

**UNITED NATIONS ECONOMIC COMMISSION
FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS**

Work Session on the Communication of Statistics
(Geneva, Switzerland, 27-29 June 2012)

Session 2: Building Credibility

**Improving Data Transparency and Accessibility by Enabling Third-Party Innovation.
Leveraging Application Programming Interfaces to Disseminate Census Data**

Prepared by U.S. Census Bureau, United States – Stephen Buckner, Lisa Wolfisch and Logan Powell

Contents

Contents	1
I. Introduction	1
II. Context	3
III. An “Information-Centric” Approach	4
V. Decoupling Data and Presentation	Error! Bookmark not defined.
IV. Conceptual Model.....	Error! Bookmark not defined.
VI. The Data Web API and Launch Process.....	5
VII. Future Plans	8

I. Introduction

1. On his first full day in office, President Obama signed a Memorandum on Transparency and Open Government:



Transparency promotes accountability and provides information for citizens about what their Government is doing. Information maintained by the Federal Government is a national asset. My Administration will take appropriate action, consistent with law and policy, to disclose information rapidly in forms that the public can readily find and use. Executive departments and agencies should harness new technologies to put information about their operations and decisions online and readily available to the public. Executive departments and agencies should also solicit public feedback to identify information of greatest use to the public.

2. Improved data access, a critical part of this Open Government Initiative, fosters transparency, participation and collaboration from U.S. Federal agencies. This is also in accord with the recently published U.S. Digital Government Strategy: Building a 21st Century Platform to Better Serve the American People¹. In this spirit, the Census Bureau has created application programming interfaces (APIs) to enable open, stable and secure access to Census Bureau statistics for use by applications, including mobile applications, user mash-ups², data visualizations and web applications. This new strategy will liberate our data and allow external collaborators to harness them and thereby assist in their dissemination.

3. An application programming interface (API) allows direct access to an organization's data resources without the need to go through front-end channels (e.g., interfaces designed for user access like websites, web applications, download centers and mobile applications) provided by the organization. This direct-link to resources enables innovation and lowers barriers to development by opening access to third parties to create their own applications built 'on top' of agency resources. This is especially important for those third parties with niche or highly segmented markets that demand specific applications that fall outside the scope of an agency's budget and services. Those developers can directly access the specific data they need to build applications for that niche purpose.

4. Third-party application developers can now create their own applications with the Census Bureau's API, launched as a beta release in April, 2012. This kind of functionality empowers citizens to use Census Bureau statistics in new ways and thereby maximizes their value and raises awareness of agency programs. At present, the API provides access to the 2010 Census Summary File 1 and the 2006-2010 American Community Survey. This is the first step towards providing broader access to agency data through a generalized API mechanism that will accommodate different kinds of data, for example geospatial data or documents. This paper discusses the goals and results of this open data initiative including the API experience to date.

5. In his seminal "Government as a Platform"³ white paper (2009), Tim O'Reilly, technology forecaster, publisher and open-source advocate, remarked on the democratic power of 'crowd-sourcing' innovation by leveraging the passions of the American citizenry or "coordinating the collective action of citizens" which is enabled by the use of APIs. He wrote:

“ Government maintains information on a variety of issues, and that information should rightly be considered a national asset. Citizens are connected like never before and have the skill sets and passion to solve problems affecting them locally as well as nationally... Citizens are empowered to spark the innovation that will result in an improved approach to governance.

This is a radical departure from the old model of government, which Donald Kettl so aptly named 'vending machine government.' We pay our taxes; we get back services. And when we don't get what we expect, our 'participation' is limited to protest – essentially, shaking the vending machine.

¹ <http://www.whitehouse.gov/blog/2012/05/23/roadmap-digital-government>

² Mashup (web application hybrid) - a web application that combines data and/or functionality from more than one source

³ <http://www.forbes.com/2009/08/10/government-internet-software-technology-breakthroughs-oreilly.html>

In the vending-machine model, the full menu of available services is determined beforehand. A small number of vendors have the ability to get their products into the machine, and as a result, the choices are limited, and the prices are high.

