DEPARTMENT OF INFORMATION TECHNOLOGY

Maryland

Larry Hogan | Governor
Boyd K. Rutherford | Lt. Governor
Michael G. Leahy | Secretary
Lance Schine | Deputy Secretary

Data Policies and Data Assessment

The Impact on Maryland's Open and Internal Data Programs

State of Maryland – Department of Information Technology

Category: Data Management, Analytics & Visualization

Contact: Michael G. Leahy
Secretary / State Chief Information Officer
Michael.Leahy@maryland.gov
(410) 697-9406

Project Initiation/End: 2018

Maryland | **DEPARTMENT OF INFORMATION TECHNOLOGY**

Larry Hogan | Governor
Boyd K. Rutherford | Lt. Governor
Michael G. Leahy | Secretary
Lance Schine | Deputy Secretary

Executive Summary

Data Management is defined as "*the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles.*"[1] The State of Maryland, seen as a national leader in emerging technologies and innovation, has been leveraging data as a mechanism for governing and transparency, starting with the introduction of the State's Open Data Portal in 2014 and the adoption of the Maryland Open Data Act in 2015. Over that time large quantities of data have been developed, maintained, and made available to the public, however an overarching data strategy and associated data policies had not been defined.

To address this, the Department of Information Technology (DoIT), solidified the State's Data Management processes by developing policies and standards for data/metadata entry, resulting in consistent data quality and completeness. This also included the development of data inspection tools to verify that the data in the portals are, in fact, meeting these standards. Additionally, the State's Dataset Freshness Report and Dashboard, which provides a daily assessment of each dataset and if it is as up-to-date as required, was redeveloped and expanded to provide a complete catalog of all open data available from the State.

It is also important to emphasize how the customer accesses the data, that they understand the context of the data, and can see how the data is being used throughout the state. This was achieved through redesigns of the open data portals to provide a more organized flow of information. This also included new "area of interest" subpages that focus on the data's use in a particular industry, applications for interacting with the data or deriving analysis, and stories and videos that explain how the data was used. This also included the adoption of a new internal/protected data portal, built on the same platform as the Open Data Portal, but in a secure environment only available to State employees.

Concept

Data transparency and accessibility have been at the forefront of activity in the State. The State of Maryland has more than 45,000 government employees plus contracted staff, all potentially needing access to data in order to carry out their jobs. Additionally, citizens rely upon access to publically available data, with the State Open Data Portals supporting 60,000 unique users, 175,000 plus data downloads, and 5,000,000 API calls annually. In order to successfully meet the needs of both government employees and the public at large, a comprehensive approach to data management, data assessment, and data distribution had to be undertaken. As data is more broadly seen as government's most valuable asset, establishing management processes for data governance, data quality, and data centralization are imperative to the State's continued effort to ensure valuable data is available to support all government decisions.

---

[1] DAMA DMBOK, 2nd Edition (2017)

DEPARTMENT OF INFORMATION TECHNOLOGY

Maryland

Larry Hogan | Governor
Boyd K. Rutherford | Lt. Governor
Michael G. Leahy | Secretary
Lance Schine | Deputy Secretary

Recognizing this, the Department of Information Technology (DoIT) implemented and updated a number of components to their Data Management program. These included:

- Establishing Data Policies – Data Management and Governance
- Implementing Data Assessment and Validation Processes
- Streamlining Data Management and Distribution across Internal and Open Data Platforms

Establishing Data Policies – Data Management and Governance

In order for agencies across state government to adopt a consistent approach to data management, clear standards need to be defined. This enhances the value of the data in the expansion of evidence based and data drive decisions. DoIT released its Data Management Standardization Report on June 1, 2018, which identified both the data management strategy and the model for data governance as the foundation for all data management processes across the state. This initiated the process for reorganizing where data fall in the Data Flow Design, determining how and where data reside in business process development. The State has adopted the "data first" orientation where identifying that the "what" (data) should always come before the "how" (application development). By standardizing this approach to data management, the State had changed to approach to how data is seen in the workflow of government and provides incentive for ensuring data quality is maintained.

