



# Open Data for Big Data

Washoe County  
Technology Services

**Washoe County  
Technology Services**

# **Open Data for Big Data**

## **Executive Summary**

The concept of open government is now becoming more of a reality in a lot of public organizations. In this era of information, governments nowadays strive to connect with citizens using the value of data. One segment of open government is about releasing data to our constituents through open data.

Open data is about releasing non-sensitive data so that it can be easily shared and is accessible to the public. This is free data that is provided to the public in a format that the citizens can use and reuse it. The aim of this is to foster collaboration not only with citizens but also with other government entities.

Generally, government websites contain information for the public but are spread out in different websites that is either difficult for the public to locate or even to peruse the data. In open data, a portal is created that is dedicated to displaying government information for various sets of data. These data are formatted in a way that it can be downloaded for analysis. Technical developers can also create applications either web-based or mobile that the data can be fed into. Users of these applications may be able to access real-time information if the data provided is current and timely.

With the open data initiative, it allows the government to work with citizens as more of a team effort. Though there will be challenges as we move through the process of creating open data, there are positive advantages such as creating transparency and accountability which are essential in governance. The focus on transparency and accountability has been magnified as a main goal in the realm of government management. This creates an atmosphere of trust as citizens will be more aware of the county's business operations. We may be able to promote innovative ideas from the citizens that will enhance the way government operates. This may allow us to tap into the expertise that our citizens can provide which may lead to economic gain within our community.

On the data.gov<sup>1</sup> website, it currently shows that 39 states and 40 U.S. cities and counties (see Figure 1) in United States have joined the effort in the open data initiative. In California, Alameda County has a data sharing site that provides information regarding its operations. Along with this, it also held special events such as the Alameda County Apps Challenge in 2012 and 2013 to get citizens involved in creating exciting applications that will benefit the county.

Figure 1 – Open Data Sites in United States



Source: Data.gov

The data collection performed by government agencies through numerous years has created a huge potential for analysis. The plethora of data harvested and evaluated can be core factors in important decision making which can eventually lead to major business values and cut inefficiencies in the government arena.

A lot of government entities are embarking on data releasing initiatives that allow citizens to extract and utilize public data. This open data effort is becoming another cornerstone of open government. As more open data is implemented within various government entities, public organizations and entrepreneurs have learned to make use of these available data in a more productive manner. The availability of these open data allowed the public to intermix data from various systems for analytical purposes which assisted in boosting business for entrepreneurs and helped cut down cost for government agencies. This collaboration of efforts between government and citizens

---

<sup>1</sup> Data.gov was launched on May 22, 2009 by the federal government to allow the public to gain access to non-sensitive government data.

can eventually lead to what is popularly known as Big Data<sup>2</sup>. Big Data projects are evolving in the entrepreneurial market and also at the federal level but local governments can also implement the same idea on a smaller scale version starting with open data.

For decades, Washoe County has collected and stored information through a variety of applications and databases. These are structured data that can be used for ad hoc query and analysis. The datasets gathered throughout the years in these databases can be extrapolated with the possibility of merging it with datasets from other entities to create more values for decision making. The integration of Washoe County's dataset with other government agencies may assist into uncovering insights that could result to improving services. Washoe County can do its share of open data to allow citizens to harness these data in ways that can promote the community. This sharing of data with our constituents can also stimulate the need to progress to a collaboration effort and analysis on a much larger scale called Big Data.

---

<sup>2</sup> *Big Data is defined by Gartner as "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making".*

## **Challenges and Opportunities**

In order to be successful in Washoe County's open data initiative, it needs the full support and sponsorship of senior management. This effort needs the departments to assist in carefully assessing and prioritizing data categories that can be released to the public. Some counties like Cook County, Illinois went to the extent of passing an ordinance mandating the county to officially embark on the open data initiative. There may be situations wherein a department may need to dedicate personnel to the open data working group effort which can pose an issue considering budget cuts and the lean number of personnel in each department.

It's not only the mere fact of trying to organize and decide which data to present to the public but the reality of what the public reactions will be to the data when it is released for use. There is a good possibility that the public will audit the data and raise questions about expenditures and business processes. This can be viewed negatively but on a positive light it can potentially result into a major improvement in the way government functions. Ideas from the public may be solutions in cutting cost and in improving government efficiencies. A proactive communication plan is highly recommended as an integral component of the open data initiative. It should be developed for organizational preparedness for issues that may arise.

