

State of California California Open Data Platform

Category: Data Management, Analytics and Visualization

Scott Gregory

Deputy Director, Office of Enterprise Technology

916.431.5449

scott.gregory@state.ca.gov

Project Initiated: December, 2016

Project Completed: Multiple Production Iterations



California
DEPARTMENT OF TECHNOLOGY
STRATEGY INNOVATION DELIVERY

Executive Summary:

California is one of the most diverse regions in the world. It boasts a unique variety of people, cultures, business, education, health care and natural resources; not to mention it is the 5th largest economy in the world. With all these attributes and accolades, the data that this state develops in support of environment, health, public safety, natural resources and the general operations of its government, are the foundational building blocks that help this great state enable its citizens to thrive.

California's data resources are a key to its healthy economy and the California way of life that its citizens have become accustomed too. The state's collective data portfolio consist of water and wetlands, ocean and coastal ecosystems, oil and gas, fish and wildlife, cultural lands, forestry and parks, natural energy, and rivers and mountains. It also consists of detailed health data, public safety related information, and general data related to the operation of its government. It is the focus of the California Open Data program to working with our citizen and government stakeholders, to manage, protect, and curate the wealth of data the State of California produces for current and future generations.

The California Department of Technology (CDT) in partnership with the California Government Operations Agency (GovOps) have collaborated to develop a culture where information and data that informs it, is readily available for further interrogation. The State of California's workforce consists of approximately 235,000 employees, which represent many different and diverse organizations that support the complicated operations of governing a state of California's size and influence. In order to achieve this mission, data and the availability of it, is required to accelerate new, creative, approaches and innovative solutions. Our approach is based on technology, science, engineering, and collaboration, but the heart of it all is centered on the state's data assets. Our current and future ability to effectively manage and deliver a government that works for its citizens hinges on the availability of these resources.

In recognition of the efficacy of accurate and available data for decisions making, CDT and GovOps, developed a series of complimentary initiatives to support this new culture of data transparency. These two organizations developed an open data policy, which delineated the rules of engagement for organizations to participate in the state's open data program. Most importantly, the two organizations developed the California Open Data Portal; the state's flagship open data site for search discovery and use of one of California's most important resources – its data.

Data.ca.gov curates a large variety of datasets that represent some of the most important initiatives at work in California government. These initiatives range from education to health and human services, to natural resources and transportation. The platform allows citizens and government alike to ask questions, download state data and explore the richness of this vital resource made available for open use. California believes in the power of unlocking government data. We invite all to search and explore our open data portal and engage with our data to create innovative solutions. We believe the California open data portal will bring government closer to citizens and start a new-shared conversation for growth and progress in our great state.

Concept:

The ability to effectively manage and share California's vast data resources has become very challenging. The problems and issues that California faces are many. The rise of sea levels, the impact of climate change, increased population, droughts, wildfires and the social and demographic changes California is managing are incredibly complex. Not to mention the massive impact of California's economy on the global stage requires us as a government to use the resources we have to understand the challenges we face, and provide viable solutions to meet them head on.

Even though California has many resources at its disposal, its data assets, and the access and availability to them, was sorely lacking. This created gaps in data authority and providence. Without authoritative data and information, the State's confidence in meeting the challenges of the future would erode, and run risk of alienating the citizens it supports. The absence of key datasets for analysis and decisions make it difficult to articulate the problems being studied.

CDT and GovOps had a very clear definition of the problem – make data accessible and available for citizens and government to use. However, there was much more to the solution, and it had much to do with how we would develop the state's open data platform. The details resided in our ability to effectively collect, share, analyze, visualize, and publish large and disparate volumes of data in a useful and usable manner in collaboration with stakeholders, researchers, local/state/federal agencies, and other public and private entities. The goal was to create an open data platform that would encourage and enhance data sharing and data use across the various stakeholders related to the utilization of data and analytics in addressing California's challenges. To meet this goal, CDT and GovOps needed to implement and maintain an Open Data platform that would be utilized to enhance data collection, data transparency and access, which would provide high-value datasets to its constituents. The Open Data platform (data.ca.gov) would need to be utilized across a multitude of stakeholders that are committed to maximizing the use and value of the data/information to drive progress toward achieving the goals of California's initiatives, and strategically and collaboratively address the issues facing the State.

Significance and Impact

The California Open Data Platform has both Open Data and Analytics capabilities, capacity, and interoperability. CDT used best practices to design and implement the open data environment for the state. The California Open Data Platform consists of interoperable components that provide both an open source open data platform and data analytical platform capabilities and features.

