



Quantified Self Public Health Symposium

April 2014

Support
provided by



Robert Wood Johnson
Foundation

Additional support was provided by
The Qualcomm Institute of Calit2 at
UCSD and the U.S. Department of Health
and Human Services, Office of the CTO.

Quantified Self/Public Health Symposium

Gary Wolf and Ernesto Ramirez¹

Contents

3	Introduction
4	What Is Self-Collected Data?
5	Who Self-Tracks
6	Example Projects
6	Mapping Activity in Winston-Salem, NC
7	Oregon Dept. of Transportation & Strava
7	The Quantified Diet
8	100K Wellness Project
8	Health eHeart Study
9	New Research Frontiers
9	Toolmaker-Driven Public Health Research
10	Public/Private Partnerships
11	Validation
11	Effectiveness
12	Discovery
14	Guidelines
14	Data Flows
14	Aggregation
16	Export
16	OAuth

Continued [↪](#)

¹ Gary Wolf and Ernesto Ramirez are the founder and program director, respectively, of QS Labs, a social enterprise whose mission is to inspire meaningful discoveries about ourselves and our communities, grounded in accurate observation, and enlivened by a spirit of friendship.

17	Participation Reconsidered
17	Personal Access as a Research Tool
18	A Permissions Model for Access
18	Reverse Data Flows
20	Participatory Research Designs
21	Research Technique
22	Access Prospects
23	Appendix A
23	References
24	Videos
24	Introduction and Opening Remarks
25	Data Flows
25	Deep Personal Data
25	Participation (1)
25	Participation (2)
26	Closing Remarks
26	Reactions
27	Attendees
31	Conveners
31	Toolmakers
39	Researchers
47	Quantified Self Labs
48	Appendix B
48	Quantified Self APIs and Data Flows
48	Introduction
48	Our Point of View
49	How to Use the Diagrams
50	Highlights
53	Diagrams

Introduction

Data we collect in the normal course of life holds clues about important questions of public health. Self-collected data can give us evidence about the effects of today's practices. It's also a source of new ideas. But many barriers stand in the way of using our data for personal and public benefit. These barriers include: poorly documented data formats; records isolated in data silos; lack of best practices for sharing data; and unanswered questions about the personal, social, and business risks.

This report has its origin in a day-long symposium organized by QS Labs in April 2014 with support from the Robert Wood Johnson Foundation. At this meeting, about one hundred self-trackers, toolmakers, public health researchers, and science funders participated. Our topic was "improving access to self-collected data." In this report we've attempted to summarize the most important lessons we took from the many contributions to the symposium, whether these contributions were in the form of public talks, background interviews, or follow-up discussions. Other attendees will doubtless have taken different lessons, and a link to additional materials and comment is included in Appendix A.

Three special instances of good fortune made this meeting possible. In 2013 we fielded a call from Bryan Sivak, the CTO of the U.S. Department of Health and Human Services, who asked if we could help connect public health researchers into the Quantified Self community. His bold support allowed us to convene a group that went far beyond our normal network. Meanwhile, at Calit2 at the University of California, San Diego, Dr. Kevin Patrick's group was completing their "Health Data Exploration" project, a survey of attitudes toward data sharing among users, researchers, and makers of self-tracking tools. Dr. Patrick and Calit2 Director Larry Smarr agreed to host our meeting and contribute their deep research knowledge. And, crucially, the Robert Wood Johnson Foundation, where many of the most important threads in public health research intersect, agreed to fund the meeting. On April 3, 2014, nearly a year of preparation culminated in a day-long meeting for scientists and researchers, government leaders, application developers, device makers, and self-trackers. A list of attendees, video documentation, citations, and references can be found in Appendix A.

We had both a modest and an immodest hope for our meeting. The modest hope was that we would expose to common view some of the most important barriers to research, highlight some novel collaborations, and, by producing a list of hard problems with examples, dispel some of the fog of abstraction that makes failures in this area appear mysterious. We didn't suppose we would find many answers. We are still at an early stage, when simply naming problems, and agreeing that they are problems, has value.

Our immodest hope lies in the supposition that our work will have additional stages after the first. After all, science has always used data. Medicine and social science have always used data coming from people. Our meeting was about using "self-collected" data. Can making this one switch, changing out this one circuit, amount to the inauguration of a new research program in the human sciences?

We hope it does.

What Is Self-Collected Data?

Some examples of self-collected data, in the way we mean it, are: step and activity data from popular devices like the Fitbit, Jawbone UP, and the Nike+ system; glucometer and CGM (continuous glucose monitor) data used by diabetics to manage their own care; sleep data from the many sleep tracking apps on mobile phones; lifelogs; food diaries; and the many other types of data originating from the dozens of other self-tracking tools now becoming popular.

Self-trackers care about this data because it is personally relevant, easy to gather, and appears to be more or less private and under their own control. The Quantified Self video archive contains documentation of hundreds of projects self-trackers have carried out using their own data,² typically in service to personal goals related to health and well-being. Public health researchers are interested in the data because it appears to be scientifically relevant, highly detailed, machine readable, and well structured. While these sets of interests overlap, they aren't identical and sometimes actively conflict. Moreover, a close look at self-collected data shows highly complex data flows that don't match the expectations of either self-trackers or researchers.

Before turning to the complexities of self-collected data, however, we should note that not all data based on individual reporting counts as "self-collected," as we understand it. Consider a public health researcher taking advantage of the aggregate information contained in U.S. tax returns to plot health disparities in the United States. Many of these returns are painstakingly filled out by individuals using private knowledge of their own financial data. However, engagement in the research process by the people whose data is aggregated is too slight to notice; as far as a researcher is concerned the work done by individuals could be done by a machine or a disinterested 3rd party, and often is. The special issues of self-collected data arise when the data is gathered in the first place by individuals for their own purposes.

If aggregate tax data doesn't count, what about data collected in population health research with the help of study participants who may answer surveys, speak with interviewers, sign forms permitting the release of their medical records, or consent to use a specialized research device for bio-monitoring? We exclude most of these cases too. In putting together our program, we realized that this distinction between self-collected and researcher-collected data applies even if the "specialized research device" is a commercially available activity tracker supplied by the researcher's lab. In other words, it is not technical instrumentation that governs the distinction, but human intentions.

Who Self-Tracks

Self-collected data can be important for public health only if self-tracking practices are common. The best data about the prevalence of self-tracking comes from the Pew Research Center's detailed "Tracking For Health" report,³ published in early 2013. Susannah Fox, an associate director at Pew Research Center's Internet & American Life Project, and co-author of the report, set the context for our work with unambiguous evidence that self-tracking is widespread. Some of the most important findings from that study include:

- 60% of U.S. adults say they track their weight, diet, or exercise routine.

² <https://vimeo.com/groups/quantifiedself>

³ <http://www.pewinternet.org/2013/01/28/tracking-for-health/>

- 33% of U.S. adults track health indicators or symptoms, like blood pressure, blood sugar, headaches, or sleep patterns.
- 12% of U.S. adults track a health indicator on behalf of someone they care for.

Most people do not use digital technology to collect any formal data about themselves. Many say they are tracking “in their heads,” and among those making a record of their observations the most common tool is pencil and paper.

- 49% of health trackers just keep track in their heads.
- 34% use pencil and paper, like a notebook or journal.
- 8% use a medical device such as a glucometer.
- 7% use an app or other tool on their mobile device or phone.
- 5% use a computer program, like a spreadsheet.
- 1% use a website or online tool.
- (More than one answer was allowed; total exceeds 100%)

More recent Pew data suggests that both the prevalence of tracking and the use of technology will grow. According to Pew, 18- to 29-year-olds are the most likely to use an app or mobile tool for self-tracking. People managing a chronic disease (45% of U.S. adults) or caring for a loved one with significant health issues (39% of U.S. adults) are especially likely to self-track. The most likely age group to self-track are people 65+, but many of these are “pencil and paper” trackers. Today’s 65+ age group reached adulthood before the advent of personal computing, but within the next two decades all people over 65 will have used computers routinely at school and work.

