Data Science for Health Decision Making

"Data science is the process of finding, developing, and communicating actionable information that stems from multiple sources."

—John Spencer,
Senior GIS Technical Specialist

Photo by Wayne Hoover, MEASURE Evaluation

Data science is a tool relatively new to the field of public health. The public health field is just beginning to explore what data science can do to improve health outcomes. Data science is the process of finding, developing, and communicating actionable information that stems from multiple sources. For instance, data science might bring together information from household surveys, routine health information systems (RHIS), and non-traditional sources like mobile phone data to model outbreaks of disease or predict effects of health interventions.

Data science takes advantage of a global tsunami of information: in volume and variety, more data is available from more places and sources than ever before in human history. Traditional sources of data (surveys, routine tracking and monitoring systems, for example) are rapidly expanding. At the same time, every year brings new data capture possibilities (such as increasingly powerful and cheap monitoring capabilities) as well as expanded mechanism for self-generated data (e.g., social media) Together, they are driving exponential growth in the global information tsunami. All indications suggest this process will only accelerate and grow ever broader.

This data tsunami contains an almost limitless amount of vital, actionable information that can improve health decision making. This information can allow us to define and answer questions previously beyond our reach, improving health and human welfare in the process. Fortunately, these changes come at a time of growing appetite and emphasis on the need for reliable health information systems. The data available and the will to use it can help us develop stronger systems and more innovative ways to address health needs.

MEASURE Evaluation is exploring the potential of data science to strengthen health systems and improve monitoring and evaluation and more generally yield previously unobtainable, policy-relevant, actionable information. We are identifying best practices for data science in global health, and developing tools that facilitate the management, analysis, and communication of data using the data science framework.

The convergence of need and the technology to address it is at the heart of data science. Data science integrates four distinct tasks that must work together to realize the

enormous value of the information currently "hidden" in the information tsunami:

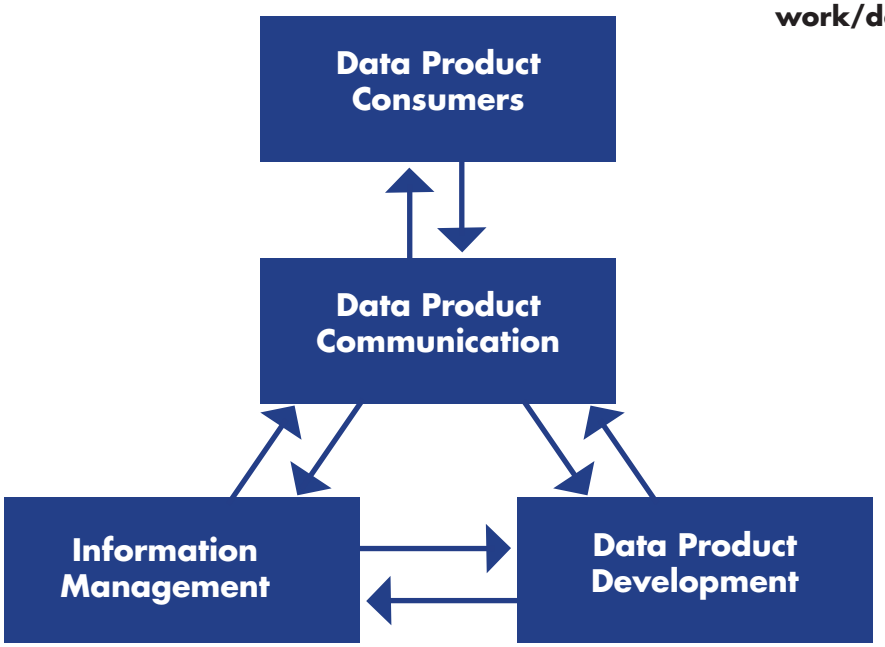
- **Define information needs**. Communicate with decision makers to understand their information needs and define what information would be valuable and actionable.
- **Manage, extract and use information**. From relevant data in existing sources of information, the next step is to identify actionable information to answer defined needs and translate it into formats suitable for analysis.
- **Create information products**. An information product such as a report, program, or dashboard synthesizes the information needed for decision making. Because technology and data evolve, data products also should evolve so information is conveyed in the most powerful and effective way.
- **Exploit the full value of the information products**. Disseminate products to decision makers, or other data consumers, in compelling ways that make them aware of new information and that suggest and support how to use it to greatest effect.

Integrating these four tasks, represented in the diagram below, is the key to the successful practice of data science. It's important to note that the process is reiterative—better data used effectively expands our understanding of the possibilities provided by the data tsunami.

MEASURE Evaluation is focusing its strength on several key data science capabilities:

- innovative analytical techniques to address issues in global health
- tools and techniques for maximizing data use to address programmatic questions
- developing a strategy for building capacity for the practice of data science.

For more information:
Visit **www.measureevaluation.org/measure/our-work/data-science**



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 Group, and Tulane University. The views expressed in this presentation do not necessarily reflect the views of USAID or the United States government. FS-15-154