DIRECT FROM CDC ENVIRONMENTAL PUBLIC HEALTH TRACKING NETWORK

Veronica Burkel, MPH

Environmental Health Tracking Rides the Open Data Wave

Editor's Note: As part of our continuing effort to highlight innovative approaches and tools to improve the health and environment of communities, the *Journal* is pleased to publish a bimonthly column from the Centers for Disease Control and Prevention's (CDC's) Environmental Public Health Tracking Network (Tracking Network). The Tracking Network is a system of integrated health, exposure, and hazard information and data from a variety of national, state, and city sources. The Tracking Network brings together data concerning health and environmental problems with the goal of providing information to help improve where we live, work, and play.

Environmental causes of chronic diseases are hard to identify. Measuring amounts of hazardous substances in our environment in a standard way, tracing the spread of these over time and area, seeing how they show up in human tissues, and understanding how they may cause illness is critical. The Tracking Network is a tool that can help connect these efforts. Through these columns, readers will learn about the program and the resources, tools, and information available from CDC's Tracking Network.

The conclusions of this article are those of the author(s) and do not necessarily represent the views of CDC.

Veronica Burkel has been an ASPPH/CDC informatics fellow with the Environmental Health Tracking Branch since February 2015. Prior to joining Tracking, she was an information analyst for the Michigan PBB Registry at Emory University.

Raise your hand if you've ever used an online map. Raise your hand if you like your phone giving you directions in the car. Did you know that the Global Positioning System (GPS), developed by the U.S. Department of Defense, only became operational for public use in 1995 and had restricted precision even then (National Coordination Office, 2014)? The U.S. military piloted and used the satellite navigation system for decades prior. Only in 1983 did President Reagan require it be provided for civilian use and, in 2000, it became available with the same precision as used by the military (Sullivan, 2012). Since then, the innovations with GPS have been and will likely continue to be limitless. Just the other day, I went for a run and my phone tracked not only my route, but also my varying running speeds along the way. How cool is that?! What other systems, tools, and data can be made available as to produce such useful innovation? Luckily, more innovation and idea syntheses are on their way with the federal government's new demand for open data.

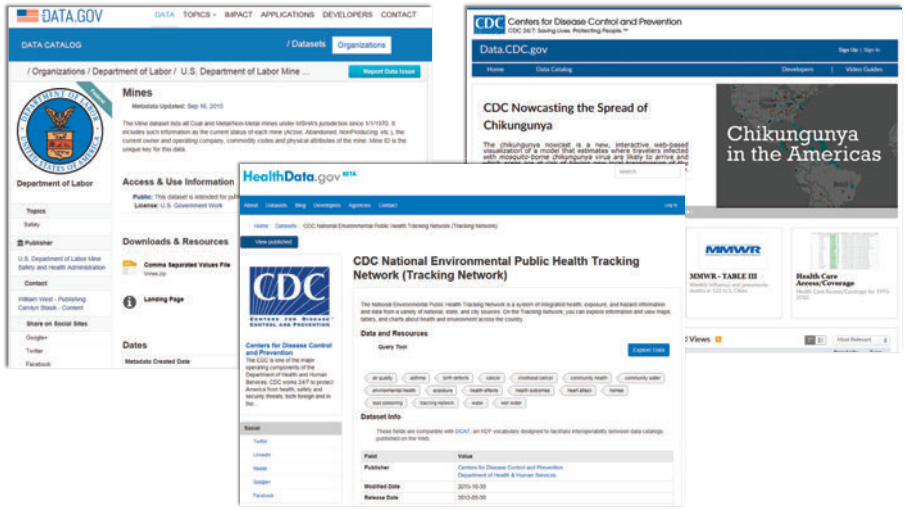
The executive order for open data was released in May 2013, 30 years after the request for public GPS. It's a federal government-wide order to make data easily available to the public in machine-readable formats. The purpose is to make appropriate data and information resources easy to find, access, and use. If history is a guide, the results of this mandate will promote entrepreneurship, innovation, and scientific discovery, which, besides creating jobs, will improve the lives of Americans in creative and unimaginable ways (Obama, 2013).

Before discussing the potential of open data, let's look at the language in the executive order. First, what is "appropriate" data? Appropriate data for public display include data that would not jeopardize individual privacy, confidentiality, or national security. Good examples include data for all planned snowmobile trails in North Dakota, or all coal mines in the U.S. since 1970. Another term to clarify is "machine readable." Do not mistake "machine readable" with "me reading on my computer"; machine readable means data are shared in a format with a standard computer language (e.g., CSV, XML, JSON files) read by a web browser, computer system, or computer program, not people-friendly formats like HTML, PDF, or DOC files.

The open data initiative is underway as I type. The federal government created data.gov, a Web site housing governmental data with standardized metadata fields (e.g., title, description, tags, last update, etc.) (Office of

FIGURE **1**

Web Sites Created as Part of the Open Data Initiative



Left to right: data.gov, healthdata.gov, and data.cdc.gov.

Citizen Services and Innovative Technologies, n.d.). Data.gov includes data from the Department of Agriculture, Department of Transportation, U.S. Environmental Protection Agency, National Aeronautics and Space Administration, Small Business Administration, Department of Justice, and other Chief Financial Officers Act (CFO-Act) agencies. Health and Human Services, also a CFO-Act agency, established healthdata.gov, a site specific to health information and datasets generated by the U.S. government. Further down the chain is data.CDC.gov, a site to centralize data from the Centers for Disease Control and Prevention (CDC) that allows in-site visualizations and machine-readable downloads. All of these Web sites (Figure 1) increase the discoverability and accessibility of government data and information. Many programs at CDC are riding this open data wave. One such program is the National Environmental Public Health Tracking Program (Tracking Program).