Along with the Pew data about self-tracking for health, there is good evidence of the spread of self-tracking with technology for other reasons, and much of this additional self-collected data may be useful to public health researchers. Last year a leading technology news site reported that app downloads in just one self-tracking category, “running apps on mobile phones,” exceed 50 million to date.⁴ Other market research supports this prediction.^{5,6}

Interest among public health researchers in self-collected data is premised on the increasingly common use of digital self-tracking tools. This premise is well founded. However, patterns in the adoption of self-tracking technologies will undermine naive generalizations about population health from self-collected data. Inequality is only the most obvious factor influencing the use of self-tracking tools; critical opposition to practices that can be seen as involving medical surveillance may also play a role.⁷ While the Pew data provides a reassuring foundation for interest in using self-collected data for public health benefit, research into the individual and cultural context of self-collected data, including the conditions that lead to a lack of data, remains much needed.

⁴ <http://mashable.com/2013/05/13/runtastic/>

⁵ <http://mobihealthnews.com/29883/prediction-17m-wristworn-activity-trackers-to-ship-in-2014/>

⁶ <http://mobihealthnews.com/28825/fitbit-jawbone-nike-had-97-percent-of-fitness-tracker-retail-sales-in-2013/>

⁷ See, for instance: Mesmerizing Maps of Where People Jog Reveal Something Odd: A City’s Rich and Poor Neighborhoods [Link]. For a description of attitudes toward participation in medical research among African American males, see Byrd, G. S., Edwards, C. L., Kelkar, V. A., et al. (2011). Recruiting intergenerational African American males for biomedical research Studies: a major research challenge. *Journal of the National Medical Association*, 103(6), 480–7. [Link]

Example Projects

Many self-trackers collect data about their weight, diet, metabolism, mood, location, and symptoms of acute or chronic disease. The intuition that self-collected data can be used in ways that would yield immense public benefit is based on the observation that it would be much easier to collect data on important health topics like obesity, allergies, asthma, metabolic disorders, addiction, and mental health if the people tracking them for their own purposes would share their self-observations with researchers. When researchers think about critical public health instruments like the National Health and Nutrition Examination Survey (NHANES), which yields foundational evidence supporting programs addressing lead poisoning in children, folic acid deficiency, obesity, and healthcare access - among many others - they naturally wonder about future benefits for a public that could see itself in many more dimensions, for far less cost. Use of such data for public health is not yet common, but collaborations do exist. We haven't yet attempted a general survey, but a few examples of research collaborations using self-collected data point to an interesting divergence of approaches.

Mapping Activity in Winston-Salem, NC

Researchers at the University of Michigan and Harvard University, among others, and data scientists at activity tracking company MapMyFitness collaborated on a research project to showcase the use of self-collected fitness data for public health research. Specifically they generated an example use-case exploring the amount and type of physical activity that occurs in public parks in Winston-Salem, NC. By partnering with MapMyFitness researchers were able to obtain data on nearly 3,100 individuals who tracked more than 43,000 unique workouts (running, cycling, walking, and hiking). The interest in their examples for our purposes lies not in any surprising conclusion but in the ability of the researchers to gather large-scale data from individual activity sessions. As the study's authors wrote in the paper they published in *Frontiers in Public Health*:

“The enormous scale of these data creates the potential to explore questions about physical activity in many more individuals, at a much more detailed level than in previous studies.”⁸

In his symposium talk, Kyler Eastman, research and development engineer at MapMyFitness, and one of the paper's authors, expressed his ongoing willingness to work with researchers using MapMyFitness data.

“I'm interested in figuring out how to partner with researchers to get this data into their hands while at the same time preserving our users' sense of trust and privacy. We want to partner with people who are devoting themselves to figuring out the best insights.”

⁸ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3949289/>

The purpose of this collaboration was to demonstrate the capacity of self-collected data to contribute to knowledge of population health over larger geographic and temporal scales than what is typically accessible. However, despite the large data sets, the researchers faced significant limitations in meaningfully interpreting the MapMyFitness data, including difficulty in determining the extent to which the data was representative and questions about the reliability of user-entered data mingled with automatically collected data.

Oregon Dept. of Transportation & Strava

Public agencies use public health data for planning and administration, and the intersection of research, civic, and personal data is an emerging area for cooperation. For instance, the Oregon Department of Transportation (Oregon DOT) recently paid Strava, a popular GPS-based bike and running tracker, \$20,000 for a one-year license to use data on more than 17,000 riders. The license agreement allows the Oregon DOT to access data about cyclists' routes and specific GPS traces of cyclists while exercising and commuting.⁹

Strava sold the Oregon DOT a license to use de-identified data as part of a new business strategy they are calling the Strava Metro platform.

“Strava Metro’s mission is to produce state-of-the-art spatial data products and services to make cycling, running and walking in cities better. Using Strava Metro, departments of transportation and city planners, as well as advocacy groups and corporations, can make informed and effective decisions when planning, maintaining, and upgrading cycling and pedestrian corridors.”¹⁰

In providing de-identified data, Strava, like MapMyFitness, attempts to protect privacy. As a result, correlating activity data collected with Strava with other personal data is difficult. Interpretive limitations of the type faced by researchers using the MapMyFitness data are mitigated here, and on the Strava Metro platform generally, by an emphasis on questions of civic management, such as traffic, that appear addressable without deep contextual knowledge. However, where de-contextualized self-collected data is used for urban planning, unexamined social biases in self-tracking patterns are a problem.¹¹

The Quantified Diet

Recently, the makers of Lift, a habit and behavior self-tracking application, embarked on an ambitious project to measure the effectiveness of many different popular diet protocols. Lift recruited current and new users to participate, and assigned each participant to one of nine different diets, one of two control programs, or a placebo program. Assignment was random after participants identified personally acceptable

⁹ <http://bikeportland.org/2014/05/01/odot-embarks-on-big-data-project-with-purchase-of-strava-dataset-105375>

¹⁰ <http://metro.strava.com/>

¹¹ See: Confronting the Data Dilemma [Link]

interventions. Participants were asked to self-report their weight, mood, and energy levels. Currently, Lift has data on over 9,000 participants in the “Quantified Diet” experiment, with all the data, experimental methods, and preliminary results freely available as a repository on GitHub.¹² In conversations with us, Tony Stubblebine, CEO of Lift, expressed enthusiastic interest in conducting collaborative experiments with public health researchers. The Quantified Diet research is strikingly different from the two projects described above. Highly active participants contribute detailed multidimensional data because of a keen personal interest in the results of the research. Context and engagement is given priority over formal rigor, and all data is shared with the public, using GitHub as the publication channel.

100K Wellness Project

Leroy Hood, the founder of the Institute for Systems Biology (ISB), has recently announced a new research venture named the 100K Wellness project.¹³ This project will be a longitudinal study of 100,000 individuals who will participate in new data collection methods as well as contribute their own self-collected data. The project’s vision is to combine genomic, environmental, lifestyle, and medical biometric data for a more complete picture of human life over time, with an emphasis on the transition from wellness to disease. As Dr. Hood explains:

“There are no metrics for wellness, but we can make quantitative measurements of wellness. Currently we study disease only after we’re sick. We don’t study the transition from wellness to disease. We don’t study longitudinal progression of disease. With our project we can start to build those models. We are going to build systems to give participants access to the critical information that they can use to help change their life.”