One of the great opportunities in releasing data is that it can bring about trust and improve the relationship between citizens and government. It can help identify accomplishments in performance and measure programs to make sure that the goals of a project are being met.

One of the main goals in implementing open data is to achieve transparency. Along with transparency is accountability. Both are key foundations of good governance. Simply releasing information to the public does not fully achieve transparency. Information that will be presented to the public should have relevant information. It should also be accessible and available in various formats so that different audiences of the public can download and reuse it. There should be enough detail but at the same time ensure that sensitive information are not compromised. As a government entity, we have an obligation to protect our citizen's sensitive data. A clear definition of the kinds of data to be published will need to be carefully determined. Data should also retain a level of disaggregation that can be useful for analysis and evaluation. Information should be made public in a timely and accurate manner for integrity purposes.

## **Data Sharing**

There are some basic but important items that need to be lined out in the deployment of open data. It is important to consider about licensing because this will describe how the released data will be accessed and used. The licenses listed below will need to be reviewed to determine which will be the viable options for Washoe County. Here is a list of the licenses:

- Open Data Commons Public Domain Dedication and License (PDDL)
- Open Data Commons Attribution License
- Open Data Commons Open Database License
- Creative Commons

There is no exact formula on which data should be released first. The county released schedule should consider its strategic plan, mission, goals and data that is currently available on our website or consider information that are requested by the public. Below is a list of datasets that have been published by various government organizations:

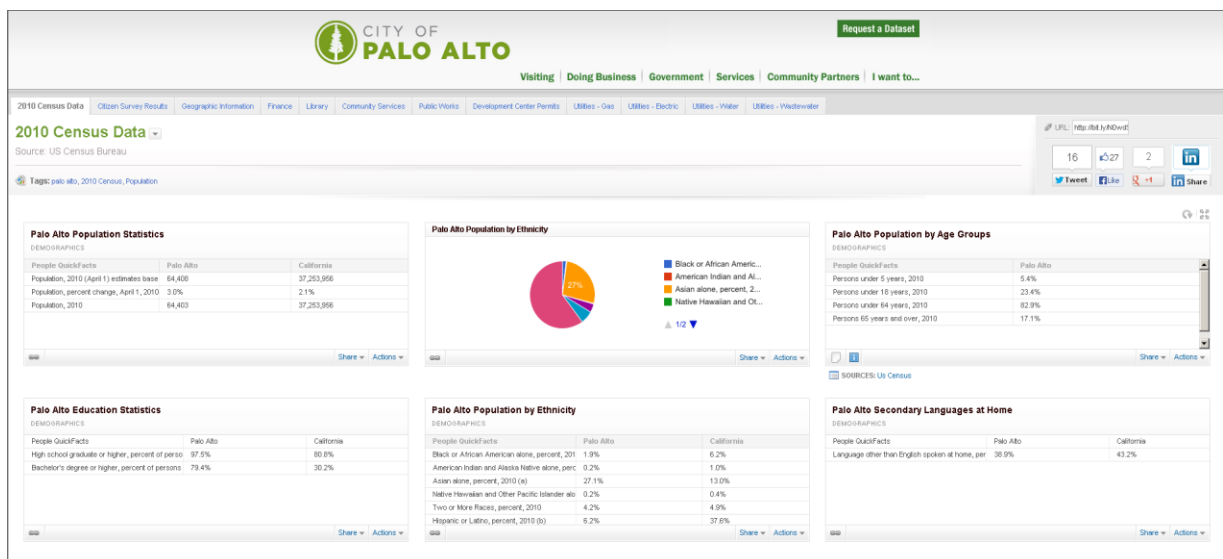
- Sheriff's crime reports
- Restaurant inspection health reports
- Vendor awarded contracts
- Foreclosure information and map
- Property tax
- Map of elected officials
- Administrative hearing on top 5 Sheriff violations
- Building and zoning permits
- General fund expenditures
- Payments to health care providers information for citizens, inmate care
- Tax collection trends
- Median household income (with US Census)
- Database of Wi-Fi hotspots
- Service requests
- Geographic data on public parks
- Social Services information
- Healthcare data

