

This forum is dedicated to personal health in all its many facets: decision-making, goal setting, celebration, discovery, reflection, and coordination, among others. We look at innovations in interactive technologies and how they help address current critical healthcare challenges. — **Gillian R. Hayes, Editor**

Exploring Personal Data for Public Good Research

Matthew J. Bietz, University of California, Irvine, **Cinnamon Bloss**, University of California, San Diego, **Job Godino**, University of California, San Diego, **Kevin Patrick**, University of California, San Diego

Lifestyle and environment are important determinants of individual health and wellbeing. In the U.S., sedentary habits, poor diets, and a lack of sleep are widespread, contributing to the epidemics of obesity, diabetes, and cardiovascular disease. Exposure to environmental pollution is on the rise, linked with increased risk for asthma, cancer, and a variety of other diseases and disorders. Studying these issues at the level of the individual has been a challenge for health researchers. However, new data streams and digital traces emerging from personal sensors and devices, social media, and other electronic systems have the potential to change this and transform our understanding of human health.

New technologies produce a different kind of data. Traditional health data relies on two primary data sources: clinical assessments (whether experimental trials or the monitoring of medical data) and surveys. Data that comes from clinical assessments tends to be highly controlled and collected by trained observers using validated instruments and protocols. However, each clinical data point is costly to collect in terms of both time and money, with data typically collected only at periodic intervals when the patient comes into the doctor's office or lab. Survey methods provide an opportunity to generate large sample sizes, but they have the potential for self-report and retrospective biases. Surveys and public health surveillance data help with discovering trends and

outbreaks, but tend to focus on the level of communities or populations.

Personal health devices and digital traces allow us to see human health at a higher resolution. Data can be collected continuously, not just at periodic visits to the doctor. The sensors are relatively inexpensive (and their cost keeps decreasing) and are being included not just in dedicated activity monitors but also in all sorts of wearable and personal technology like cellphones and clothing. The data is collected where the person lives and works, in the course of everyday activities, lending a higher level of ecological validity to the measurements. This data can shine a light onto aspects of lifestyle and health that were previously difficult, if not impossible, to measure.

There is a growing call for "precision medicine," the tailoring of healthcare to individual patients based on a better understanding of individual characteristics. While the term is most often used in reference to genetic profiling (e.g., for cancer risk

or drug efficacy), individual lifestyle and environment are also key factors. Knowing, for example, whether someone leads an active lifestyle, is getting enough sleep, or is exposed to high levels of pollution is important in diagnosing and treating many conditions. In the U.S., the National Institutes of Health (NIH) has launched a Precision Medicine Initiative, with goals including the combination of genomic and medical record data with "lifestyle data, such as calorie consumption and environmental exposures tracked through mobile health devices" to conduct research into human health (<http://www.nih.gov/precisionmedicine/future.htm>).

Unlike data collected explicitly for research, personal health data is often the byproduct of the Internet and the Internet of Things. In a person's individualized contexts, an activity tracker can help someone get more exercise, a sleep monitor can track insomnia, a smart fridge might remind its owner to buy more milk, or a social media post can communicate with friends and family. But when brought together and analyzed along with similar data from thousands (or even millions) of other people, this data can provide a powerful new way to understand the relationships among, for example, exercise, sleep, diet, and mood. Accessing, aggregating, and interpreting this data in a meaningful and responsible fashion is, however, not straightforward. The Health Data Exploration project was formed in 2013 to identify challenges and catalyze the use of personal health data for public good research.

Insights

- Personal health devices and digital traces can provide valuable data for human health research.
- With appropriate protections, individuals are generally willing to share personal data if it will contribute to the public good.
- Human-centered design is necessary to address issues of privacy, adoption, informed consent, and social justice.



STAKEHOLDER CONCERNS

There are three key sets of stakeholders whose interests must line up to use personal health data for research. First, the individuals who generate personal health data must be willing to share it with researchers. Second, researchers must be willing to use these new kinds of data. Finally, the companies and organizations that collect and manage the data from these devices and apps must be willing to give access to researchers.

In mid-2013, we conducted surveys and interviews to better understand the attitudes and concerns of each of these three groups of stakeholders. We surveyed 465 individuals and 134 researchers interested in tracking their own health and conducted interviews with 11 of the individuals, 9 researchers, and 15 key informants from companies

working in this area. The results of these investigations gave us confidence that while there are a number of challenges that need to be addressed, there is significant opportunity to better understand human health and wellness. Our findings include:

- *Individuals are willing to share their data for research.* However, it matters to individuals how the data will be used. Many of our participants indicated that they did not want their data used for

marketing or by for-profit companies.

- *Privacy is a significant issue.* For our respondents, privacy was a complex set of concerns about what data would be shared, how the data would be used, who would have access to the data and for how long, what regulations and legal protections were in place around the data, and the individuals' ability to know each of these factors and control the fate of their data. Individuals want to know that identifiable data will not fall into the wrong hands or be used against them by employers, insurance companies, government entities, or others.

