata.org/>

<sup>19</sup> <https://sunlightfoundation.com/policy/documents/ten-open-data-principles/>

<sup>20</sup> <https://project-open-data.cio.gov/principles/>

<sup>21</sup> <https://sunlightfoundation.com/2017/05/04/dc-data-policy-balances-privacy-security-and-openness/>

<sup>22</sup> [https://twitter.com/JoshData/status/860290841444847620?ref\\_src=twsrc%5Etfw&ref\\_url=https%3A%2F%2Ftechnical.ly%2Fdc%2F2017%2F05%2F08%2Finnomaytion-announcements-open-data%2F](https://twitter.com/JoshData/status/860290841444847620?ref_src=twsrc%5Etfw&ref_url=https%3A%2F%2Ftechnical.ly%2Fdc%2F2017%2F05%2F08%2Finnomaytion-announcements-open-data%2F)

<sup>23</sup> <http://cfoinfo.dc.gov/cognos/finance.htm>