The California Open Data Platform is a comprehensive solution that is flexible, scalable, and addresses complex datasets and business functionality required by state departments and their constituencies, stakeholders, partners, and the public. Data.ca.gov has over 17,000 data/information sets, 37 contributing data partners, 2,500 monthly resource downloads, 9000 monthly visitors, 25 different types of data available and 4 partners portals actively harvested currently. The California Open Data Platform is an interoperable environment that supports the entire data information value chain including collection; analytics; visualization; and publishing. Some key features include:

- Built on DKAN and MySQL open-source platforms
- Strong API capabilities and customization for special requirements
- Addresses large and small data volume capacity
- Data transformation and data extractor capabilities
- Flexible data capabilities: spatial, time-series, tabular, documents, maps, etc.
- Federation and data harvesting capabilities
- Powerful analytical and visualization capabilities
- Data import and export capabilities
- Rich search capabilities and rich set of metadata for each dataset and data dictionary
- Machine-readable format and role based security

California's constituency and defined collaboration partners can utilize this platform to perform complex analytical analysis and modeling of large and disparate datasets. The California Open Data Platform enhances data sharing across state program areas and changes the way they interface with and provide information to constituencies, stakeholders, partners, and the public. The California Open Data Platform features data, showcases, analysis, and information visualizations related to a variety of topical areas. Making data open and accessible has a significant value to the management and decision-making processes of California's government. The following benefits were identified as valuable for the implementation of the California Open Data Platform:

- Improve Information Transparency
- Improve Data Accessibility
- Make Data Useful to Everyone (all interested parties)
- Improve External and Internal Data Contributions
- Improve Data and Information Collaboration
- Provide Optimization of Open and Analytical Data features and technical capabilities

The California Open Data Platform is significant because it is an invaluable resource for all parties and stakeholders interested in California's data resources. It provides organizations, citizens, stakeholders, and interested parties better access to holistic data and allows them to use it in a manner that fits their needs. It helps local, federal, and state agencies make data driven decisions by giving them access to data that was previously locked in silos.

The California Open Data Platform was developed to provide information to California citizens, agencies, and interested stakeholders and parties in a transparent and useful manner.

It supports state program missions by providing an environment to collect, share, analyze, visualize, and publish data that can be utilized effectively. Recognizing the significant benefits and value from an active open data program, California launched into a new era of utilizing a multitude of datasets to set direction, analyze impacts, make adjustments, and communicate with affected and interested parties related to California's future. By utilizing data on data.ca.gov and engaging stakeholder participation and contributions, all parties are able to assist in addressing California's challenges.

The success of the California Open Data Platform was brought about as a result of business and information technology alignment that created enabling capabilities and collaboration for the

state's data and information stakeholders. In a broader perspective, and in alignment to the California State Information Technology Strategic Plan, the California Open Data Platform functions as a statewide resource to do the following:

- **Improve Government:** Open data and analytics improves government, by increasing transparency, and enhancing services and direction.
- **Empower Citizens:** Opening access to government data gives citizens and stakeholders the ability to make informed decisions and impact change.
- **Create Opportunity:** Open data creates new opportunities for interested parties and organizations by fostering innovation.
- **Solve Public Problems:** Open data plays an increasingly important role in solving large problems, primarily by allowing citizens and policymakers access to new forms of data-driven assessment of the problems at hand.

It also enables data-driven engagement, producing more targeted interventions and enhanced collaboration. The California Open Data Platform is an invaluable resource for all parties and stakeholders interested in California related data and information. It gives program areas, agencies, citizens, stakeholders, and interested parties better access to holistic data and allows them to view and use the information in a manner that fit their needs. The California Open Data Platform is a collaborative data environment that will continue to grow and progress over time. The more data that is collected, analyzed, and published the more valuable the California Open Data Platform, as a resource, will be. The use of the new California Open Data Platform is open to all state organizational programs and all California's stakeholders or interested parties. CDT and GovOps plan on continuing to work with departments, commissions, boards, conservancies, local agencies, researchers, and others to add much more data, visualizations, and features to it. The Open Data Ecosystem facilitates government improvement by assisting officials in critical decision-making based on data, transparency, innovation and stakeholders' engagement. With the launch of the California Open Data Platform, the CDT and the GovOps have already seen improvements in data collection, data analysis, data access, data sharing and collaboration, and data driven decision-making processes.

Links:

California Open Data Platform - <https://data.ca.gov>