Defining Electronic Health Technologies and Their Benefits for Global Health Program Managers

Data Science and Big Data

DEFINITION

Data science: combines "aspects of statistics, computer science, applied mathematics, and visualization," using "automated methods to analyze massive amounts of data and to extract knowledge from them"

Source: New York University. 2013. "Data Science at NYU." Available at: http://datascience.nyu.edu/what-is-data-science/.

Big data: "typically refers to data on the scale of terabytes (10 to the 12th power) and petabytes (10 to the 15th power). A petabyte is a million gigabytes."

Source: School of Information, University of California at Berkeley. 2015. "What Is Data Science?" Available at: https://datascience.berkeley.edu/about/what-is-data-science/.

The world is generating an ever-growing mountain of data. According to an estimate by IBM, 90 percent of the world's data were generated in the past two years (1). Moreover, in 2020, the amount of data the world generates will be 20 times greater than the amount generated in 2011 (2).

Data science is an emerging field that capitalizes on this growing data environment and emphasizes creative uses of data to discover the stories that data contain. Practitioners of data science draw from three main skill sets:

- Data curation
- Data analysis
- Communication of findings

Data scientists are adept at finding relevant data, analyzing them using a wide range of techniques, and then creating data products that lead to action. Voice, text, transactional, locational, and positional data can be overlaid with other data—for example, income, health, and education—generated by official sources to produce new insights into complex issues.

The increasing speed of computers to analyze data and novel hardware that lowers computational barriers to handling large amounts of diverse data have brought the high-power analytics of data science within the reach of more users. This type of analysis reveals patterns in data that lead to new perspectives on complex problems. Rayid Ghani, the Director of Data Science for Social Good Fellowship, describes data science as "helping humans in discovering new knowledge that can be used to inform decision making, or through automated predictive models that are plugged into operational systems and operate autonomously" (3).

Voice, text, transactional, geospatial, and positional data can be overlaid with other data—for example, income, health, and education—to produce new insights into complex issues

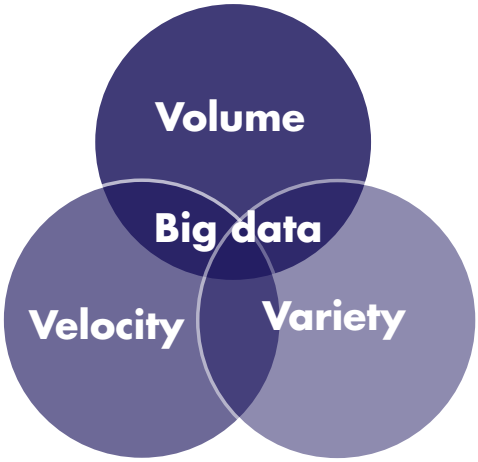
A special type of data that has gained attention in recent years is "big data." This term refers to "data sets so large or complex that traditional data processing applications are inadequate" (4).

Data Science and Big Data

While hard to quantify, big data is in the magnitude of petabytes or terabytes, while the average electronic medical record database is in megabytes—four times smaller than a big data repository.

Big data can be understood in terms of the "Three V's": "higher volume, greater variety and quicker velocity" data, coming from "sensors, social media sites, online photos and videos, online purchase records, mobile phone record signals and call records" (5, p. 16).

Data science's methods of collection, analysis, and communication are suitable not only for high-volume and -velocity data but also for data of lower volume and velocity.



For more than 20 years, MEASURE Evaluation, funded by the U.S. Agency for International Development (USAID), has helped countries build their capacity to collect and use high-quality health data for evidence-based decision making and programming. Building on this history, we are conducting a proof of concept activity on how data science techniques can help answer global health questions, including improving health program management.

What Can Data Science and Big Data Do for Global Health Program Managers?

Data science has been applied most commonly in sectors other than health. An example of this type of work is a nongovernmental organization's use of data science to identify program areas, by analyzing satellite images. GiveDirectly provides direct cash transfers via mobile payments to people living in extreme poverty. To make sure their programs reach those most in need, they used an automated computer model to scan Google Earth images and identify areas with many thatched roofs. Because metal roofs cost more than thatched, this data can be a proxy indicator of concentrations of poverty. GiveDirectly used the model to decide where to focus the program (6).

Data science is beginning to be applied in global health, and one of the data sets with most potential is data from mobile phones (7). In Kenya, for example, Harvard University's Engineering Social Systems lab successfully used mobile phone data and census data to model how slums are growing. The lab provided this information to urban planners to inform their decisions about where to allocate new latrines and water pumps (8). Data science techniques were helpful both in analyzing the data and communicating the findings.

The Malaria Atlas Project (MAP) applies data science methods to cull data from many sources, estimate data gaps, and provide trustworthy data to program managers. A challenge facing malaria program managers is that the available malaria data are highly clustered; epidemiological information on Central Africa is sparse. Moreover, variability in the data that exist makes them challenging to use. Efforts to account for this variability are under way, but much remains unexplained. Data science addressed the problem of converting this patchy dataset, with substantial unexplained variation, into maps that allow program managers to evaluate malaria risk. To simulate possible malaria risk, the research team applied a geostatistical model to climate data, health system data, parasite prevalence data, as well as survey data such as USAID's Demographic and Health Surveys. This approach gave the decision makers confidence to design malaria prevention and treatment programs (9).

Spatial distribution of malaria endemicity in Nigeria in 2010. Malaria Atlas Project website.