Dr. Hood envisions a scientific model of human wellness and disease that forms the basis for individuals learning. Attempts to integrate detailed self-collected data in the context of large-scale scientific studies like the 100K Wellness Project attempt to work on both side of the distinction between official prescription and individual self-efficacy.

Health eHeart Study

Matching Dr. Hood, Dr. Jeffrey Olgin and his colleagues at the University of California, San Francisco, have announced the Health eHeart Study, a massive program to understand heart disease with the help of one million research participants.¹⁴ Enrolled through an online process, participants are asked to commit to a ten-year study that will include diverse data gathering procedures using self-tracking tools such as the sleep and heart rate monitoring apps from Azumio, Inc., and the “iPhone ECG” device made by AliveCor. Since

¹² <https://github.com/liftcode/lift-research/tree/master/quantified-diet>

¹³ <http://research.systemsbiology.net/100k/>

¹⁴ <http://www.health-eheartstudy.org/>

the range of self-tracking tools is expanding rapidly, the study is designed to be agnostic about what type of tools participants use. As the recruitment materials state:

“Collect data at home. If you want, you can use your own scale, blood pressure machine, and more to collect measurements and send them to us using our secure system. We might even mail you a ‘spit kit’ to collect your DNA.”

The vision of the Health eHeart Study is data-centric, in the sense that the stated emphasis is almost entirely on recruitment and data collection, management, and analysis systems. With commitments in place from the Patient-Centered Outcomes Research Institute (PCORI), a Congressionally established funding source for innovative research involving patients, and with an announced intention to recruit more individuals than have ever before been part of a voluntary health research program, the Health eHeart Study is destined to be proving ground for new ideas related to the barriers to research described below.

New Research Frontiers

Already, research involving self-collected data is going beyond the accelerometers and mobile phone sensors most commonly associated with Quantified Self practices. New tools available to individuals who are curious about themselves include brain monitoring devices, sleep tracking devices, medical-grade ECG monitors, at-home blood and metabolic testing, environmental sensors, and systems to track nutrient components in food. As the potential sources of self-collected data multiply, it’s impossible not to notice a tension between a desire by researchers to have access to “all the data” so that they can address canonical topics in public health and the unbounded creativity of individuals who insist on exploring individualistic and idiosyncratic questions using novel types and combinations of data. Projects that have the advantage of being highly legible from an administrative perspective have the disadvantage of low meaning for participants, while projects that attempt to gather rich contextual data, such as the 100K Wellness Project and the Health eHeart Study, by contrast, confront a rapidly changing landscape of tools and practices that makes the establishment of standard protocols difficult. In this regard, the situation of researchers attempting to use self-collected data on a large scale resembles that of technologists whose hardware and software can quickly be rendered obsolete - even prior to release - by changes in context.

Toolmaker-Driven Public Health Research

Makers of self-tracking tools are today’s de facto stewards of self-collected data. Many engage in research that resembles work traditionally done by government researchers and university faculty. This is part of a broader pattern of privatized research in science. New business models have made it commercially attractive to sell tools and services based on personal data for improving health outcomes in an employee or managed care population. There’s high confidence in the healthcare industry, and particularly among investors, that access to detailed self-collected data can be used to decrease—or at least slow the increase—in medical costs. Many companies that make self-tracking tools sell them in bulk to corporations, along with data analysis services to support employee health. For these companies, outcomes research linked to lifestyle and fitness improvements is worth money. Companies primarily focused on serving individual users also may have

an interest in the kind of statistical analysis characteristic of population health research, as they search for evidence of the benefits of their tools, and ideas for improving them.

A realistic approach to the question of access to self-collected data for personal and public benefit requires recognizing the stewardship role of private companies in aggregating health-related data. Private companies control most personal data flows. Many of these companies do research on aggregate user data that seeks to understand classic public health questions, such as the factors influencing obesity, depression, allergies, and other conditions, and the relative effectiveness of interventions. Much of this work is invisible to the traditional public health research community.

Commercial stewardship creates particular access challenges. From a self-tracker's perspective, access to our data is insecure when it is controlled by commercial stewards with conflicting interests whose corporate lifespan may be brief. For example, on the Quantified Self Forum, one of the most active threads is a discussion of how to retrieve sleep data from the Zeo sleep monitoring system,¹⁵ a retail product whose maker went out of business more than a year ago, shuttering its web site, closing access to its data archive, and selling all its user data to a buyer whose identity has never been disclosed.

The challenges of serving the public interest under conditions of private stewardship are compounded by the fact that research universities, traditionally the guardians of public scientific knowledge, are now commercial purveyors of intellectual property through their technology transfer offices. The role of the research university as a commercial competitor makes it difficult even for a highly motivated private researcher to share aggregate data.

Public/Private Partnerships

David Van Sickle is the founder of Propeller Health (formerly Asthmapolis),¹⁶ whose GPS-enabled asthma inhalers feed data into a system for understanding asthma triggers and reducing asthma incidents. Propeller Health research projects are already generating knowledge related to public health; for instance, by measuring and mapping the ratio of at-home/away-from-home incidents, which may lead to a re-evaluation of the relative contribution of pollution and polluters to asthma prevalence. Van Sickle is an experienced public health researcher, whose dissertation research examined the rising prevalence of asthma and allergy in India, a topic he previously studied among Native Americans in Alaska, Arizona, and New Mexico. He is motivated to share data with public health researchers, but in practice this is difficult. “We got an inquiry from a group at University of Wisconsin that wants to use data we collect,” Van Sickle recounted in one of our interviews. “When we get to the actual contracting with the university, their IP/TTO [Intellectual Property/Technology Transfer Office] compels total confrontation with our legal team and our business model. They have odd contractual and data sharing issues that I have to pay someone \$500/hr just to understand. Then I have to think about the possibility I’m going to have to go to my board and say I messed up and licensed a valuable business asset to a random university for no dollars.”

It's important to note that “licensed,” in Van Sickle's scenario, does not simply mean “allowed to use.” It means “allowed to commercialize.” In other words, the university is perceived by Van Sickle as a potential

¹⁵ <https://forum.quantifiedself.com/thread-zeo-shutting-down-export-your-data>

¹⁶ <http://propellerhealth.com/>

force of privatization, taking knowledge that he wants to share with the public and placing it back into the stream of commerce to his disadvantage.

Interestingly, researchers and toolmakers at the symposium mostly agreed that, given time and attention, it is possible to design fair collaboration agreements allowing research access to data under the control of toolmakers. In fact, Propeller Health did share data with the University of Wisconsin, thanks to the intervention of Dr. Patty Brennan, professor of nursing and industrial engineering at the university and a leading thinker about the use of self-collected data for public benefit. The key to success in this case was personal authority. As Van Sickle put it: “Patty went off and used her muscle to get a much more company-friendly data sharing agreement.” The bespoke nature of the solution indicates the complexity of a research environment in which aggregate data about population health is both under the control of commercial stewards and valued as a commercial asset by managers of research institutions.¹⁷

With private companies focused on understanding population health and making money from evidence of improved outcomes, it’s not surprising that some collaborations related to public health are initiated not by academic or government researchers, but by the toolmakers themselves. These research projects are motivated by toolmaker interest in validating their instruments, proving their effectiveness, and discovering new benefits. Toolmakers focused on activity and nutrition are also interested in contributing to realistic national guidelines, as these guidelines influence user goals.