## Data Portal

Instead of having datasets presented on various web addresses, governments are using one portal web address to release open data. The site provides users the option to select a dataset through a variety of format as in geospatial, spreadsheet, chart, calendar and others. It also has interactive functionalities to query the database, filter, sort, compare data and others. Data can be exported into a variety of format like CSV, JSON, PDF, RDF, RSS, XLS, XLSX, XML and others. For developers, there are also Application Programming Interface (API) that can be provided to access open data through custom applications. There are sites that will request a user to log in prior to extracting data. Information regarding frequency of the data updates is also provided.

Below are some of the platforms used in creating the open data portals:

- Junar is an open source platform used by City of Palo Alto.




## Use the City of Palo Alto API.

The API allows direct access to the data published in the City of Palo Alto Open Data site from your applications. It uses a RESTful interface and returns the data in JSON format.

Data streams invoked through the API provide online and standardized access to data contained in HTML pages, XLS, CSV and other similar file types available on Internet.

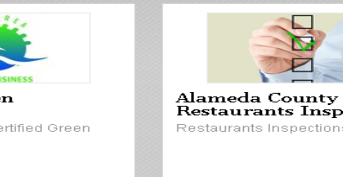


Get an Auth Key! 1	Create a Request 2	Process your response 3
<p>The Junar API requires a key to operate.</p> <p>Get one and access to any resource through it.</p> <p>Jump right into the <a href="#">API resource documentation</a> to get in touch with the requests and methods you can use.</p>	<p>A typical request will look like this:</p> <pre>http://paloalto.cloudappx.junar.com/datastreams/invoke/GENER-RESID-WATER-SERVI-UTILI?auth_key=YOUR_API_KEY</pre> <p>A few things to note in the request above:</p> <p><b>/datastreams/invoke/</b> This is an invoke request, that will return a json response. The entire resources descriptions are available in the <a href="#">API resource documentation</a>.</p> <p><b>GENER-RESID-WATER-SERVI-UTILI</b> This is the GUID of the data stream you are accessing. You will always find this GUID in the Data Stream Details page, making a data stream search through the API, or in many other places.</p> <p><b>auth_key=YOUR_API_KEY</b> This is the key from step 1. You'll need it for every request.</p>	<p>Your response will be returned as JSON. You can also obtain the response in XML, CSV and HTML...</p> <p>To understand the format of the result field, read the <a href="#">API resource documentation</a>.</p>

- 

Alameda County Data Sharing Initiative


[Home](#)
[Suggest a Dataset](#)
[Community Apps](#)
[Terms of Use](#)




**Got code?**

**Apps Challenge 2013.1**


On April 27, 2013, Alameda County hosted its second hackathon. The Alameda County Apps Challenge 2013.1 took place at Berkeley High School and was a huge success. Visit our website for more details.




**Certified Green Businesses**  
Alameda County Certified Green Businesses



**Alameda County Restaurants Inspections**  
Restaurants Inspections



**Alameda County Renewable Generation July 2011 To June 2012**  
Alameda County Renewable Generation July 2011 To June 2012 in 15 minutes intervals



**Alameda County Sheriff Crime Reports - Map**  
Alameda County Sheriff Crime Report - Map










**View Types**

  - Datasets
  - Charts
  - Maps
  - Calendars
  - Filtered Views
  - External Datasets
  - Files and Documents
  - Forms