- *Current ethical norms and regulatory procedures are inadequate.* We do not yet have a good understanding of the risks associated with revealing personal health data. Many standard procedures for informed consent and safety monitoring are cumbersome, if

Algorithms tuned to the gait of an athlete may not correctly register the shuffling steps of an older person.

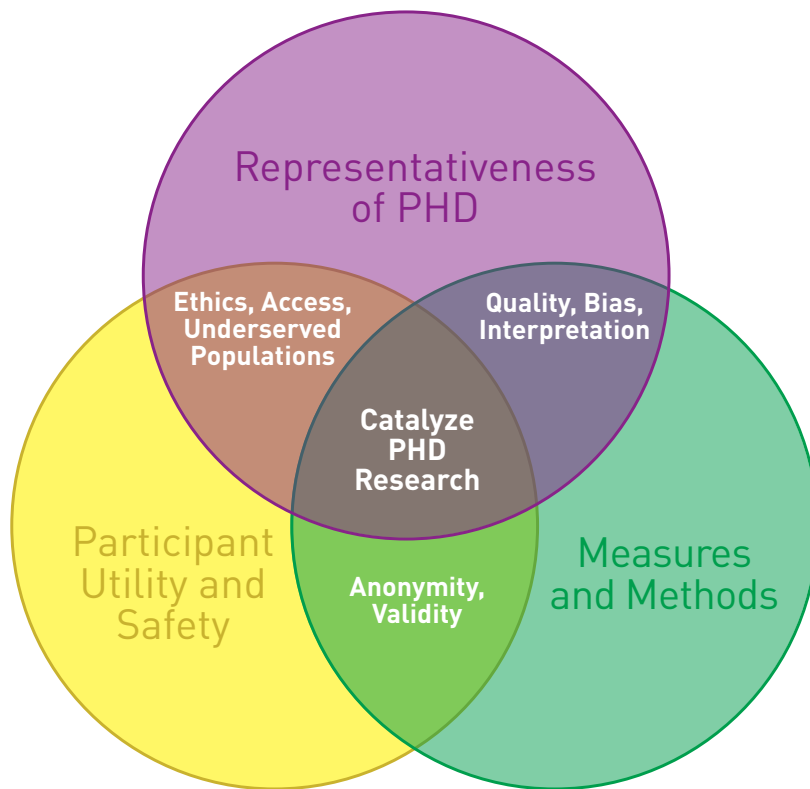


Figure 1. Areas of concern for personal health data (PHD).

BUILDING A NETWORK

With the support of the Robert Wood Johnson Foundation, we are building a network of researchers, companies, and other stakeholders to advance the use of personal data for public good research. The network is a space to foster new collaborative research, engage in policy debates, share best practices and emerging themes, and develop agendas for future research. Membership in the network is open to anyone who wishes to join. We currently have more than 100 individual members and 15 organizational members, and the network continues to grow. We invite anyone who is interested in this area to join our network at <http://hdexplore.calit2.net>.

not impossible, to implement in this new research medium.

- *Researchers see personal health data as complementary to traditional health data.* These new data sources won't replace the data that researchers already use, but can rather allow them to answer questions in new ways.

- *Researchers are willing to use personal health data.* While personal health data presents new quality and interpretation issues, the researchers in our study generally did not see these as insurmountable obstacles.

- *Legal agreements to share data are a significant hurdle.* Many of the researchers reported that the biggest roadblocks they faced involved developing workable data-sharing agreements between companies and universities.

- *Many companies would be interested in sharing their data with researchers.* Working with a researcher can provide expertise in data analysis that the company may not have in-house. Companies also gain status when researchers use their data. However, preparing data for sharing is costly, and many companies do not have the resources

available to provide data for free.

- *Companies want to protect their customers.* Companies place a high value on the trust they have with the users of their products. Breaches of privacy or any other use of personal data that makes customers lose trust in the company must be avoided.

Full results from the surveys and interviews are available in the *Personal Data for the Public Good* report [1].

CATALYZING PERSONAL HEALTH DATA RESEARCH

Based on the results of our initial investigations, we have identified three broad areas of concern that must be addressed in order to advance the use of personal data in public good research (Figure 1). Many of the challenges we identified may be addressed through human-centered design.

The first area involves the utility and safety of personal health data. Many of the issues that arise in this area have to do with the challenges of moving data among individuals, companies, and researchers. First, individuals must feel safe sharing their data. We need better mechanisms that allow participants to control who has access to their data and how it will be used. Recent innovative experiments are proposing new models like profile-based privacy controls, or micro-payment architectures that allow users to be paid for their datasets (e.g., [2,3]). Similarly, data-sharing agreements between research institutions and device and app companies need to protect the business interests of the companies while allowing enough access to raw data that researchers feel confident in drawing conclusions. Work in these and similar areas will have a significant impact on the ability to conduct research with personal health data.

We also need to develop a clearer policy environment so individuals trust that sharing data does not place them at significant risk of harm in terms of employment decisions, insurance coverage, and so on. Using data in research depends on informed consent processes. Many users have granted companies almost unlimited use of their data by agreeing to device and app terms of service. But given how few individuals bother to read,

let alone understand, “click-through agreements,” it is not clear that this qualifies as “informed” consent that would satisfy researchers’ ethical obligations. Thus, there is a design opportunity in creating ethical and human-centered informed consent processes.