Validation

Data sensing companies are concerned with reflecting the reality of human behaviors back to their users. This requires efforts to make sure that those reflections are accurate. For instance, research into the nature of physical activity is ongoing at companies who use proprietary algorithms for interpreting accelerometer output as steps, METs, or calories. While toolmakers use internal research to test their products, they also seek research collaborators to demonstrate accuracy and, at times, to discover new potential uses of their tools. For instance, Max Utter, vice president of wellness product research at Jawbone, is currently collaborating on validation studies with Matthew Buman and Eric Heckler at Arizona State University aimed at demonstrating the accuracy and reliability of the Jawbone UP device. Academic researchers at Maastricht University in the Netherlands collaborated with researchers at Philips to validate the Philips DirectLife activity monitor, and published their results.¹⁸ Such validation studies are probably the most common form of collaboration between researchers and toolmakers.

Effectiveness

As toolmakers increasingly market their ability to effect health outcomes, they are also interested in collaborating with research partners in order to understand how effective their tools and techniques are.

¹⁷ See Appendix A for an example data sharing agreement from Propeller Health and the University of Wisconsin.

¹⁸ Polzien, K. M., Jakicic, J. M., Tate, D. F., & Otto, A. D. (2007). The efficacy of a technology-based system in a short-term behavioral weight loss intervention. *Obesity* (Silver Spring, Md.), 15(4), 825–30. [Link]

For example, a study published in 2012 in *Obesity* compared a standard in-person weight loss intervention to a technology-based program that included the BodyMedia Fit system.¹⁹

The authors found that the addition of the device, with and without in-person coaching sessions, was related to statistically significant reductions in weight, stating that “the commercially available technology used in this study appears to be effective for weight loss, reducing energy intake, and increasing physical activity in the absence of in-person intervention contact.” Well-executed research to determine whether or not the technology is likely to work as promised benefits self-trackers greatly, as there will always be many competing ideas about what to try. When results are positive, toolmakers can use the research to increase credibility, while researchers can better understand how to use new tools to improve health outcomes. Effectiveness research dominated by toolmaker collaborations has risks, such as corruption of research ethics, as infamously occurs in the pharmaceutical industry.²⁰ However, compared to the pharmaceutical industry, with its approximately \$500 billion in global revenue in 2013, the stakes of effectiveness research using self-collected data are still very small, which makes this a good time for frank discussion of the practices we’d like to see evolve.

Better effectiveness research could lead to public health breakthroughs. We’ve been intrigued to watch the progress of the partnership, launched in 2013, between Philips and Weight Watchers, which led to the integration of the Philips DirectLife activity tracker into the Weight Watchers program, under the name ActiveLink. Weight Watchers is the most popular commercial weight loss program in the world. Its effectiveness is controversial. A systematic review of commercial weight loss programs, published in 2005, found evidence of slight but significant long-term effectiveness, with the largest of the studies analyzed in the meta-analysis reporting a loss of 3.2% of initial weight at 2 years.²¹ The largest randomized controlled trial involved 423 participants. There have now been more than 400,000 ActiveLink devices put into circulation by Weight Watchers, and participants with ActiveLink are collecting activity data in the context of tracking their weight and diet. Given the intense public health attention on weight loss, and Weight Watchers’ role as a leading program combining dieting, exercise, self-tracking, and social support, accurate and detailed knowledge of its effectiveness would be valuable.

Discovery

Toolmakers may also benefit from learning about new ways to use their tools and data to support meaningful change in the lives of their users. In the last few years there has been a steady growth of research projects, conferences, and workshops dedicated to the areas of personal informatics, Quantified Self, and self-collected data. Preliminary talks and publications from a handful of scholars in design, human-computer interaction, sociology, and anthropology investigating the Quantified Self movement have just begun to

¹⁹ Pellegrini, C. A., Verba, S. D., Otto, A. D., Helsel, D. L., Davis, K. K., & Jakicic, J. M. (2012). The comparison of a technology-based system and an in-person behavioral weight loss intervention. *Obesity* (Silver Spring, Md.), 20(2), 356–63. [Link]

²⁰ See “Institutional Corruption and the Pharmaceutical Policy.” Special Issue of *The Journal of Law, Medicine & Ethics* 41.3 (2013) [Link]

²¹ Tsai, Adam Gildea, and Thomas A. Wadden. “Systematic review: an evaluation of major commercial weight loss programs in the United States.” *Annals of internal medicine* 142.1 (2005): 56–66. [Link]

appear.^{22,23,24} While these approaches may seem distant from many of the concerns of public health research, there is in fact significant overlap, especially in the area of community health.

The research of Katherine Kim of the Health Equity Institute at the University of San Francisco focuses on how self-collected data can help young people living with obesity and depression, and on enhancing tribal health and food security in the Klamath Basin of Oregon. The patterns of use and meaning that she sees in her work are normally hard for toolmakers to notice. “The usage and adoption model for the underserved is important. Start that conversation, and it might open up different revenue streams,” said Kim during a pre-symposium interview. “What I really want to know is how we can impact individual and public health. How can we take the bits and put them into products that can help the public? Someone has to build real services that are sustained after the grant is over.”

In a fascinating talk given at the symposium, Mathew Kay of the University of Washington demonstrated that even a biometric as simple and seemingly well known as the weight reading delivered by a bathroom scale is not well understood. Kay wanted to know what people thought they were seeing when they looked at a scale, and how this connected, or didn’t connect, to the limitations and uncertainty of measuring weight. Kay became curious about human intraday weight fluctuations, and was surprised that the topic had never been studied. Experts were interested in trends over time, not more unpredictable single measurements. “But the way the data looks to the individuals is different than what the health professionals care about.” As Kay wrote in his paper:

“[W]e found that consumers’ perception of weight scale behavior is often disconnected from scales’ capabilities and from clinical relevance, and that accurate understanding of weight fluctuation is associated with greater trust in the scale itself. We propose significant changes to how weight data should be presented and discuss broader implications for the design of other ubiquitous health sensing devices.”²⁵

Health research that explores the experience of self-tracking in individual and community contexts exposes social and cognitive factors that influence learning and can function as product research for toolmakers.

²² Lupton, D. (2013). Understanding the Human Machine [Commentary]. *IEEE Technology and Society Magazine*, 32(4), 25–30. [Link]

²³ Nafus, D. & Sherman, J. (2014). This One Does Not Go Up to Eleven: The Quantified Self Movement as an Alternative Big Data Practice. *International Journal of Communication*. [Link]

²⁴ <http://quantifiedselfresearch.org/>

²⁵ Kay, M., Morris, D., Schraefel, M., & Kientz, J. A. (2013). There’s no such thing as gaining a pound. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp ’13* (p. 401). New York, New York, USA: ACM Press. [Link]

Guidelines

Expert guidelines matter to users. Current federal physical activity guidelines suggest that adults should participate in 150 minutes of moderate-to-vigorous aerobic physical activity per week (or 75 minutes of vigorous activity per week). These guidelines are common references in medical and public health literature as well as markers for researchers as they study and evaluate activity patterns. Common sense suggests that if minimum or optimum physical activity levels are proven to exist in a meaningful way for individual users, these levels will vary greatly across the population. Since the fitness programs packaged with commercial tracking tools are often pegged to user goals, toolmakers are curious about the standard physical activity guidelines and would like to see them improved. Margaret McKenna, head of data & analytics at RunKeeper, told us that current activity guidelines do not reflect the cyclic nature of activity and exercise they observe in their users and this can lead to inaccurate and unrealistic benchmarks used by medical professionals and researchers. “We know that our most engaged users can lapse up to 6-8 weeks at a time,” McKenna said. She continued:

“The idea that someone can work out 150 minutes per week, 52 weeks a year, is just unrealistic. We’re wondering if there is a way we can feedback our data into some of the processes that go into making these guidelines so that they’re more realistic for people when they’re trying to achieve them.”

