



SHORT SUBJECTS

OPEN GOVERNMENT SERIES

May 2012

S-12-01

Public-use Data: Safe, Secure, and Supporting Open Government

Organizations prepare public-use data with the intent of making them public and ensuring that they do not expose confidential information. A common misconception about this type of data is that general access means opening the door to database corruption. This fact sheet provides information about the ways that public-use data are safe, secure, and support open government.

PUBLIC-USE DATA ARE SAFE

In most cases, public-use data do not contain individually-identifiable information and maintain anonymity and confidentiality. Prior to releasing data for public use, organizations remove identifiable information such as social security numbers and names that would make it possible to identify an individual or steal an identity. Advanced statistical techniques may also be used to mask, collapse, and recode sensitive data to further protect the identities of individuals while maintaining the validity of the data.

Some data do not need to be deidentified prior to their public use. These data may offer information about counties, states, or countries instead of about individuals. The information is aggregated and neither protected nor necessarily confidential. Some data may offer information about individuals, but the information is already in the public domain. For example, the Sacramento Bee maintains a searchable database that provides the voting records of state legislators. These individual-level data are public record.

PUBLIC-USE DATA ARE SECURE

Making data available to the public does not mean giving the public user direct, unfettered access to databases. Providing public-use data is a process of extracting data—with confidential information removed—and duplicating them so that members of

the public can access the data without interfering with business operations. The original data remain with the host organization on a different server, and that organization's policies and procedures govern the integrity of the information.

The public-use copies may be analyzed, cited, quoted, and otherwise processed by the public even as the original remains exactly as an organization stores it. When done correctly, providing public-use data does not expose the original databases to threats of unauthorized access to confidential information.

Making data available to the public does not mean giving the public user direct, unfettered access to databases.

PUBLIC-USE DATA SUPPORT OPEN GOVERNMENT AND INNOVATION

Public-use data from government agencies offer an avenue toward open and transparent government. They also stimulate business innovation, help government agencies be more efficient and publicize their work, and encourage researchers to ask and answer questions that might otherwise have gone unexamined. Already entrepreneurs and businesses are creating mobile device applications (apps) and dashboards to manage data and assess outcomes. Cross-agency transferability of data identifies duplication and might inhibit costs and widen the scope of outcomes agencies measure. Researchers may access previously inaccessible data to uncover relationships among data and develop and test new hypotheses.

EXAMPLES AND USES OF PUBLIC-USE DATA

The federal government, California, and other state and local governments already grant the public electronic access to some information. The federal government launched data.gov as a part of its Open Government Initiative and also provides public-use census data files. These public-use data may be tabular datasets or charts, graphs, interactive tools, and dashboards informed by datasets.

Public-use data inspire innovation even as they inform.

The Bay Area Rapid Transit (a.k.a. BART) has a public data feed from which entrepreneurs have created apps such as trip planners. California's Department of Justice offers the public crime trend data (see Figure 1) as well as data about domestic violence-related calls for assistance. Cities such as Seattle and states such as Maryland and Tennessee provide the public data about annual budget expenditures, real-time 9-1-1 calls, and/or public employee salary records.

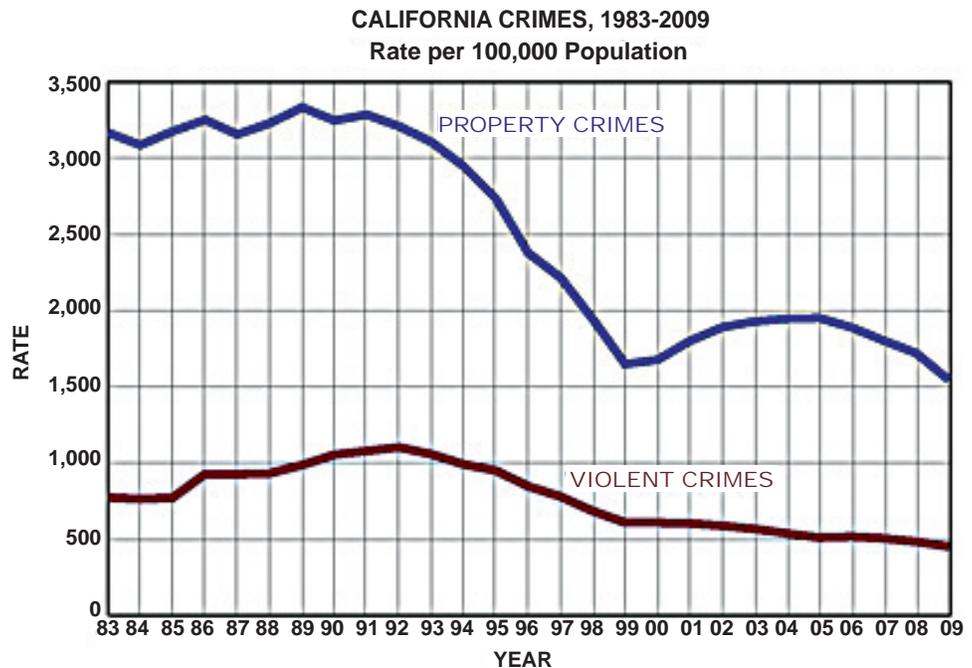
TO LEARN MORE:

McDermott, Patrice. (2010). "Building Open Government." Government Information Quarterly. Elsevier. <http://www.sciencedirect.com/science/article/pii/S0740624X10000663>.

Meijer, Albert J., Deidre Curtin, and Maarten Hillebrandt. (2012). "Open Government: Connecting Vision and Voice." International Review of Administrative Sciences. Sage. <http://ras.sagepub.com/content/78/1/10>.

Walker, Karen E., Chelsea Farley, and Meredith Polin. (2012). "Using Data in Multi-agency Collaborations: Guiding Performance to Ensure Accountability and Improve Programs." Public/Private Ventures.

Figure 1: Example of California Public-use Data



Source: <http://oag.ca.gov/crime>

ENDNOTES

1. Park, Todd. (2012). Pioneering Innovation through Health Data Transparency. <http://www.whitehouse.gov/blog/2012/03/15/pioneering-innovation-through-health-data-transparency>.
2. Open Government Dashboard "Leading Practices" for Agency Open Government Plans. <http://www.whitehouse.gov/open/documents/leading-practices-open-govt-plans>.
3. "United States Open Government Initiative Advances." (2012) Journal of E-governance 34 (1). IOS Press: Amsterdam, The Netherlands. See <http://www.data.gov/> and <http://www.whitehouse.gov/open> for more information.

The California Research Bureau, a division of the California State Library, provides in-depth research and non-partisan public policy analysis to members of the State Legislature and the Governor. Online at www.library.ca.gov/crb.

Tonya D. Lindsey, Ph.D., author of this Short Subject, can be reached at tlindsey@library.ca.gov.