Yet there is an alternate model, which is much closer to the kind of government envisioned by our nation's founders... In this model, government is a convener and an enabler – ultimately, it is a vehicle for coordinating the collective action of citizens.”

II. Context

6. The work being done at the Census Bureau surrounding APIs is aligned with the recently published Digital Government Strategy and complements several initiatives aimed at building a 21st century government that works better for the American people. The Digital Government Strategy outlines a path for all U.S. Government agencies to reduce redundancies among and increase the flexibility of IT systems (both inter and intra-agency wide) by shifting from a document-centric to an “information-centric” paradigm.

7. The strategy notes that in order to accommodate such a shift to a more open and participatory model – to serve as a platform for innovation –, governments must shift how they think about digital information. Instead of focusing the majority of their effort in formatting the presentation of web pages, applications or other publications of content⁴, agencies can better spend their resources ensuring that their data are accurate and secure. By reframing content as structured data and verifying that it is associated with valid metadata⁵, agencies move closer to “Gov 2.0”⁶.

8. In concert with this strategy, by providing information through web APIs, agencies can build architectures that foster interoperability and openness, making data assets freely available for use within and between agencies, by entrepreneurs, academics and others. This approach also supports device-independent security and privacy controls, since such features are applied directly to the data (via metadata). This structural change enables agencies to focus on securing the data so that it is usable on a variety of platforms rather than specifically securing any individual device or application.

9. With foresight in predicting the inevitability of a transition such as that outlined in the Digital Government Strategy, the Census Bureau has made progress towards the ends of “increasing returns on IT investments, reducing waste and duplication, and improving the effectiveness of innovative solutions (as defined in the Federal Shared Services Strategy⁷).”⁸ Above this, what the Census Bureau seeks to accomplish with this transition is to serve as a model of open government that provides more accessible government data and information assets to the American public.

⁴ For the purposes of this document, the term “content” will refer to all unstructured information, while the term “data” will refer to all structured information unless otherwise noted.

⁵ Metadata are information used to describe certain attributes of a piece of digital information, such as page title, author, date updated, and other classifications. Consistent quality metadata tagging can improve search results and also be used to structure content so that it can be more widely disseminated.

⁶ Gov 2.0 definition: The employment of the Internet and the world-wide-web for delivering government information and services to the citizens.’ (United Nations, 2006; AOEMA, 2005).

⁷ http://www.cio.gov/documents/Shared_Services_Strategy.pdf

⁸ From: <http://www.wh.gov/digitalgov/pdf>

III. An “Information-Centric” Approach

10. A necessary precursor to enabling APIs as a part of a comprehensive technology strategy is the separation or decoupling of information from its presentation. This separation represents a structural paradigm shift from content management to the management of structured data that is presentable in a wide variety of formats (e.g., social media, mashups, web site pages, web applications, mobile applications, etc.). In effect, this model represents a complete overhaul of the way our government currently provides such digital services and in a fashion that also makes it more efficient.

11. Like other agencies and organizations,, the Census Bureau has created dissemination systems for specific uses as needed. These systems inherited a tight coupling of presentation and information optimized for the specific context of the system. However, by delineating these systems so specifically, they become difficult to adapt to changing internal and external needs. This results in duplication of efforts and the building of multiple systems to serve different audiences where a single would suffice.

12. We can expose data to other computers (internal and external) via Web APIs to decoupling information from its presentation, This functions to open the data, which also means our customers can build applications that leverage the same web APIs provisioned to provide services to ourselves thus dramatically reducing development costs simultaneously for the Census Bureau and our stakeholders.

13. By leveraging an information-centric model, we can design our systems’ for portability (build once, deploy anywhere), modernize our content publication model, and deliver better, cross-platform (any device: mobile, tablet, pc) digital services at a lower cost. Before discussing how we leverage APIs at the Census Bureau, it is helpful present a well-established conceptual model of the three “layers” of digital services (see Figure 1), as also used in framing the new digital strategy.⁹

- At the base, the **information layer** contains digital resources. It includes structured information, such as census and employment data, plus unstructured information – which tend to be more text heavy - (e.g., content), such as reports, fact sheets and press releases.
- At the top, the **presentation layer** provides the user interface for the information (web sites, mobile applications, web applications, etc...)