MARYLAND

**OPEN DATA
POLICY AND STANDARDS**
AUGUST 2018

Version 1.0

The release of the Data Management Standardization Report resulted in the development of child-policies specific to the way data is managed and served. As data transparency had been an important mission throughout the State, ensuring that the data that is being made available to all customers, through Open Data Portal, the Department of Information Technology developed Open Data Policy and Standards (https://opendata.maryland.gov/Administrative/Open-Data-Portal-Policy-and-Standards/axj6-iyhf) in August 2018. The specifications and guidelines in the Policy and Standards improve data consistency and availability of information. It ensures that all levels of government and the public have access to the most up-to-date information, reduce or eliminate overlapping data requests and redundant data maintenance, ensure metadata is consistently created, and ensure that data services can be displayed by the consumer with the software of its choice.
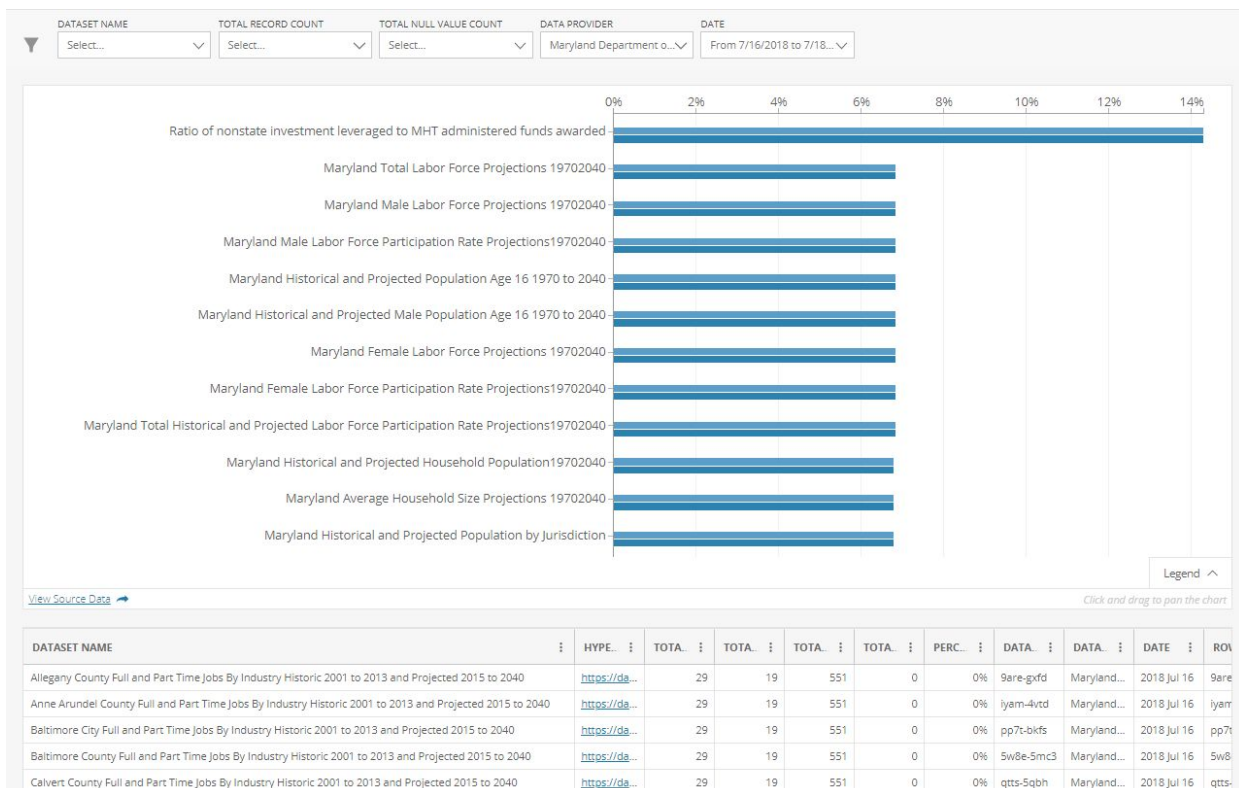
DEPARTMENT OF INFORMATION TECHNOLOGY