Data Flows

Perhaps no phrase is more useful, in a negative way, in thinking about the intersection of self-tracking practices and public health research than Clive Humby’s famous statement: “Data is the new oil.” The utility of the cliché lies in its neat expression of the fantasy that data is the kind of thing that exists out there already in the world, a pool of valuable raw material awaiting discovery and refinement into a fungible source of power.

Self-collected data originates with individuals. Much of it is recorded by hand. Even where recording occurs automatically, the nature of the observations is closely shaped by the interests and intentions of people making them, often in ways that are hard to understand at a distance. Technology can increase rather than decrease dependence on contextual knowledge, because quantitative data is highly abstract, and doesn’t come packaged with complete rules governing where it is relevant and how to apply it.

Aggregation

A common strategy for lack of contextual knowledge about self-collected data is to collect data from many users and look for general patterns. For many researchers, access means “access to aggregate data.” In early 2014 we created a set of diagrams of data flows from a sample of self-tracking tools and systems. These sample diagrams and a more detailed analysis are included in Appendix B. The goal of the diagrams was to understand how self-collected data looks as it moves away from its original context, and to find out how

self-trackers and public health researchers can access data, in the simplest sense of “acquire a copy.” We were interested both in how data could be accessed manually by self-trackers, and how developers working with public health researchers could access data via APIs. Readers interested in the details of this research can look at the appendix. Here we’ll give just one example of how data flows through self-tracking APIs, and of the way the ambiguity of self-collected data can increase as the data travels away from its source.

RunKeeper is among the most popular fitness tracking apps, and the company has devoted unusually expert resources to facilitating data flows for users and partners. RunKeeper has a well-maintained API that serves data aggregated from other self-tracking systems, called the Health Graph API. A glance at the RunKeeper diagram [page 53] shows that there are many devices and services that can flow data into the RunKeeper Health Graph and therefore will show up in data accessed via the API. However, as data flows through the Health Graph system, important elements about data provenance can be lost, such as information about what tool or system was used to gather a measurement.

This is not a special problem for RunKeeper. The last few years have seen the rise of many aggregation pipelines that impose order on heterogeneous data. Validic, Human API, and Fitabase all manage API access for academic and corporate clients. Many of their tools for self-tracking include features that help their customers store and view data collected from other sources. The recently announced HealthKit from Apple promises to integrate data flows from multiple self-tracking services and give its customers the ability to explore their own data using a variety of apps.²⁶ Many of these systems succeed in easing the burden of managing data flows for individuals and institutions. Researchers appreciate getting data that is more tractable. Self-trackers like being able to view diverse data in a clear and logical way. But as data travels and is combined into aggregates, the types of queries it can help answer change.

Some examples of what this means in concrete terms: Today most data collected by self-trackers includes neither geolocation data nor time stamps that accurately capture what time it was for the participant. Important questions on topics like medication, diet, exercise, or metabolism can’t be answered without accurate times. Most self-tracking systems do not publicly log the effects of changes in firmware that may change the measurement outcome. Activity trackers typically cannot distinguish between periods of inactivity and periods where the device was not worn. Methods of logging changes to data by self-trackers—who may add an entry at a later date, or revise a previous entry—are diverse and often undocumented. Individuals can often adjust and analyze their records based on personal knowledge, but in aggregate such adjustments are impossible.

Anne Wright, the co-principal investigator and director of operations for the BodyTrack project in the CREATE Lab at Carnegie Mellon University, has extensive experience troubleshooting aggregation issues. BodyTrack is a system for integrating diverse data types into a timeline that supports reflection and problem solving by individuals.²⁷ “For any arena where your activities generate data that is potentially valuable for self-reflection you should be able to direct it to conveniently flow to destinations of your choice where you can meaningfully reflect on it,” Wright said. But, she continued:

²⁶ <https://www.apple.com/ios/ios8/health/>

²⁷ BodyTrack has combined with the open-source Fluxstream project: <https://fluxstream.org/>

“...there are a lot of assumptions about aggregated data. Everyone assumes that’s the easy thing, the low-hanging fruit we can pluck and then knowledge will be ours. It’s only really useful [with] context. If you have people living their lives and you don’t have their context then taking that data and clubbing it into a story is fairly meaningless.”

Nobody who works with data can be very surprised by these challenges. Expert researchers and the most skilled self-trackers handle data challenges in a similar way. They do not refine their data into a fungible, universal source of power. Instead, they adjust their questions, rework their data, and make new observations in a back-and-forth craft of knowledge making. With this goal in mind, we’d like to call attention to two practical ways to support access.

Export

Many toolmakers already provide a way to download a complete record of our data into a file under our own control. Some toolmakers place the export feature behind a paywall, and still more provide no export at all. In an academic public health context, the situation is worse: Few researchers offer participants access to their own data. Data export should be standard. Export gives self-trackers with time and skills the chance to annotate or augment data in a way that is unavailable in the original data collection system; it allows self-trackers to collaborate with people who have more skills in data visualization; it permits data to be uploaded into apps and web pages offering useful analytic capacities; and, most important for this discussion, export allows self-trackers to easily share their data with researchers. This is not today’s reality. There is still no “Blue Button” for self-collected data.²⁸

OAuth

Export, while fundamental, limits access to discrete, manually triggered data transfers. One of the main reasons self-trackers want access to their data is so they can take advantage of visualization, analytics, and other services offered by third parties. The standard way self-trackers integrate their data into third-party services is via OAuth, a widely used protocol that allows us to authorize our data to travel from one service (for instance, Fitbit), to another (for instance, RunKeeper), without exposing usernames and passwords. OAuth keeps the individual self-tracker at the center of data flows by establishing a border crossing where permission is asked for and granted. Nearly everybody reading this report will have seen an OAuth request; in fact, it is now so widespread that there’s a risk of taking it for granted. We mention it here because OAuth embodies the principle of *rescindable individual permission for ongoing access*, which in combination with universal ability to export our data could power a data-sharing architecture founded on individual choice.

²⁸ <http://www.healthit.gov/patients-families/blue-button/about-blue-button>

Participation Reconsidered

The phrase “research subject” has been replaced, in the public health lexicon, by “participant.” If yesterday’s research subjects were production factors in a scientist’s workshop and today’s participants are meant to be fully informed volunteers with interests worthy of protection, then the spread of self-tracking tools and practices opens the possibility of a new type of relationship between public health researchers and individual self-trackers, in which participants bring valuable craft knowledge along with their data.

Personal Access as a Research Tool

Doug Kanter’s work on his blood glucose data exemplifies skillful practice by a self-tracker. Using more than 90,000 individual blood glucose measurements, along with other manually and automatically gathered personal data, Kanter produced an image of a year-long project to improve his health and gain a better understanding of the relationship between his metabolism and various life events, including his pattern of physical exercise.²⁹ As a type 1 diabetic, Kanter began his education in self-tracking as a patient following conventional protocols, but his discovery of personal benefit has led him—as it has led many diabetics—beyond standard approaches.

“I get to see my endocrinologist every three months,” Kanter told us. “When she sees I’m under control, she says to just keep doing what I’m doing. But I want to do better.” Kanter’s A1c level improved significantly during the year he was working on this project. Within the overall pattern of improved health are interesting personal discoveries, including both the dramatic effect of exercise on his insulin dose and the upper bound, at about 30 miles of running per week, where the benefits of increased exercise cease. Although Kanter’s motive in going beyond standard practice is not to make an original research discovery, but to improve his health, his methods and results border on research questions of broad interest.

Diabetics and people caring for family members with diabetes have been forced to develop advanced techniques for learning from their own data because their measurements influence the consequential decisions they make many times a day about diet, activity, and insulin dosing. Some of these techniques run far in advance of methods commonly used in the academic research community, both in terms of understanding their own data and providing tools for supporting discoveries relevant to public health.