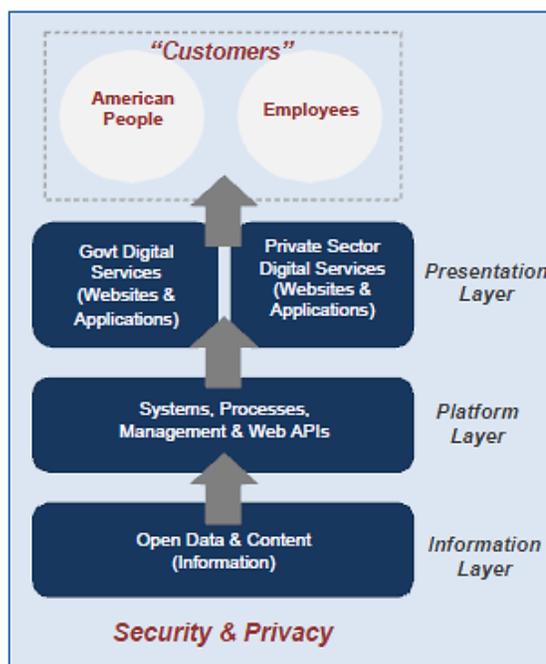


Figure 1

⁹ <http://www.wh.gov/digitalgov/pdf> page 4

- Between these, the **platform layer** includes all the systems and processes that mediate this information and serve it to the presentation layer. The API lives in this intermediary platform layer.

14. Separating our future information efforts into these three layers will allow the Census Bureau to create content and data once then use it in different ways.

15. As stated in the Digital Government Strategy:

“Government agencies must enable the public, entrepreneurs and our own government programs to better leverage the rich wealth of federal data to pour into applications and services by ensuring that data is open and machine-readable by default... (This separation helps) serve to unlock the power of government data, spur innovation, and thereby improve the quality of services for the American people.”

VI. The Data Web API and Launch Process

16. The API built by the Census Bureau (Figure 3) is a service built on the agency’s DataWeb¹⁰, which is a network that links data sources together - using a single infrastructure - across a variety of governmental agencies. The DataWeb provides easy access to data from distributed locations across the internet. It brings together demographic, economic, environmental, health, and other datasets that are usually separated by geography and/or organization.

