PENNSYLVANIA OPEN DATA PROGRAM

Category: Cross-Boundary Collaboration and Partnerships

State: Pennsylvania

Contact: John MacMillan, Chief Information Officer

210 Finance Building

613 North Street

Harrisburg, PA 17111

cio@pa.gov

717.787.5440

Initiated: April 2016

Complete: Nov. 2017

Executive Summary

The Commonwealth of Pennsylvania's open data program is the collaborative result of partnerships with the civic hacking community, institutions of higher education, business partners, state agencies and the Governor's office. Working with these stakeholders, Pennsylvania was able to quickly establish a robust and successful statewide program where none had previously existed.

While commonwealth agencies provide a myriad of data on their public websites, it can often be difficult for citizens and other users to find the data they are seeking. The decentralized nature of the data also limits efforts to engage stakeholders across multiple disciplines who may also benefit from the availability of the data.

In April 2016, Governor Wolf signed an executive order to release agency data to the public in open, machine readable formats and tasked the recently created Office of Data and Digital Technology (ODDT) to establish the state's open data program.

The team collaborated with agency data owners, program areas and executives to publish nearly 100 datasets to a centralized open data portal while commanding quality and well-defined data. The open data portal provides quality, easily accessible, valuable data that can be used by citizens, businesses, academia, government, researchers, media, data evangelists, civic and social entrepreneurs and others to fuel business, create jobs and foster collaboration. Each dataset receives layers of data quality, policy and legal reviews, surrounded by a data governance process. The portal also provides a mechanism for users to recommend new data sets, further supporting the focus on providing high-value data and not just data for data's sake.

Building on the success of the new open data portal, Pennsylvania partnered with Harrisburg University of Science and Technology, Code for Philly, Technology Council of Central PA and industry partners to host a statewide hackathon event in the fall of 2017, known as CODE4PA, in Harrisburg and Philadelphia. Over 200 technology professionals, designers, entrepreneurs and students competed, using the latest technology to create new solutions based on the commonwealth's newly-published open datasets. The locations were connected live via a virtual classroom platform provided by one of the sponsors, which enabled participants from both locations to interact in real time.

The event culminated a month later with a "pitch day" during which 20 teams presented their projects – which included ideas, proofs of concepts, applications, dashboards and websites - to a panel of judges from the commonwealth, higher education and the private sector. Pennsylvania is currently in the process of planning for CODE4PA 2018, which will focus on using open data to understand and combat the opioid addiction crisis.

The open data portal and the hackathons provide a platform for state agencies to collaborate with each other and with stakeholders. They are a source of innovation to state agencies by using data in new and exciting ways for public transparency and engagement. The open data program has allowed Pennsylvania to build new relationships and strengthen existing ones with external stakeholders through collaboration on solutions and partnerships on initiatives.

Concept

Pennsylvania state agencies provide a wealth of data to the public via their websites, including reports, charts, maps and spreadsheets. However, hosting data on siloed websites makes it a challenge for constituents to find the data they want. Once the data is finally found, it is often presented in an aggregated report or PDF format. Not only is the aggregation potentially an issue, but the data is not broadly accessible or usable in analysis or other applications, creating usability barriers for constituents.

No centralized oversight or governance existed to monitor or promote the opening of state government data or prioritize the release of new data. State agencies were left to their own devices and initiatives regarding the release of open data to the public. This resulted in diverse methodologies for selecting datasets for publication, platforms and file formats, as well as a lack of standardization in the maintenance and updating of released data. Additionally, without centralized guidance or insight into the information released, there were no comprehensive methods to track the use of the data, measure the ROI or provide information to help guide customers to the data.

In April 2016, Governor Wolf signed an executive order to release agency data to the public in open, machine readable formats. Thus, Pennsylvania needed an enterprise solution that provides a central place for our citizens to look for data, with a dynamic enterprise governance and quality process, while providing transparency benefits. Additionally, Pennsylvania needed to foster collaboration around data through new efforts to engage the public. It also recognized that there were opportunities to partner with, and learn from, cities, universities and civic hacking organizations with established open data programs and initiatives.

Pennsylvania selected an open data platform with nominal costs, that is easily maintained and provides a user-friendly interface. This platform is used by most states and cities across the United States, which provides consistency and a familiar feel for researchers who review government public data. The portal provides analytics, such as top datasets, top referrers, top stories, search terms, applications, embeds and downloads. It also provides visibility into what is being searched on, giving insight into other data sets to potentially add. The entire platform was implemented by a team of two, one full-time and one part-time complement.

