



# The Computerworld Honors Program

Honoring those who use Information Technology to benefit society

## Final Copy of Case Study

**Status:**

Laureate

**Year:**

2013

**Organization Name:**

U.S. Census Bureau

**Organization URL:**

<http://www.census.gov/>

**Project Name:**

Census Bureau Open Data API

**Please select the category in which you are submitting your entry:**

Human Services

**Please provide an overview of the nominated project. Describe the problem it was intended to solve, the technology or approach used, how it was innovative and any technical or other challenges that had to be overcome for successful implementation and adoption. (In 300 words or less.)**

Like many federal agencies and bureaus facing continued budget reductions and limited resources, the Census Bureau is adapting its operating model to be a smarter, faster, and leaner 21st century organization. As the authoritative source for statistical data about the U.S. and its citizens, the Census collects, analyzes, and disseminates massive amounts of information to a widely diverse audience spanning localities, companies, even countries. These large volumes of data, coupled with the rapid pace and democratization of technology presented the Census with a challenge in how it should expand access to its data while ensuring quality and security and minimizing costs. To address this complex challenge the Census Bureau launched the Open Data Strategy project. The

project aimed to deliver an application programming interface (API) to the public that can provide building blocks for third-party developers to design online and mobile applications using Census Bureau statistics. The API empowers developers with easy web-access to a single pipeline of data with full flexibility and control over which data variables to request and receive, without the need to create copies. To achieve this, Census had to develop innovative approaches to harnessing data from two of the largest federal statistical datasets available for public consumption, comprising nearly 400GB of storage. A key challenge was the ability to improve performance given the API needed to handle up to a thousand concurrent users and return data requested with an acceptable response time (sub-second more than ninety-nine percent of the time). While APIs help expand computing potential at Census, their real value comes through enabling innovation and fostering a 21st century economic model. The Census Open Data API allows developers both inside and outside the agency to create new, and perhaps unexpected, applications based on Census Bureau statistics, thereby extending their reach.

**When was this project implemented or last updated? (Please specify month and year.) Has it incorporated new technologies and/or other innovations since its initial deployment? (In 300 words or less.)**

The Census Bureau open data API was implemented in July 2012 with the release of the 2010 Decennial Census data and the 2006-2010 5-Year American Community Survey Summary (ACS) File. It was last updated in December 2012 with the release of the 2007-2011 5-Year ACS Summary File. The ACS is an ongoing statistical survey that samples the U.S. population every year -- giving communities the information they need to plan investments and services. The Census Bureau Open Data API has incorporated continuous updates and enhancements as suggested by our user community of developers. For example, when it was originally deployed, there was a limitation of five variables which could be obtained per request. Several users asked if it could be increased, despite the fact that it could severely impact the performance of the API. The development team modified the API to handle fifty variables in a request, and avoided the performance issue by mimicking the action of making multiple requests. This modification improved the overall performance to the user community by eliminating the network lag from making multiple requests from the client.

**Is implementation of the project complete? If no, please describe the project's phases and which phase the project is now in. (In 300 words or less.)**

No, the project is not complete. There are four phases including: remainder of Phase I: Incorporating additional aggregate datasets; Phase II: Normalizing time



series datasets; Phase III: Normalizing public use microdata datasets; and Phase IV: Normalizing longitudinal datasets so that the complexity of their use is transparent to the developer/user community. The project is currently in maintenance of Phase I, where additional aggregate datasets are being included, and in the design stage of Phase II.

**Please provide at least one example of how the technology project has benefited a specific individual or organization. Feel free to include personal quotes from individuals who have directly benefited from the work. (In 300 words or less.)**

One example of how the open data API has benefited an organization is the Sunlight Labs' Sitegeist application, the goal of which is to show how government data can be useful to the average person. "[The] majority of the data used in Sitegeist comes from the US Census Bureau and their wonderful API...I've never worked with the Census bulk downloads, but I hear from colleagues that it can be a daunting task. Their API makes it incredibly easy to slice and dice the numbers as needed, combining data across 'tables' for any geography you are working with." Other examples include platitudes from general public developers: "Having used these APIs for a while, I find them very useful and working well. Thank you." And: "First, let me applaud the Census Bureau for doing this; it's an awesome project. I'm glad to see JSON as the response format for the API."

**Would this project be considered an innovation, a best practice or other notable advancement that could be adopted by or tailored for other organizations and uses? If yes, please describe that here. (In 300 words or less.)**

The Census Bureau open data API is considered a notable advancement that could be adopted by or tailored for other organizations, especially smaller federal agencies that may not have the resources or funding to establish the necessary infrastructure to fully support the potential users of their data. The open data API is an integral part of the Census Bureau's response to the comprehensive federal Digital Government Strategy, of providing data and services that are available "anywhere, anytime, on any device."

**If there are any other details that the judges should know about this project, please note them here. (In 300 words or less.)**

In addition to the ability to allow external developers to create new applications, internal developers are leveraging the same technology, and analysts are using the open data API to disseminate complex information to the public in an easily digestible format through web and mobile applications and newly developed data visualizations.