For instance, Lane Desborough, Jason Calabrese, and other contributors to the open-source Nightscout project have learned how to collect data from Dexcom’s proprietary continuous glucose monitor; store it in a well-structured, open format; and wirelessly broadcast it to a mobile device so that it can be remotely viewed.³⁰ Desborough and Calabrese are parents whose children with diabetes would like to participate freely in activities like riding bikes and having sleepovers without having to stop frequently for an adult to check their glucose readings. With Nightscout, Desborough and Calabrese can keep an eye on things remotely, even streaming the glucose data to a Pebble smart watch. “Until I started doing this, I hadn’t had a real night’s sleep in four years,” said Calabrese recently, talking to an activist meeting organized by DiabetesMine.³¹

³⁰ <http://nightscout.github.io/> ; <http://www.nightscout.info/>

³¹ <http://www.diabetesmine.com/2014/06/the-nightscout-revolution-cgm-in-the-cloud.html>

There are already more than 1,200 people in the Nightscout Facebook group,³² and between 50 and 100 new people are joining daily. Meanwhile, the Tidepool Project, a nonprofit devoted to reducing the burden of type 1 diabetes, founded by Howard Look, himself a parent of a child with diabetes, is building an open infrastructure for personal access and research contribution of diabetes data, including data passing out of Dexcom continuous glucose monitors via Nightscout. Importantly, public health researchers attempting to use data from Nightscout will not find it in official clinical use, as it isn't approved by the FDA and inhabits a regulatory gray zone. Collaborations using data gathered via Nightscout are more likely to come through direct support of diabetics (or parents of diabetics) like Desborough and Calabrese and activist projects like Tidepool, whose common slogan is “We are not waiting.”

The goals, instruments, and techniques used by Kanter, Desborough, and Calabrese are advanced, but they are categorically similar to those used by many other self-trackers. Services like Tictrac, Addapp, Zenobase, Withings Health Mate, TRAQS.me, Health Graph, Exist.io, and many others, are meant to help individual self-trackers store and view data from multiple connected devices and apps. While the accessibility to self-trackers of their own data from these systems varies, all are simpler than hacking the Dexcom continuous glucose monitor.

The prospect of active intellectual collaboration between public health researchers and communities of self-trackers who have access to their own data opens a productive middle ground between the isolated personal experiments and grandiose aggregation schemes based on the assumption that benefits of scale have no upper limit. Continuously collected data can be shared with researchers on a permissions basis only where self-trackers themselves have good access.

A Permissions Model for Access

Imagine a researcher would like to understand the relationship between geolocation and physical activity. A researcher could set up a portal that would allow participants to authenticate their personal data tracking services for activity (i.e., Fitbit, Jawbone) and location (Moves, OpenPaths). This would allow the researcher to follow the participant over time, gathering and storing data as long as the participant allowed. OpenPaths,³³ a geolocation tracking application for iOS and Android phones, already employs this type of permissions model. It acts as both a service and a data broker, allowing users to access their own geolocation data in a variety of formats, as well as allowing users to authenticate researchers to access their historical and ongoing geolocation data. Additionally, users can deny access to any ongoing data collection at any time.

Reverse Data Flows

If self-trackers can permit researchers to access self-collected data, how about the reverse? Can researchers share study data back with participants?

Jason Bobe is the director of community for the Personal Genome Project (PGP).³⁴ Genomic data is considered a particularly sensitive data type because it can't be easily de-identified and its future significance

³² <https://www.facebook.com/groups/cgminthecloud/>

³³ <http://www.openpaths.cc>

³⁴ <http://www.personalgenomes.org/>

is hard to predict; therefore Bobe and his colleagues have had to come up with new ways of conceptualizing participation and protecting participants from unconsidered risks. For instance, the recruitment process for contributing personal genomic data requires learning and reflection.³⁵

And yet, sharing access to data can appear controversial even in the context of this highly vetted participant community. In his symposium presentation, Bobe spoke of being contacted by Andreas Keller, a leading researcher in the field of olfactory perception. Keller was interested in using genetic data from participants in the Personal Genome Project in concert with an olfactory phenotype assay he was using to profile variations among individuals' senses of smell. Bobe's condition for sharing the data was that participants would receive the results of their phenotyping when the study was completed. However, the Institutional Review Board (IRB) at Rockefeller University, which supervised the study, had never ruled on the ethics of returning data back to the research participants. After careful consideration the IRB at Rockefeller decided to make an institution-wide decision to allow for research studies to give their participants a copy of their research data. However, as Bobe noted, there is no central IRB for university science. Decisions about sharing research data with participants will have to be made one IRB at a time, sometimes with more controversial facts to consider.

The orientation of the research community today is away from sharing data back to participants. For instance, the sample language offered by the National Human Genome Research Institute reads:

“In general, we will not give you any individual results from the study of the samples you give us. This is because it will probably take a long time for this project to produce health-related information that we will know how to interpret accurately. However, we will tell you if we find that you have a communicable disease that we are required by law to report. We will also periodically summarize interesting general findings from this project and how they are contributing to our understanding of health and disease on our project website and through a periodic newsletter.”³⁶

Recognizing that evaluating risk of harm is nontrivial in some cases, we'd also like to call attention to the fact that in many cases this evaluation is indeed trivial. Moreover, self-collected data that originates with individuals seeking knowledge of themselves, such as data gathered from any of the personal data tools mentioned above, has already been integrated into what can fairly be characterized as a personal research process led by the individuals concerned. Public health researchers seeking access to participants' own data and results ought to be oriented toward returning data and study results back to participants where the risk of harm is not evident.

³⁵ <http://www.personalgenomes.org/harvard/howitworks>

³⁶ <http://www.genome.gov/27026589#a1-7>; for a case study that gives a sense of the ethical landscape, see: Returning Individual Research Results: Development of a Cancer Genetics Education and Risk Communication Protocol: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3159194/>

The largest health funding body in the United States, the National Institutes of Health (NIH), has a long history of paying special attention to protecting the privacy and integrity of human subjects' data. The introduction of regulations and rules in the form of Data Safety Monitoring Boards (DSMB) has spread across all institutes of the NIH.³⁷ While we are outsiders to research governance, we'd like to ask if there is a place within the structure of the DSMB guidelines for assuring research participants receive timely access to their research data.

Participatory Research Designs

Active collaboration with self-trackers for the purpose of public health research is changing the role of the researcher and leading to the development of new research designs. Ian Eslick's recent PhD dissertation at the New Media Medicine group at MIT focused on the practical and clinical value of self-tracking and self-experimentation. He is the creator of [PersonalExperiments.org](http://Personalexperiments.org), associated with the Collaborative Chronic Care Network run out of Cincinnati Children's Hospital, and a partner at technology incubator Vital Reactor.³⁸

Eslick's research explores a terrain between anecdote and placebo-controlled trial, where access to data counts as both a source of new ideas and supports fast, relatively easy ways to test those ideas. The results of such tests should not be evaluated using the standards of population science. The goal is not a generalizable theory of cause and effect, but a useful personal estimate. Yes, it is easy to make an error, but this should not be understood as a methodological deficiency in isolation from other factors, such as ease of experimentation, which allows more variation and repetition. "I think we can work in the hinterland between anecdote and double-blind random placebo-controlled trial," Eslick says. "Ninety-five percent confidence intervals are about scientific causal proof, but what I want to know is: Am I making a better decision? We want to value personal significance over statistical significance."