The second set of issues addresses the representativeness of personal health data. These concerns operate at multiple levels. At the level of an individual’s data, we wonder whether personal health data truly represents what we want to measure. Is a step counter worn sporadically representative of an overall level of activity? Are traces left by a single account on a single device representative of the ecology of devices and activities that an individual might use during a day?

We are also concerned with the representativeness of the personal health datasets as a whole. In other words, who is generating personal health data, and how does that affect our ability to generalize from those datasets? Many wearable devices remain beyond the financial reach of lower-income individuals. Adoption of these technologies and willingness to share their data may also be influenced by a variety of social, cultural, demographic, and geographic dynamics. In some cases, the design of the technologies themselves may prevent data collection from particular groups of people. For example, individuals whose bodies fall outside a device manufacturer’s size ranges may be unable to use wearable technologies. Algorithms tuned to the gait of an athlete may not correctly register the shuffling steps of an older person. Understanding the patterns of inclusion and exclusion that shape these datasets is essential to ensure that research is not based on biased samples.

These technologies can also serve as healthcare delivery platforms. The push for precision medicine involves using this data to better diagnose and treat health problems whose symptoms may be elusive during brief office visits. Many researchers and clinicians also hope that by providing real-time feedback, these devices

can prompt users to make healthier decisions about activities, diet, and personal interactions. However, it is important to avoid creating new digital and medical divides. For both research quality and social justice reasons, we must work to minimize and address obstacles to participation.

The third area of concern for personal health data has to do with the kinds of data that are collected and how we make sense of it. Traditional sources of health data for research tend to be highly standardized, collected by trained professionals, and carefully tuned to specific research questions. Personal health data, on the other hand, tends to be much less standardized, collected in uncontrolled situations, and generated by a heterogeneous collection of general-purpose sensors that may not have been clinically validated. While traditional datasets tend to be small collections of relatively high-quality and controlled data, PHD datasets tend to be much larger, much noisier, and often more difficult to interpret. Researchers need new methods for integration and analysis of large volumes of unstandardized data.

While not an exhaustive taxonomy, these three areas of concern and their overlaps create a useful classification of the challenges in using personal health data for public good research.

DESIGN OPPORTUNITIES


A clear takeaway from this project is that there is a huge need for human-centered design in this area. Many of the challenges that we outlined here have HCI and user experience components. How do we create usable privacy solutions? How can we account for the variety of human shapes, sizes, needs, and values that affect the adoption of personal health technologies and the willingness to share personal data? What are appropriate infrastructural solutions to address issues like privacy and data integration?


The Health Data Exploration Network brings together people from a variety of different viewpoints to engage in solving some of these problems. The size and scope


of the challenges demands an interdisciplinary approach that grapples with the tensions of individual privacy and security, corporate business interests, and high-quality, evidence-based research. Human-centered design must be part of the conversation.


ENDNOTES

1. Health Data Exploration Project. *Personal Data for the Public Good: New Opportunities to Enrich Understanding of Individual and Population Health*. Calit2, San Diego, CA, 2014.
2. Liu, B., Lin, J., and Sadeh, N. Reconciling mobile app privacy and usability on smartphones: Could user privacy profiles help? *Proc. of the 23rd International Conference on World Wide Web*. ACM, New York, 2014, 201–212. DOI: <http://dx.doi.org/10.1145/2566486.2568035>.
3. Tasidou, A., Efraimidis, P.S., and Katos, V. Economics of personal data management: Fair personal information trades. In *Next Generation Society. Technological and Legal Issues* (Vol. 26). A. Sideridis and C. Patrikakis, eds. Springer, Berlin Heidelberg, 2010, 151–160. DOI: http://dx.doi.org/10.1007/978-3-642-11631-5_14.

 **Matthew J. Bietz** is an assistant research scientist in the Department of Informatics at the University of California, Irvine, where he studies the design of sociotechnical systems for data sharing and distributed collaborative knowledge work.
→ mbietz@uci.edu

 **Cinnamon Bloss** is an assistant professor in psychiatry and family medicine and public health in the Division of Health Policy at the University of California, San Diego and a policy analyst at the J. Craig Venter Institute. Her current research focuses on the individual and societal impacts of emerging health and biomedical technologies.
→ cbloss@eng.ucsd.edu

 **Job Godino** is a research associate in the Department of Family Medicine and Public Health at the University of California, San Diego. His current research focuses on the development and evaluation of interventions that utilize mobile and wearable technology to promote healthy changes in physical activity, sedentary behavior, and diet.
→ jgodino@ucsd.edu

 **Kevin Patrick** is a professor and director of the Center for Wireless and Population Health Systems at the University of California, San Diego Qualcomm Institute/Calit2. His research explores how to use mobile and social technologies to measure and improve the health of individuals and populations.
→ kpatrick@ucsd.edu