¹⁰ For more details, see <http://dataferrett.census.gov/about.html>

2010 SF1 DataAPI Servlet Operation Diagram

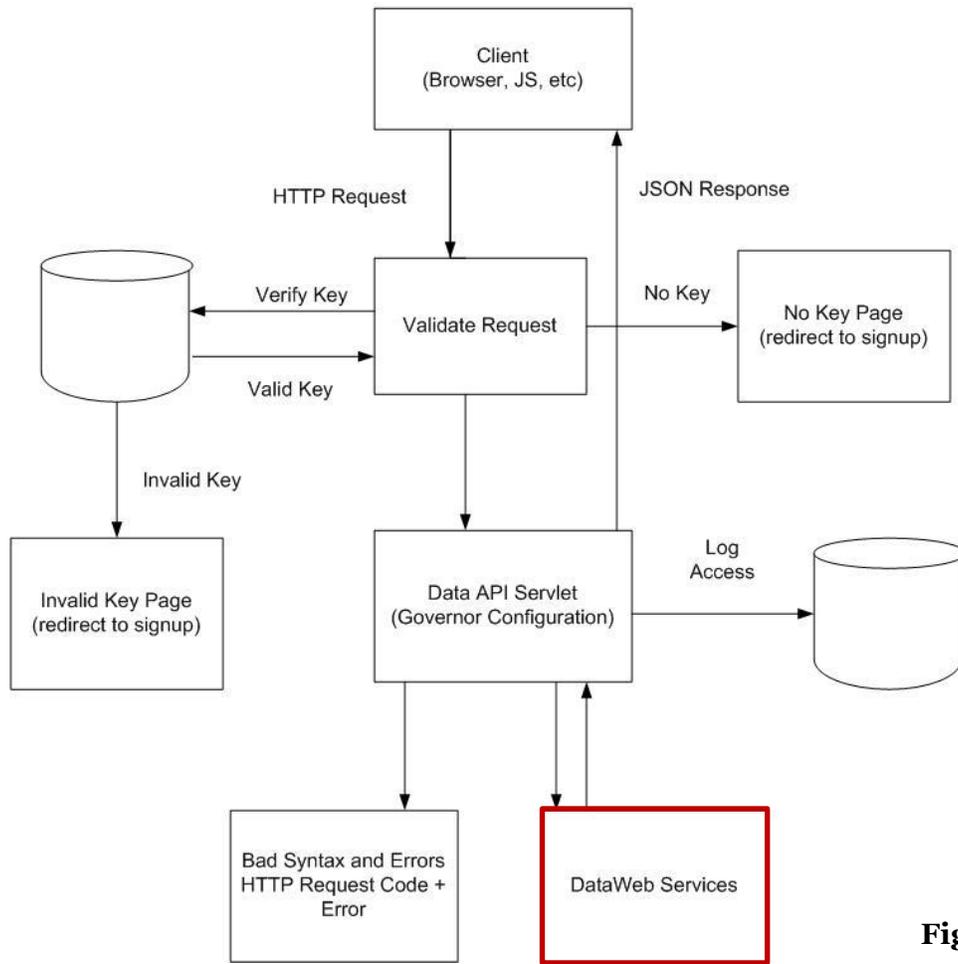


Figure 3

17. Via the DataWeb, the Census Bureau has built its first API to provide access to the American Community Survey (ACS) and our 2010 Census Summary File 1 (SF1). This API opens access to high value data, ranging from basic demographic statistics to key socioeconomic and housing characteristics like income, poverty, education, housing and transportation for every neighborhood nationwide. The architecture accommodates the addition of other data types (e.g., geospatial) and sets by using a standards-compliant metadata format.

18. A phased-launch approach allowed the agency to progressively release ‘outward’ to a broader audience after vetting by more intimately associated stakeholders. The initial launch was limited to our State Data Centers, a network of partners in each of the 50 states representing their state government as a liaison with the Census Bureau. We also sent invitations to other selected beta testers and stakeholders.

19. On April 23, the Census Bureau launched a beta version of this API. Within six hours, a developer from Cornell University ‘mashed’¹¹ an application on top of it (Figure 2). At present, the site allows third parties to develop apps using the 2010 Census Summary File 1 and the 2006-2010 American Community Survey (five-year data).