An example of big data in global health comes from the 2014 Ebola outbreak. Researchers used data sources such as cellphone data and social media postings

Data Science and Big Data

from multiple countries to help track and respond to the quickly moving virus. Because cellphone calls are routed through the nearest cell tower, individual location can be approximated. These types of data helped program managers see where people were moving, which was critical for making decisions to combat a disease spread by human contact. Using advanced computing, they synthesized data from these sources to create a virtual city, in which they ran multiple simulations of the spread of the disease. This information helped them to plan data-informed prevention and mitigation programs (10).

Data scientists used mobile phone data and census data to model the growth of slums, helping urban planners decide where to allocate new latrines and water pumps

For more information on data science, go to:
http://search.proquest.com.libproxy.lib.unc.edu/docview/1288621449?pq-origsite=summon

For more information on big data in global health program management, visit:
http://www.who.int/bulletin/volumes/93/3/14-139022/en/

For more information on MEASURE Evaluation, visit:
www.measureevaluation.org

BIBLIOGRAPHY

1. School of Information, University of California at Berkeley. 2015. "What Is Data Science?" Available at: https://datascience.berkeley.edu/about/what-is-data-science/.

2. New York University. 2013. "Data Science at NYU." Available at: http://datascience.nyu.edu/what-is-data-science/.

3. Shan, C. "How Data Science Can Be Used For Social Good." Carl Shan [blog], January 8, 2015. Available at: http://www.carlshan.com/2015/01/08/data-science-social-good.html.

4. Janssen, C. "Big Data: Definition." Techopedia, n.d. Available at: https://www.techopedia.com/definition/27745/big-data.

5. Raftree, Linda and Bamberger, Michael. 2014. Emerging Opportunities: Monitoring and Evaluation in a Tech-Enabled World. Itad and the Rockefeller Foundation Evaluation Office: Rockefeller Foundation.

6. Abelson, B., Varshney, K.R., and Sun, J. 2014. "Targeting Direct Cash Transfers to the Extremely Poor," Pp. 1563–1572, in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: Association for Computing Machinery.

7. Boor, Paul van der and Wesolowski, Amy. "Mobile Phone Data Are a Treasure-trove for Development." SciDev.net, June 19, 2013. Available at: http://www.scidev.net/global/icts/opinion/mobile-phone-data-are-a-treasure-trove-for-development.html.

8. Bhushan, Aniket. 2012. "Big Data." Pp. 152–156, in *USAID Frontiers in Development*, edited by Rajiv Shah and Steve Radelet. Washington, DC: United States Agency for International Development.

9. Patil, A. P., Gething, P. W., Piel, F. B., and Hay, S. I. 2011. "Bayesian Geostatistics in Health Cartography: The Perspective of Malaria." *Trends in Parasitology* 27(6), 246–253.

10. Caulderwood, K. "Ebola Outbreak 2014: How Researchers Use Big Data To Help Stop The Disease." *IBT Pulse*, September 25, 2014. Available at: http://www.ibtimes.com/pulse/ebola-outbreak-2014-how-researchers-use-big-data-help-stop-disease-1694907.

Defining Electronic Health Technologies and Their Benefits for Global Health Program Managers

Data Science and Big Data

ALL ABOUT eHEALTH

Electronic health (eHealth) refers to the health sector's use of information and communication technologies (ICT) such as mobile phones, portable and handheld computers, Internet and cloud-based applications, open source software, and data warehouses. Advances in ICT have increased exponentially the amount of data that health information systems can collect, synthesize, and report. Expansion of these technologies in low- and middle-income countries (LMICs) promises to revolutionize the global health sector's response to these countries' most pressing health issues.

MEASURE Evaluation—funded by the U.S. Agency for International Development—seeks new ways to exploit such eHealth solutions as data dashboards and geospatial data analysis, as part of its mandate to strengthen health systems in low-resource settings. Even though health program managers in LMICs—as everywhere—are increasingly expected to use and invest in such strategies, many lack information about how the strategies work and how they can benefit the management of health programs.

To address this problem, we developed this glossary of eHealth strategies most likely to enhance data access, synthesis, and communication for health program managers at all levels of a health system who are eHealth novices. The list has been vetted and revised by an advisory group representing the World Health Organization, the Free University of Free Brussels/European Agency for Development and Health, the University of Oslo, the Public Health Foundation of India, and the National Institute of Public Health Mexico.

The complete set consists of fact sheets on the following eHealth strategies, in addition to this one:

- **Dashboards**
- **Crowdsourcing**
- **Hackathons**
- **Open data**
- **Geospatial analysis**
- **Integration and interoperability**
- **App competitions**

In each fact sheet, you'll find the following information:

- eHealth strategies that have been used in health information system strengthening efforts to improve access to and synthesis, presentation, and communication of health data for program management
- How the strategies have been adapted (or not) from their application in resource-rich country settings to health programs in LMICs
- An example of the strategy for global health program management
- Links to additional resources for more in-depth details on the strategies

USAID FROM THE AMERICAN PEOPLE

MEASURE Evaluation

MEASURE Evaluation is funded by the U.S. Agency for International Development (USAID) under terms of Cooperative Agreement AID-OAA-L-14-00004 and implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International, John Snow, Inc., Management Sciences for Health, Palladium, and Tulane University. The views expressed in this presentation do not necessarily reflect the views of USAID or the United States government. FS-15-165c