



Using big data to solve big problems

November 14, 2019

“The ability to collect information far outpaces the ability to fully utilize it,” says Laboratory researcher Sara Del Valle. “Yet that information may hold the key to solving some of the biggest global challenges facing the world today.”

Take, for instance, the frequent outbreaks of water-borne illnesses as a consequence of war or natural disasters. What if we could better understand the environmental factors that contributed to the disease, predict which communities are at higher risk, and take action to stem the spread?

Answers to these questions—and others like them—could help avert catastrophe.

Data is already collected about virtually everything, from birth and death rates to crop yields and traffic flows. IBM estimates that, each day, 2.5 quintillion bytes of data are generated—equivalent to producing all the information in the Library of Congress more than 166,000 times every 24 hours.

“The power of all this information is not fully harnessed,” says Del Valle. “It’s time to change that—and thanks to recent advances in data analytics and computational services, we finally have the tools to do it.”

Data scientists at Los Alamos National Laboratory study data from wide-ranging, public sources to identify patterns, aiming to predict trends that could threaten global security.

For example, knowing mosquito incidence in communities would help public health officials predict the risk of mosquito-transmitted disease such as dengue, the leading cause of illness and death in the tropics, or West Nile virus, which has been found in New Mexico each year since 2003. However, mosquito data at a global (and even national) scale is not available.

To address this gap, the Laboratory is using other sources such as satellite imagery, climate data and demographic information to estimate risk. Using these data streams, as well as clinical surveillance data and Google search queries that used terms related to the disease, The Laboratory has developed a model that successfully predicts the spread of dengue in Brazil at the regional, state and municipality level.

While the predictions aren’t perfect, they show promise. The researchers’ goal is to combine information from each data stream to further refine the models and improve their predictive power.

The potential exists for big data to solve big problems. Los Alamos and other national laboratories that are home to some of the world’s largest supercomputers have the computational power augmented by machine learning and data analysis to take this

information and shape it into a story for not only one state or even nation, but the world as a whole.

As Del Valle says, "The information is there. It's time to use it."

An extended version of this story first appeared in the [Science on the Hill](#) series of articles in the Santa Fe New Mexican

Los Alamos National Laboratory

www.lanl.gov

(505) 667-7000

Los Alamos, NM

Managed by Triad National Security, LLC for the U.S Department of Energy's NNSA