The purpose of the Tracking Program is, as indicated in the Pew Environmental Health Commission report in 2000 (Environmental Health Tracking Project Team, 2000), to create a nationwide system to track environmentally related exposures and diseases such that the link between environment and health may be more readily monitored. The cornerstone of this program, the National Environmental Public Health Tracking Network (Tracking Network), is a multi-tiered, Web-based surveillance system. Tracking Network data include environmental exposures, hazards, and health effects data from national, state, and city sources. Tracking provides standardized data on content areas such as asthma, air quality, community water systems, heart disease, birth outcomes, climate change, lead poisoning, cancer, lifestyle risk factors, and more. Maintaining quality, standardized, understandable, up-to-date, and precise data is Tracking's priority. The intent is for others to use this network of information to identify trends, target interventions, and explore the links between environment and health status.

How does Tracking ride the open data wave? For starters, Tracking already exists as a store with shelves full of available data (National Center for Health Statistics, 2015). A visit to the Web site will reveal environmental and health data at your fingertips. Until now, Tracking has provided data in primarily people-friendly displays with cookie-cutter visualization tools across the data measures. You can look up data based on location; understand topic areas such as carbon monoxide poisoning or water quality; explore Tracking data in map, table, or chart views; and share your data results via Facebook, Twitter, or URL. Tracking data are machine readable, too. You can download a CSV (Microsoft Excel) file. And, touching on the Tracking Program's latest project, users can access the Tracking Application Program Interface (API).

An API is a tool for accessing data in machine-readable format. It provides machine-friendly data for a person to program a machine to read and create people-friendly data tools. The typical audience for an API is programmers, as they use APIs to build software applications more easily. The applications built directly request information, machine-to-machine, from an API's code-based library to create people-friendly tools. If you've seen a YouTube video embedded in a non-YouTube Web site, this is an example of API in action: the YouTube API that allows a video to be viewed and for that view to count towards Web site analytics is integrated into the host site's page. For the user, the experience is seamless, but in the background, the Web site is dynamically interacting with other machines and systems to deliver this functionality.

The Tracking API is simply a different interface for accessing Tracking data, especially for entrepreneurs, innovators, researchers, and others (Figure 2). In contrast to the Tracking Network's people-friendly maps, charts, and tables, the API provides results in coded, machine-friendly tables. Software that retrieves these tables can more efficiently create specialized visuals and results that the Web site does not currently provide. The Tracking Network's Web site serves as a vehicle for idea and hypothesis generation, offering data for your exploration. The power of Tracking is in its ability to handle large amounts of data across a broad range of health and environmental topics, and, as such, generates the one-size-fits-all results mentioned above. Embracing open data through the Tracking API empowers users to find new ways to display health and environmental data, to make it more discoverable, and, as is the mission of Tracking, to utilize the data to improve the health and knowledge of individuals and communities.

How will you use the Tracking API? Will you create a mobile app that cross-visualizes asthma hospitalizations with air quality? A program that models how many children in each county should be tested for blood lead, based on the age of housing, and compares it to the actual number of children tested?

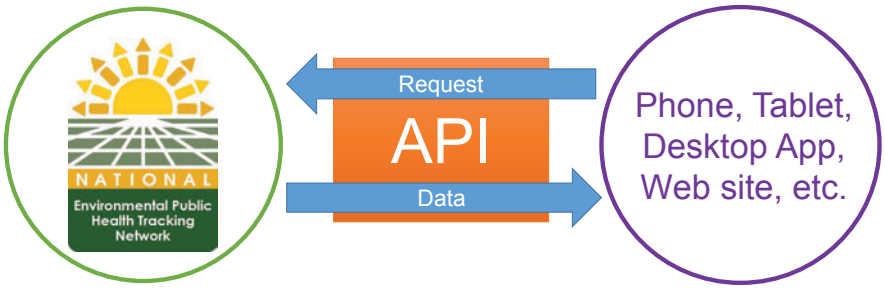
Can you find a pattern in Tracking data that nobody has yet realized? The potential for innovation with open data, similar to that of GPS and its uses, is limitless. The impact this and other open data efforts will have on public health are unimaginable.

For more information on our API, please visit http://ephtracking.cdc.gov/apihelp.

Corresponding Author: Veronica Burkel, ASPPH/CDC Informatics Fellow, Environmental Health Tracking Branch, Division of Environmental Hazards and Health Effects, National Center for Environmental Health, CDC, 4770 Buford Highway, NE, MS F-60, Atlanta, GA 30341. E-mail: xee5@cdc.gov.

FIGURE **2**

Requesting Machine-Readable Data From the Environmental Public Health Tracking Network's Application Program Interface (API)



References

Environmental Health Tracking Project Team, Johns Hopkins School of Hygiene and Public Health, Department of Health Policy and Management. (2000). *America's environmental health gap: Why the country needs a nationwide health tracking network*. Retrieved from http://healthyamericans.org/reports/files/healthgap.pdf

National Center for Health Statistics. (2015). *Interactive tables and databases*. Retrieved from http://www.cdc.gov/nchs/data_access/data_tools.htm

National Coordination Office for Space-Based Positioning, Navigation, and Timing. (2014). *Frequently asked questions*. Retrieved from http://www.gps.gov/support/faq/

Obama, B. (2013). *Executive order—making open and machine readable the new default for government information*. Retrieved from https://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

Office of Citizen Services and Innovative Technologies, U.S. General Services Administration. (n.d.). *About data.gov*. Retrieved from http://www.data.gov/about

Sullivan, M. (2012). *A brief history of GPS*. Retrieved from http://www.pcworld.com/article/2000276/a-brief-history-of-gps.html