Eslick's interest in using computation to support new forms of knowledge-making for health was inspired by what he learned in tracking and experimenting to manage his own chronic disease. Like other scientifically trained self-trackers managing their own chronic disease or caring for a family member with a chronic disease, Eslick can stand on the border between personal and professional research designs and forge a practical compromise. As both a researcher and a self-tracker, Eslick conceptualizes his emerging field of study in terms of individual human relationships, calling upon researchers to identify themselves clearly to participants when they use self-collected data:

"For any use of data collected there is moral requirement to document what has been done with the data, so that the participant can see what's going on. Even just a picture of the researcher and the fact they accessed their data would be a tremendous value, not because many people would use it, but because enough people would use it and we would have some real review on the entire process. Having a researcher putting themselves

³⁷ http://grants.nih.gov/grants/policy/hs/faqs_aps_dsm.htm

³⁸ <http://www.personalexperiments.org>; <http://c3nproject.org/>; <http://vitalreactor.com/>

out there as a person by saying, ‘This is who I am and how I’m putting your data to work’ would solve an interesting piece of the puzzle.”

We’d like to call attention to the fact that Eslick is far from the only professional researcher who also self-tracks for health. Self-trackers with advanced research skills can play a leadership role. For instance, the Nutrition Research Cohort (NRC), led by Andre Boorsma, a biostatistician at Maastricht University, specifically aims “to bridge the Quantified Self movement with professional research and healthcare innovation” by recruiting self-trackers from the scientific community.³⁹ By constituting themselves as a voluntary research cohort whose members use common self-tracking tools to generate data, including genomic data, that they themselves study, members of the NRC hope to benefit from an increased freedom to learn. The principles of the NRC give a vivid picture of this hybrid research model, including a component of self-funding:

- Each study participant owns and has continuous access to his/her data and results, embedded in the results obtained from the cohort study.
- Each participant provides all data and analysis at his/her own expense.
- Analysis is primarily based on “do-it-yourself” noninvasive or minimally invasive methods.
- Informed consent includes unlimited exploitation of all anonymous data and results, including any future use, covered by an open access license.
- The NRC is a cohort growing in size (number of participants) and complexity, where technologies will continuously adapt to new development, aiming to establish a new standardized methodology of self-quantification serving both research and (personal) health maintenance.
- The NRC cohort is composed of (nutrition) researchers. This allows optimal involvement of participants in shaping all aspects of the cohort (ownership) and makes aspects of the ethics handling more flexible.⁴⁰

The principles of the NRC explicitly address some of the key issues we’ve outlined, and could be used as a basis for experiments in access that attempt to significantly broaden participation.

Research Technique

The future of research design using self-collected data will involve techniques that support meaning-making at various scales. The size and complexity of self-collected data sets can grow very large, and technologically advanced methods are necessary for collecting, analyzing, and exploring data even for a single individual attempting to solve a personal problem.⁴¹ As self-collected data flows into the professional research community, tools for complex analysis can flow back out to make personal use of data more powerful. The work of James Fogarty and his colleagues, including Julie Kientz and Sean Munson in the department of Human Centered Design & Engineering at the University of Washington involves using advanced techniques in computer science to support self-understanding and reflection. As an illustration of how advanced techniques can help solve common problems for self-trackers, Fogarty proposed ‘ForgetBit’—software that

³⁹ <http://nrc.dbnp.org/>

⁴⁰ <http://nrc.dbnp.org/home/info?item=1>; see also: The nutrition researcher cohort: toward a new generation of nutrition research and health optimization [Link]

⁴¹ For new ideas about using embedded data, see: Kientz, Julie A. “Embedded capture and access: encouraging recording and reviewing of data in the caregiving domain.” *Personal and Ubiquitous Computing* 16.2 (2012): 209-221. [Link]

uses mobile phone data to fill in what's missing when a user's data is compromised by gaps caused by device failures, such as forgetting a tracker, breaking a device, or running out of battery charge.

This research direction interests us both for its practical value and its theoretical framework. Normally, when machine learning is referenced in research relating to public health, it's in the context of understanding large aggregated data sets.⁴² When the individual participant is seen as an active contributor to research, machine learning can be put to a categorically different use. As Fogarty puts it: "Machine learning is an opportunity to develop a language for interacting with your data." In a recent paper by Daniel Epstein and others from the University of Washington (including Fogarty) advanced analytical techniques were applied to complex lifelogging data in a research exploration carried out in collaboration with self-trackers, who were interviewed about their goals, and then asked to validate the analytical tools through rating them according to their personal usefulness.⁴³ Engaging individual self-trackers as qualified interpreters of the meaningfulness of their own data resembles the research carried out by toolmakers in the context of product development. In this case, however, the goal of the research is to expand a common toolset for discovery.

Access Prospects

In the course of preparing for our meeting, listening to the lively exchanges, and having dozens of follow-up conversations, we've come to conclude that access means more than simply being able to acquire a copy of relevant data sets. The purpose of access to data is to learn. When researchers and self-trackers think about self-collected data, they interpret access to mean "Can the data be used in my own context?" Self-collected data will change public health research because it ties science to the personal context in which the data originates. Public health research will change self-tracking practices by connecting personal questions to civic concerns and by offering novel techniques of analysis and understanding. Researchers using self-collected data, and self-trackers collaborating with researchers, are engaged in a new kind of skillful practice that blurs the line between scientists and participants. To us, improving access to self-collected data for personal and public benefit means broadly advancing this practice.

⁴² A sample of recent citations where both "machine learning" and "public health" appear gives a sense of the normal approach: Lopez Pineda, A., Tsui, F.-C., Visweswaran, S., & Cooper, G. F. (2013). Detection of Patients with Influenza Syndrome Using Machine-Learning Models Learned from Emergency Department Reports. *Online Journal of Public Health Informatics*, 5(1); Gupta, S., Tran, T., Luo, W., Phung, D., Kennedy, R. L., Broad, A., ... Venkatesh, S. (2014). Machine-learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry. *BMJ Open*, 4(3), e004007; Zhang, K., Li, Y., Schwartz, J. D., & O'Neill, M. S. (2014). What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods. *Environmental Research*, 132C, 350-359.

⁴³ See: Taming Data Complexity in Lifelogs: Exploring Visual Cuts of Personal Informatics Data [Link]

Appendix A

In this appendix you will find key references, links to videos of the 18 talks from the Quantified Self Public Health Symposium, and a link to a list of the attendees.

References

Health Data Exploration Project

Personal Data for the Public Good Report

- Executive Summary
- HDE Report
- Annotated Bibliography
- Appendices

Tracking for Health (Pew)

- Complete Report
- Topline Questionnaire
- Corresponding Data

Example Data Sharing Agreement

- University of Wisconsin & Propeller Health [\[pdf\]](#)

Additional References and Scholarly Articles

We've compiled a list of research papers that deal specifically with mobile health, personal data, and the quantified self. This list reflects contributions from the HCI, behavioral health, and social science communities and includes work by many of the researchers present at that symposium. All references footnoted in the preceding report are included in the linked Mendeley Group:

Quantified Self Public Health Symposium Research List

Notable papers include:

Hirsch, J. A., James, P., Robinson, J. R. M., Eastman, K. M., Conley, K. D., Evenson, K. R., & Laden, F. (2014). Using MapMyFitness to Place Physical Activity into Neighborhood Context. *Frontiers in public health*, 2, 19. [\[Link\]](#)

Fritz, T., Huang, E. M., Murphy, G. C., & Zimmermann, T. (2014). Persuasive Technology in the Real World: A Study of Long-Term Use of Activity Sensing Devices for Fitness. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 487–496). New York, New York, USA: ACM Press. [\[Link\]](#)

Thomaz, E. (2013). A Human-Centered Conceptual Model for Personal Health Informatics Data. *Personal Informatics in the Wild: Hacking Habits for Health & Happiness — CHI 2013 Workshop*. [\[Link\]](#)