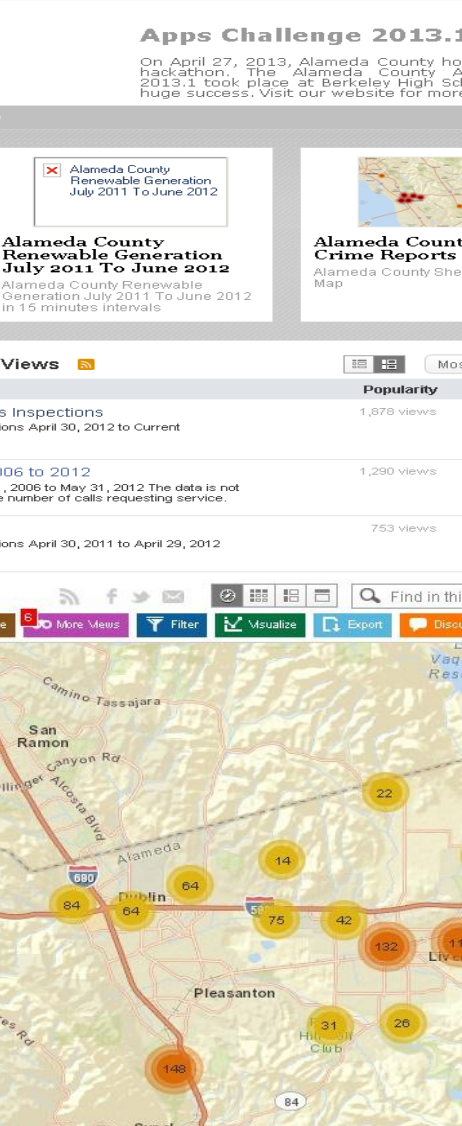
**Categories**

Business








**Search & Browse Datasets and Views**

Name	Popularity	Type	RSS
 <b>Alameda County Restaurants Inspections</b> Alameda County Restaurants Inspections April 30, 2012 to Current	1,878 views		
 <b>Alameda County Bed Bug 2006 to 2012</b> Bed Bug Requests for Service June 1, 2006 to May 31, 2012 The data is not confirmed cases of bed bugs, just the number of calls requesting service.	1,290 views		
 <b>Taco Bell</b> Alameda County Restaurants Inspections April 30, 2011 to April 29, 2012	753 views		

**Alameda County Sheriff Crime Reports...**  
Based on Alameda County Sheriff Crime Reports  
Alameda County Sheriff Crime Reports from January 1, ▶



Find in this Dataset



## **Next Steps – Realization of Open Data Initiative**

As open data is evolving more in the government arena, Washoe County needs to start laying out the foundation for an open data field. The path to gaining a successful open data architecture involves a lot of departmental collaboration and coordination.

The initial launch of open data initiative in Washoe County will need the sponsorship of Information Technology Advisory Committee (ITAC). An approval from the Board of County Commissioners to designate ITAC as the main sponsor of this project will be the primary onset to Washoe County's open data initiative.

In a cooperative effort, a team consisting of Department of Technology Services, Community Relations Office and Organizational Development Manager will work closely to push the open data project implementation through. As an enterprise project, not only is the technical aspect essential but the internal and external communications are keenly needed to the success of the project.

The Department of Technology Services will lead the open data implementation with the core responsibility of project planning. It will also provide updates and recommendations to ITAC as the project progresses. The Organizational Development Manager will create a communication plan and change management plan to assist the departments in understanding open data, its value to the county and provide status updates on project phases. Through this process, the departments are more likely to engage and commit to the open data initiative. With the current role of the Community Relations Office as a liaison between Washoe County and our constituents, the expertise of this office is also essential to ensure that the community outreach functions are performed effectively. A proactive communication plan will need to be developed by this agency. The Community Relations Office has the knowledge and skill set to promote open data to our citizens as well as communicate back to the project team regarding feedbacks or data requests.

Defining high-value data and making it available to our constituents is a complex process. The protection of sensitive data and confidential information of our constituents remains a high priority and will continue to be protected. As open data is still a new concept in government, Washoe County will need to consult with professionals on this subject matter for assistance and guidance in forming a framework and an implementation plan.

The need to embark on an open data initiative can no longer be delayed. The untapped potential of data can assist further in government strategic decisions. As federal governments work on Big Data projects to achieve transparency and accountability, local governments like Washoe County can do its share of Big Data project through open data.

## **Websites**

Here is a list of some government open data portals:

- Alameda County, CA  
<https://data.acgov.org/>
- City of Palo Alto, CA  
[http://www.cityofpaloalto.org/gov/depts/it/open\\_data/default.asp](http://www.cityofpaloalto.org/gov/depts/it/open_data/default.asp)
- King County in Washington  
<https://data.kingcounty.gov/>
- Cook County, Illinois  
<https://datacatalog.cookcountyil.gov/>
- San Francisco, California  
<https://data.sfgov.org/>
- Austin, Texas  
<https://data.austintexas.gov/>