

Big Data Sharing for Better Health

October 10, 2014 |

Researchers at University of California, San Diego School of Medicine have been awarded a \$9.2 million grant to help modernize and transform how researchers share, use, find and cite biomedical datasets.

The 3-year project, in collaboration with The University of Texas Health Science Center at Houston, is part of a federal initiative to increase the utility of biomedical research data, launched this week by the National Institutes of Health (NIH) through its Big Data to Knowledge (BD2K) program.

The agency plans to invest nearly \$656 million through 2020 to encourage biomedical data sharing and re-use, accelerating the pace of new medical discoveries at lower cost to U.S. citizens who fund basic research.

“Data creation in today’s research is exponentially more rapid than anything we anticipated even a decade ago,” said NIH Director Francis S. Collins, MD, PhD. “Mammoth data sets are emerging at an accelerated pace in today’s biomedical research and these funds will help us overcome the obstacles to maximizing their utility. The potential of these data, when used effectively, is quite astounding.”

Data sharing and the ability to include multiple “big data” datasets in research studies could help scientists find patterns among diseases, genes and lifestyle that might easily go unnoticed in smaller datasets. These patterns could have virtually endless applications in advancing health, for example, by helping to identify those at higher risk for breast cancer, heart attack or other diseases and conditions. Researchers might also more rapidly identify rare side effects of certain medications or the benefits of new drugs to small subsets of individuals.

As part of this vision, the UC San Diego-led team will develop a strategy for cataloging and indexing biomedical datasets, coined “big data” because of the volume, variety and speed at which information – anything from whole genome sequencing to social media tweets – are being collected in the digital era.

Dataset indexing is considered a vital step toward being able to build a searchable online digital library, much like the highly successful online PubMed directory, but for health-related datasets.

“You can’t go online right now and search for datasets on, say, a particular type of brain tumor,” said Lucila Ohno-Machado, MD, PhD, professor of medicine and lead investigator on the Biomedical and healthCARE Data Discovery and Indexing Ecosystem (BioCADDIE). “These specialized search engines don’t exist. We are starting almost from scratch. I think this might surprise people.”

Because much of the BioCADDIE project is a modern derivative of library science, UC San Diego researchers will be collaborating with UC San Diego biomedical librarians with expertise in cataloging and indexing digital publications.

The BioCADDIE project will also address key differences between digital publications and biomedical datasets, notably in the importance of maintaining patient privacy.

“Before we start extracting and compiling information from research studies and electronic health records, we have to ensure methods for protecting privacy,” said Ohno-Machado, who is also founding chief of the Division of Biomedical Informatics at UC San Diego School of Medicine. “This is a major focus of the project.”

Yet another focus is overcoming the sociocultural hurdles to data sharing. Specifically, many researchers don’t want to share their data until they have mined it for multiple publications, which can take years.

The BioCADDIE project will examine strategies for incentivizing data sharing. One option under consideration is to assign a type of authorship to datasets, which could be included in citation indices and counted toward funding, “tenure points” and other forms of career advancement. Stakeholders in the science community, including archivists, librarians, journal editors and publicists, are invited to share other ideas.

“The culture of science is currently based around peer-reviewed publications and the idea that the primary product of research is a peer-reviewed narrative,” said Maryann Martone, PhD, a professor-in-residence in the Department of Neurosciences at UC San Diego who will be leading outreach efforts to the broad scientific community on BioCADDIE goals.

“We are looking at ways to change this,” said Martone, who is also president of Force 11, a non-profit working to enhance scholarly communication through information technology. “One idea is to begin to credit scientists who produce datasets that are used in multiple studies.”

The BioCADDIE project is part of the BD2K Data Discovery Index Coordination Consortium, one of four main components of the BD2K initiative.

Other components include the creation of 11 multi-institutional Centers of Excellence for Big Data Computing, which were also announced by the NIH as part of its initial \$32-million investment for the 2014 fiscal year in BD2K.

These centers will work to develop tools and methods for advancing various aspects of big data science, such as managing data from electronic health records or conducting analyses of genomic datasets.

Kevin Patrick, MD, and Jacqueline Kerr, PhD, both professors of family and preventive medicine at UC San Diego School of Medicine, are part of the National Center of Excellence for Mobile Sensor Data-to-Knowledge, through their affiliation with the Center for Wireless and Population Health Systems in the California Institute for Telecommunications and Information Technology's Qualcomm Institute. The Calit2 researchers have pioneered the use of GPS data to monitor and improve health behaviors.

The BD2K initiative, launched in December 2013, is a trans-NIH program with funding from all 27 institutes and centers, as well as the NIH Common Fund. The NIH program is being developed in the context of a number of related projects elsewhere in the world, including those under development in the United Kingdom and Australia, and by the European Union.

#

Media contacts: Scott LaFee or Christina Johnson, 619-543-6163, slafee@ucsd.edu

Share This Article

Related News

[Three UC San Diego Researchers Receive Top Honors with NIH Director's Awards](#)
10/1/2019

[With Landmark Gift, UC San Diego Will Map Compassion in the Brain, then Prove its Power](#)
7/22/2019

[UC San Diego Ranked Ninth in World in Biomedical Sciences](#)
5/16/2019

[UC San Diego Researcher Gets \\$4 Million NCI Award to Study Cancer Drug Resistance, Spread](#)
12/7/2018

[View All News >](#)

Follow Us

 Follow @ucsdhealth