Maryland

Larry Hogan | Governor
Boyd K. Rutherford | Lt. Governor
Michael G. Leahy | Secretary
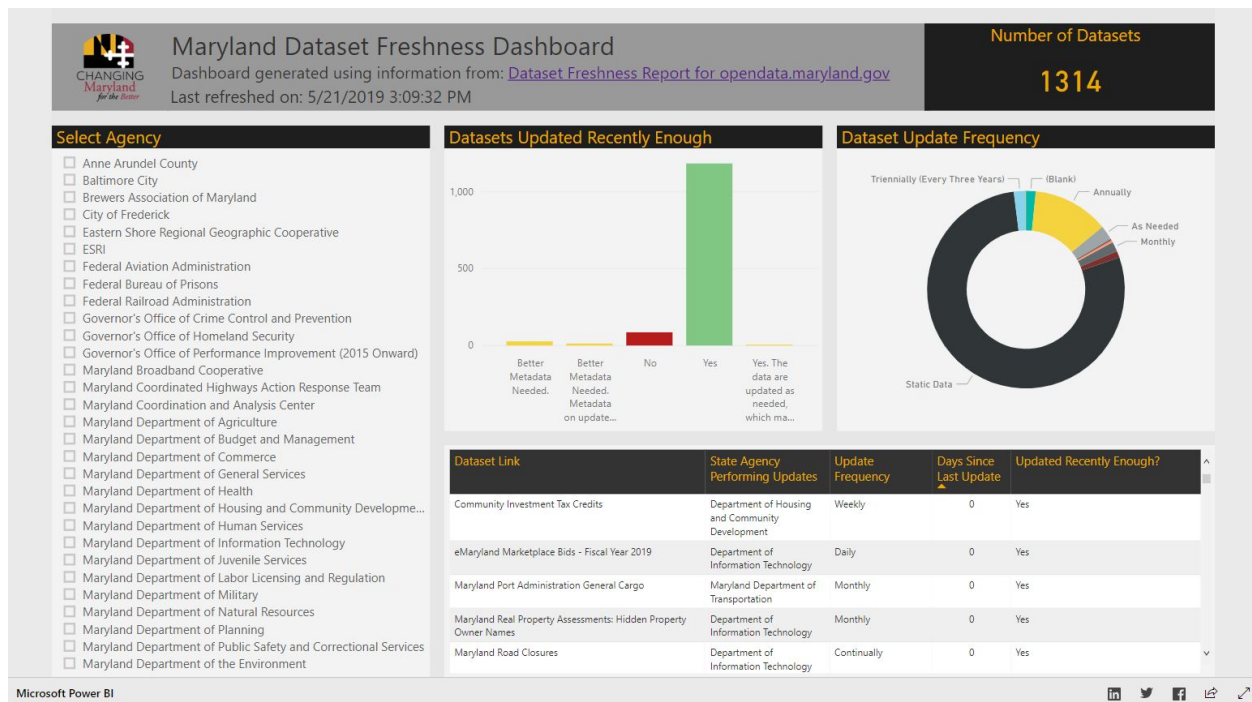**Lance Schine** | Deputy Secretary

Implementing Data Assessment and Validation Processes

In addition to establishing data policies, validating data accuracy, currentness, and completeness are important factors in data management. As part of the process for making data available on the Open Data Portal, each dataset is reviewed for completeness adhering to the Data Management Standards that have been developed. This process is done manually by the Open Data Portal administrator as part of the approval process prior to the data being made available to the public. To compliment this process, the Department of Information Technology developed data inspection analysis, which are performed using automated scripts. This is done to examine all data available on the open data portals, by collecting statistics where additional information can be derived. The analysis is performed on a dataset level, inspecting each record/value in the dataset, as well as an analysis on the specific fields to assess valid data schema. The analysis looks for data quality, data quantity, data duplication, as well as changes to the data. Information such as blank or null values, inconsistent cell characters, and count of changed records are made available to assess overall quality value of the data. This analysis occurs on a monthly basis and runs on all 1,300 plus datasets available on the Open Data Portals.