Because this open data platform used by most of the states, Pennsylvania was able to connect with teams of users to from their implementations. The ODDT staff found much value in these networking opportunities among colleagues in other states and cities. They also invested time to socialize open data within commonwealth agencies, as many of were not familiar about open data and the value it can add. The team answered questions, listed to concerns and provided examples of what the platform could do for their business and how it could help them to gain efficiencies. The portal launched with 11 datasets, enduring evaluations for quality, consistency, added detailed metadata and descriptions, following a data governance process, which includes business owners, legal counsel, security officers, policy and communications directors, information officers, Governor's office and legal advisory team as signatories approving the public release of the dataset.

With the open data portal as its foundation, Pennsylvania held its first hackathon, known as CODE4PA, in the fall of 2017 in partnership with Harrisburg University and community civic hacker groups. The event took place in both Harrisburg and Philadelphia, attracting over 200 individuals. The theme was open, with the only requirement being to use at least one dataset from the open data portal. Teams were also free to pursue projects based on use

cases provided by agencies or develop their own ideas and add in any other open dataset to build their concepts or prototypes. The teams were a mix of professionals and students who worked over a month on their own time. Live virtual classroom technology was used to connect the Harrisburg and Philadelphia host locations, paving the way for teams to pull chairs up to a screen and collaborate while in different physical locations. The teams returned the following month for fast-paced pitch presentations. The winning project used historical crash data from the PA Department of Transportation, along with weather data and other sources, to predict the most likely locations for future vehicle accidents over the next seven days. The project team suggested that state officials can use the data to address public safety and that motorists can use it to plan their driving routes. After action reviews of the event were positive and seasoned civic hacker participants remarked about how well the event had been run and that they planned to return for the next one.

During the CODE4PA event, networks and relationships were formed, providing a forum for government and industry leaders, technology experts and grass-root advocates to brain storm during a fiercely competitive programming/idea contest and share innovative ideas to strengthen our communities and government. These connections are strong and continue to thrive today.

Significance

The open data portal represents innovation and integration of data and provides the platform for feedback from community, media, academia, business, civic hackers, students and entrepreneurs. It is the beginning of community engagement, crowd-sourcing and collaboration between government, its partners and the public stemming from data as the raw source of information and the government's most valuable and strategic asset. At the same time, the open data effort emphasizes the critical importance of fully respecting privacy, protecting confidentiality and ensuring quality.

Government to citizen/community experience

The open data platform provides a customer-centric portal for citizens to interact with their state government – not only with downloadable datasets, but with an online form to suggest additional datasets. It is the foundation for agencies to make data available to the public and it provides stories and visuals that enhance their understanding of what government programs and services are accomplishing.

Improve business processes, programs, and efficiencies

The open data portal enables agencies to bring together different data islands and collaborate on the outcomes from viewing the integrated data points. Thus, agency experts spend their time discussing the results and outcomes from the analysis of the integrated data rather than spend it to bring the data together in a star schema to create reports.

Pennsylvania continues to focus on improving how agencies leverage existing data to facilitate their own programmatic work and to better serve our constituents. Sharing and analyzing datasets can lead to better information to evaluate programs and services. High-quality and reliable statistics provide the foundation for government and private research, evaluation and analysis to address societal issues and to understand how public needs are changing, how well state policy and programs are addressing those needs and where great progress can be achieved.

The Commonwealth of Pennsylvania collects and uses data every day to make decisions and provide services to our constituents. Most of this data are public records and subject to the federal Freedom of Information Act (FOIA) or state Right to Know Law (RTKL), but not everyone knows where to make requests or what data are available. Open data takes the approach of proactively publishing these records on the web for people to access and analyze.

The ODDT is very connected with communities, businesses and academia to listen to their needs and wants for data, gather feedback on the validity of the portal and to share future strategic plans.

Impact

Enterprise Memorandum of Understanding (MOU)
In the past, agencies' legal counsel spent many hours creating Memoranda of Understanding (MOU) documents for each data sharing initiative. This process could take anywhere from six to 12 months to finalize. The ODDT initiated the Enterprise Memorandum of Understanding (eMOU) to standardize the agreements and streamline their creation. This allows the project team to immediately start discussion on the data to be shared versus spending time questioning if data will be shared. It also creates consistency and acts as a catalyst to our data-driven culture by bringing agencies together to agree to share data.

Right to Know Law (RTKL) Request Improvements
Fulfilling individual requests for public records can represent a significant amount of time and resources for state agencies. The ODDT has reviewed the most frequently requested data for prioritization to be published on the open data portal, where possible. This provides one place for requestors to first look for the information they are seeking and then retrieve the data on their own, thus saving precious staff time. Datasets are listed with how often they are updated, so the requestor knows when the next update will be available.

Media Outlets
Representatives are turning to the Open Data Portal to look for data before submitting a RTKL request and to monitor what is new. This gives another place to request data of interest and to communicate with the state.