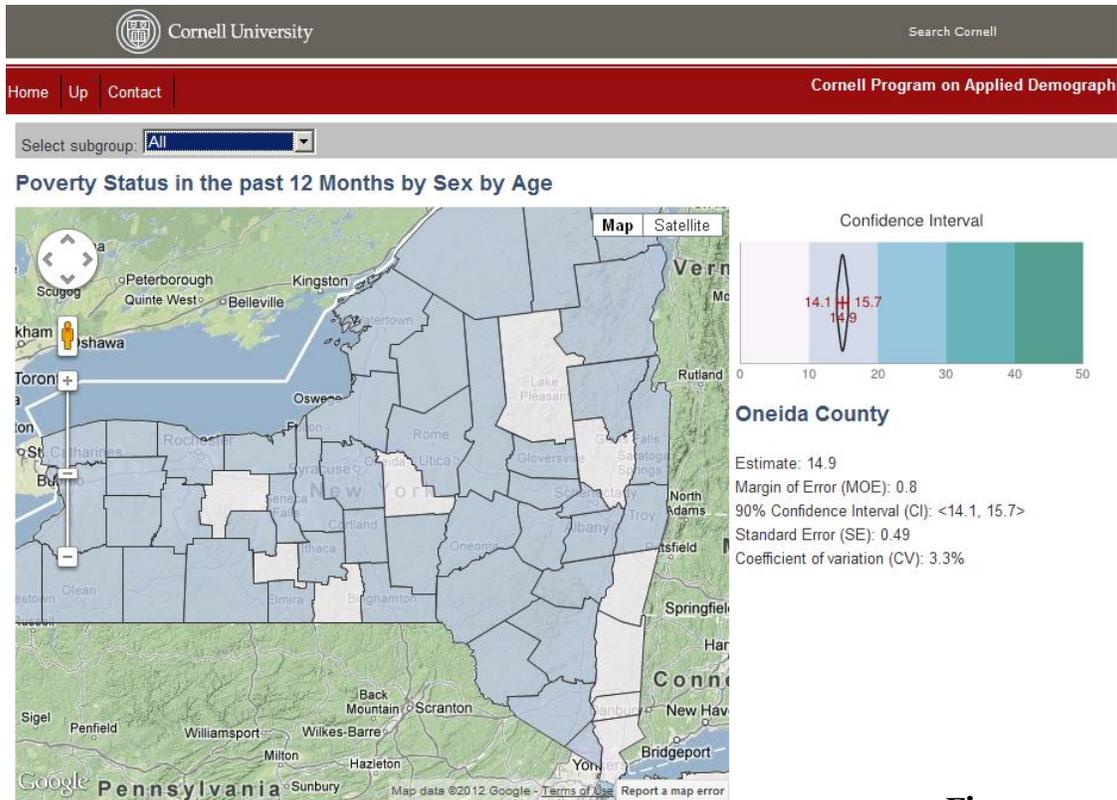


Figure 2

Source: Table B17001: American Community Survey, 2006-2010
This product uses the Census Bureau Data API but is not endorsed or certified by the Census Bureau.

20. In order to capture and organize feedback from beta testers, we set up an online collaboration space (Ideascale) that allows developers / external parties to propose applications and improvements to the APIs as well as ‘vote-up’ those they deem most salient. This forum is intended to build a community around Census Bureau APIs. Above all, this site allows the project team and subject matter experts to engage with beta testers and foster program participation around: getting/offering support, sharing ideas and suggesting improvements. Popular ideas for new features are considered by the project team in planning future enhancements and offerings.

21. During the second phase of our testing, we opened access to data reporters, federal contacts, targeted developers, and others. During this phase, one application mapped American Community Survey data across U.S. county data (<http://acsmapper.appspot.com/D3>). By June 12, developers had requested over 360 API keys.

¹¹ “Mashed”: from Mashup (web application hybrid) - a web application that combines data and/or functionality from more than one source

22. To support developers, a website hosting the technical documentation, example code and applications built on top of the API was created. While the site's Application Gallery hosts a Census Bureau example that provides access to age data, it will primarily be used to shed the spotlight on applications created by third parties.

VII. Future Plans

23. The Census Bureau API roadmap seeks to develop a broad-spectrum strategy for data provisioning ranging from large scale data delivery (as in the current use of FTP and ftp2.census.gov) through automated query-based data delivery to value-added, formatted reports/maps/etc. Future plans call for broadening access to additional census data repositories via API adding additional data sets via the current architecture. If sufficient demand surfaces, we could build a software development kit (SDK) which would abstract this multitude of variables and make the interface more user friendly for developers. Capacity may also need to be increased as usage and demand grow.

24. Internally, the use of APIs creates a culture shift from siloed software development and replicated data storage to a write-once, read-many infrastructure. The use of APIs as the foundation of future internal development of applications will allow us to separate front-end logic from back-end data access in a manner that would dramatically reduce redundancy of processing and programming effort especially with regard to database application development, which drives the majority current applications.

25. Through expanded API implementation as an internal and public service, the Census Bureau can better protect the integrity of the statistics and reduce costs associated with data storage. Furthermore, the Census Bureau can create a 'just in time' pipeline from the agency out to users that enables the most recent statistics to be available everywhere upon initial release and encourages consumers to continue to view the Census Bureau as *the* authoritative source of information about the nation's people, places and our economy.