Beyond data validation, it is important to identify how current the data available is and if it is being updated frequently enough, based on the data type. When a State agency creates a dataset, there is maintenance timeframe that is selected. This may include daily, weekly, monthly, or annual data updates. A continuous goal has been to improve the overall timeliness of the data updates on the portal.

DEPARTMENT OF INFORMATION TECHNOLOGY
Maryland

Larry Hogan | Governor
Boyd K. Rutherford | Lt. Governor
Michael G. Leahy | Secretary
Lance Schine | Deputy Secretary

The Data Freshness Dashboard was launched in 2016 to allow data owners to identify which of their datasets were up-to-date and which were in need of data updates. The Dashboard displays data from a daily script that is run to inventory all open datasets. The Dashboard reflected the datasets found on the Open Data Portal, however did not include to comprehensive list of datasets that are available to the public, including all geospatial and non-geospatial data. In 2018, this process was updated to include all 1,300-plus datasets to provide accountability for dataset updating and maintenance. Additionally, a new Data Freshness Dashboard was developed, using commercial-off-the shelf software to replace the prior custom-developed dashboard. This will allow for easier maintenance and management of the dashboard moving forward, moving away from the reliance on custom development and programmers.
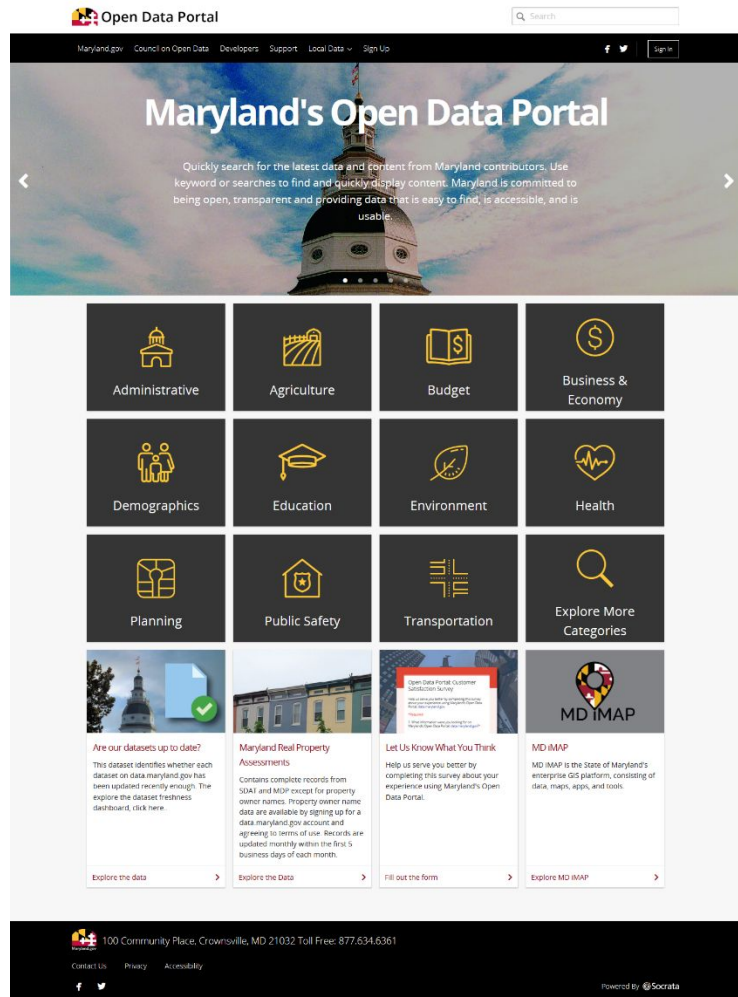

Streamlining Data Management and Distribution across Internal and Open Data Platforms

In 2018, improving the end user experience when exploring data, as well as streamlining the data management was a priority. The State's Open Data Portal (opendata.maryland.gov) received updated home pages to better streamline and organize access open data, as well as provide additional context to the information available to the public and the programs these datasets support.