CODE4PA Hackathon
Civic hackers volunteered their time, skills and knowledge to enrich communities and the lives of citizens. Their reward is the satisfaction of giving back with their unique ideas while networking and learning with others that have the same ambitions and passions. The commonwealth's reward is hearing these innovative and new ideas, networking with the individuals and then exploring how these ideas/solutions can be used to improve the lives of citizens. Universities, business partners, students and IT professionals provided 20 conceptual ideas, prototypes and applications as ideas for the state. The grand prize was a public safety application with the potential to provide value to residents and travelers. Agencies gain quick wins on new ideas, applications, wireframes and conceptual models. This is a very creative and fast way to get feedback from innovators and early adopters. Teams explore data-driven concepts for decision-making and capitalize on the advantage of data sharing. Citizens and government officials working together side-by-side, making a personal connection, goes a long way in building understanding and trust in the people and process.

Open Data Portal

5

Data sharing across organizations and communities can enable more informed decision-making to improve coordination across government, provide transparency and enhance networking with our communities, academia and businesses. Open data is the groundwork for effective collaboration among all interested parties using governmental data to build applications, websites and other tools to help solve community problems, with our end goal of making life better and richer for residents. We enhance our communications with even more people by providing dynamic, rich visualizations and storytelling abilities within the open data portal directly tied to the supporting data. The portal also offers a suggestion form for anyone to submit their dataset requests. The ODDT reviews and considers every suggestion and engages with requestor for deeper conversation around the suggestion, as needed.

Making datasets available to the public also makes them available to other government departments that may have difficulty accessing them otherwise. This access can enable more informed decision making and improve coordination between departments. Within a year, PA's open data portal grew from 11 datasets to over 100 datasets and visuals. The ODDT continue to grow the open data portal with careful consideration of the value and quality of the platform rather than the quantity of datasets. In 2017, the site experienced over one-million browser page views and averaged 81,000 API calls per month. Agencies are using the portal's API through R and SAS to gain access to extremely large datasets from their own agencies because of its ability to organize multiple years of data into an easy to access format. The portal also gives agencies the option to discontinue their own open data websites and other public reporting tools, freeing up staff for other projects.

Open data meets the Governor's objective of increasing transparency, encouraging citizen engagement and using state data for social and economic good. Data is a valuable resource and asset to Pennsylvania, its partners and the public. Managing data as an asset and making it available, discoverable, and usable – in a word, open – not only strengthens our democracy and promotes efficiency and effectives through all agencies, but also has the potential to create economic opportunity and improve quality of life. For example, open data about vehicle crashes can be used for safe driving applications and analysis for proactive action to risky roads and intersections. Through the CODE4PA event, analyzing data, strategically integrating data and creatively thinking with our community leaders and citizens directly impacts citizen engagement and transparency.

The identification and prioritization of the publication of data and the establishment of expected outcomes from the release data are crucial. Among private sector groups, some of the common benefits of publishing open government data include:

Increase government transparency, accountability, and responsiveness: Publishing data about government budgets, expenditures and other operations provides visibility into how public resources are being used. Tools developed by the private sector (or provided directly by the public sector) can help individuals and businesses analyze government performance on issues of interest to them.

For example, a team of citizens that participated in CODE4PA used PA Department of State data on elections and contributions to show voters where political money comes from and how it is being spent.

- **Promoting citizen self-empowerment, social participation, and engagement:** Open data may be published to focus on specific issue-related problems and enable citizens to help propose solutions and make more informed individual decisions through analysis and crowdsourcing of information. It may also lead to greater collaboration and

data sharing between the private and public sectors.

In addition to benefits for the private sector, opening government data may result in benefits for the public sector as well. Some high-level benefits include:

- **Greater inter-agency and inter-governmental collaboration:** Making government data open allows for the breakdown of traditional institutional silos and may encourage government entities to work together, maximize resources, and reduce duplication of services.

  For example, in addition to using predictive analytics to forecast vehicle accidents, the winning CODE4PA team also integrated data from the Departments of Environmental Protections, Health, Education and Labor and Industry to create a public safety dashboard that demonstrates the value in data sharing and utilizing the data together.

- **Faster and more efficient service delivery to customers:** Proactively releasing frequently requested data allows customers to "self-serve" and reduce the time and resources needed to process individual requests.

  Example – Data is available on PA's uninsured population through the U.S. Census Bureau website. However, the data is very difficult to locate within the bureau's vast amounts of data. Several community partners thanked ODDT for providing the information on the open data portal

- **Data used for idea generation at hackathons CODE4PA:** Gains quick wins for the commonwealth on new ideas and shows create innovative solutions and creative ideas from our communities (aka crowd-sourcing). Increases exposure of the commonwealth's data and activities. CODE4PA produced 20 prototypes from data.

Employees, colleagues, citizens, and community partners are realizing the value in providing open data and organizing hackathons. We continue to deepen relations and connections with our constituents by listening to their ideas and sharing our work. PA's partners are a tremendous asset to our ability to continue to work within tight budgets, increase knowledge and improve offerings for the benefit of the people we serve.