The Portal has been updated with a focus on user experience, by streamlining and reorganizing data category tiles to drive the end user to the data they need. Additionally, the Portal now includes enhanced story features, which combines narrative content, data visualizations, and media inputs to tell a better story of how data driven and evidence-based information are supporting initiatives to better the lives are citizens in Maryland and beyond. Users can explore data, applications, and stories about

DEPARTMENT OF INFORMATION TECHNOLOGY

Maryland

Larry Hogan | Governor
Boyd K. Rutherford | Lt. Governor
Michael G. Leahy | Secretary
Lance Schine | Deputy Secretary

particular areas of interests, including Emergency Management, Environment, Community Investment and Planning, and Transportation which was done to better serve customers providing resources based on the data topic or category, rather than the agency that maintains the data.

More than 1,300 authoritative datasets are available to the public, providing a direct connection to data that is collected, maintained, and managed by the State of Maryland. The public can easily identify data through search, using dataset names, keywords, or categories. When available, related datasets are identified providing a more comprehensive source of information to help answer questions. The data can be downloaded in many file formats, including csv/spreadsheet, shapefile, json, geojson, xml, kml, and rss. Additionally, citizens can directly connect to the data through an application programming interface (API) reading the state data resources directly into their web and desktop applications. The Portal also offers built-in tools for visualizations of the data as tables, graphs, dashboards, and maps, as well as integration with other business intelligence and analytics platforms.

Additionally, the State implemented a new internal data portal (Socrata's Connected Government Cloud), using the same platform as the Open Data Portal that will serve as a self-service centralized data repository for state employees to access controlled or protected data, which they are given access to. The internal data portal provides a secure environment for interagency data collaboration, while streamlining data management processes by providing a single data repository where the data owner distribute data internally or to the public. The internal data portal also includes the identical built-in API and visualization tools as are available on the open data portal, further providing analytical capabilities to state employees.

DEPARTMENT OF INFORMATION TECHNOLOGY

Larry Hogan | Governor
Boyd K. Rutherford | Lt. Governor
Michael G. Leahy | Secretary
Lance Schine | Deputy Secretary

Significance

With the size, complexity, and purpose of programs and projects throughout state government varying significantly, adopting and implementing standardized data strategies and management processes can prove challenging. By establishing baseline policies and concepts, state agencies are able to, at the very least, start the implementation process using standard terms and identifying where their existing procedures may align with the state based policy. The implementation will always be ongoing as policy modifications are required and the validation procedures in place are iteratively improved. The adoption of the Data Standardization Policies, affirms the end users confidence in the data they are using authoritative data as part of their business process.

This can be seen in the 1,300 plus authoritative datasets are available to the public on the Open Data Portal. By establishing policies, and governing the data to affirm that the data available to the public had met the criteria associated with the policies, the data users can be confident that they are working with the best data available to them at that time. By making available dashboards and visualizations that allow the public to easily identify when data has been updated, by whom, and if it's on schedule, provides a level of transparency that is now expected of government. Additionally, by establishing routines, and making available what data had changed from update to update and where data may need to be improved, state agencies are provided an automated quality assessment quickly alerting them where further data investigation is required.

Impact

The enhancements to Maryland's data management processes will continue to evolve in the years to come. Establishing governance processes and demonstrating the value of data governance can be measured by the improvements in decision making with better information/analytics and high quality data, as well as gaining efficiencies and lowering costs with data centralization and process streamlining. This involves a significant change to the culture of government, but a necessary change to support a mature, successful data program.