Nilsen, W., Riley, W. T., & Heetderks, W. (2013). News from the NIH: using mobile and wireless technologies to improve health. *Translational Behavioral Medicine*, 3(3), 227–8. [\[Link\]](#)

Saranummi, N., Spruijt-Metz, D., Intille, S. S., Korhone, I., Nilsen, W. J., & Pavel, M. (2013). Moving the science of behavior change into the 21st century: novel solutions to prevent disease and promote health. *IEEE Pulse*, 4(5), 22–4. [\[Link\]](#)

Kumar, S., Nilsen, W., Pavel, M., & Srivastava, M. (2013). Mobile Health: Revolutionizing Healthcare Through Transdisciplinary Research. *Computer*, 46(1), 28–35. [\[Link\]](#)

Chen, C., Haddad, D., Selsky, J., Hoffman, J. E., Kravitz, R. L., Estrin, D. E., & Sim, I. (2012). Making sense of mobile health data: an open architecture to improve individual- and population-level health. *Journal of Medical Internet Research*, 14(4), e112. [\[Link\]](#)

Fleurence, R. (2012). The Rise of Patient-Centered Outcomes Research. *Health Outcomes Research in Medicine*, 3(1), e1–e2. [\[Link\]](#)

Backonja, U., Kim, K., Casper, G. R., Patton, T., Ramly, E., & Brennan, P. F. (2012). Observations of daily living: putting the “personal” in personal health records. *Nursing Informatics: Proceedings of the International Congress on Nursing Informatics, 2012*, 6. [\[Link\]](#)

Brennan, P. F., Downs, S., & Casper, G. (2010). Project HealthDesign: rethinking the power and potential of personal health records. *Journal of Biomedical Informatics*, 43(5 Suppl), S3–5. [\[Link\]](#)

Videos

All of the talks given at the Quantified Self Public Health Symposium were recorded and are available online. You can view all of the talks [here](#), or use the links to the individual talks below.

Introduction and Opening Remarks

Gary Wolf

Gary Wolf introduces the Quantified Self Public Health Symposium.

Susannah Fox

Susannah Fox shares research from the Pew Internet and Life Project and describes the challenges ahead for promoting self-tracking.

Stephen J. Downs

Stephen Downs from the Robert Wood Johnson Foundation discusses their new focus on a “Culture of Health” and their reasons for supporting the symposium.

Bryan Sivak

Bryan Sivak, CTO of the U.S. Department of Health and Humans Services, gives his opening remarks and discusses the challenges to using self-tracking data for public health benefit.

Data Flows

Anne Wright

Anne Wright (Fluxstream/Body Track) discusses data access and the hard problems worth solving in personal data systems.

Margaret McKenna

Margaret McKenna (RunKeeper) explores the issues, challenges, and ideas large-scale self-tracking applications have in mind when they consider working with the research community.

Andy Hickl

Andy Hickl (ARO/Saga) describes the role of context in designing and using personal data systems.

Ida Sim

Ida Sim describes Open mHealth and their process of creating standardization for personal health data systems.

Deep Personal Data

Larry Smarr & Leroy Hood

Larry Smarr (UCSD/Calit2) and Leroy Hood (Institute for Systems Biology) talk about their experience with Quantified Self data and new personal data projects.

Participation (1)

Jason Bobe

Jason Bobe (Personal Genomes Project) talks about the lessons learned from involving research participants in the data ownership and discovery process.

Eric Hekler

Eric Hekler (Arizona State University) discusses new ideas and methods for involving participants in the research process.

Matthew Kay

Matthew Kay (University of Washington) talks about how researchers are exploring the design of personal data tracking tools and systems.

Participation (2)

Ian Eslick

Ian Eslick (Vital Reactor) discusses the role of personal experimentation and personal data in the medical and scientific process.

Doug Kanter

Doug Kanter (Databetes) describes what he's learned from tracking and visualizing his diabetes data.

James Fogarty

James Fogarty (University of Washington) describes the role of machine learning for helping individuals understand and make sense of their personal data.

Closing Remarks

Rachael Fleurence

Rachael Fleurence (PCORI) talks about the role of PCORI and patient-centered research groups and projects.

Kevin Patrick

Kevin Patrick (UCSD/Center for Wireless and Population Health Systems) gives his closing remarks and takeaways from the Quantified Self Public Health Symposium.

Bryan Sivak

Bryan Sivak (HHS) gives his closing remarks at the 2014 Quantified Self Public Health Symposium.

Reactions

The following links contain reactions and thoughts from some of the attendees present at the symposium.

Quantified Self Public Health Symposium by Susannah Fox

Quantified Self Public Health Symposium by Doug Kanter

Quantified Self Public Health Symposium by Edison Thomaz

I Have The Problem in My Grasp by Roni Zeiger

Personally Generated Health Data—The Next Frontier by Bryan Sivak

Quantified Self Public Health Symposium tweet collection by Joyce Lee

Attendees

The following attendees were present at the 2014 Quantified Self Public Health Symposium. We've categorized attendees into four categories: Conveners, Toolmakers, Researchers, and Quantified Self Labs staff. Contact information for each attendee can be found [here](http://quantifiedself.com/symposium/Symposium-2014/attending.php). (<http://quantifiedself.com/symposium/Symposium-2014/attending.php>)

Conveners

Gary Wolf Quantified Self Labs	Gary Wolf is the director of Quantified Self Labs, a social enterprise whose mission is to support the Quantified Self movement worldwide.
Bryan Sivak CTO, HHS	Bryan Sivak joined HHS as the Chief Technology Officer in July 2011. In this role, he is responsible for helping HHS leadership harness the power of data, technology, and innovation to improve the health and welfare of the nation.
Larry Smarr Director, Calit2	Larry Smarr is the founding director of the California Institute for Telecommunications and Information Technology (Calit2), a UC San Diego/UC Irvine partnership, and holds the Harry E. Gruber professorship in Computer Science and Engineering (CSE) at UCSD's Jacobs School.
Steve Downs Robert Wood Johnson Foundation	Steve is the Robert Wood Johnson Foundation's chief technology and information officer. In this role, he ensures that the Foundation's technology strategy and operations are aligned with its organizational plans. Current Projects: OpenNotes Health Data Exploration Project HealthDesign

Toolmakers

Andy Hickl A.R.O., Inc.	Andy is the cofounder and CEO at A.R.O., Inc. He is interested in ubiquitous computing, machine learning, and natural language processing. He is also a passionate lifelogger.
Aaron Coleman Small Steps Labs	Aaron is the creator of Fitabase by Small Steps Labs. Fitabase is a research platform that collects data from internet-connected consumer devices. It is designed for anyone wanting to aggregate, analyze, and export data gathered from many device wearers.

<p>Anne Wright BodyTrack, Fluxstream</p>	<p>Anne Wright is an MIT-trained roboticist and former lead systems engineer for prototype rovers at NASA Ames Research Center. She leads the BodyTrack Project at CMU, building tools and culture to support individuals' efforts toward personal health empowerment.</p>
<p>Tony Stubblebine Lift</p>	<p>Tony is the founder and CEO of Lift, a mobile- and web-based habit tracking and coaching platform.</p>
<p>Ian Eslick Vital Reactor, LLC</p>	<p>Ian has been working as a technology entrepreneur and academic for nearly 25 years. He recently completed a PhD exploring how personal data can be used to improve decisions and advance scientific discovery. His new company, Vital Reactor, is working with a variety of healthcare organizations on products that use personal data to transform clinical care.</p>
<p>Robert Evans Google, PACO</p>	<p>Bob develops tools to support analysis and exploration of daily experience. For the past three years as a software engineer at Google, Bob has been working on PACO (Personal Analytics COmpanion). PACO allows individuals and behavior scientists to easily create and conduct behavior studies and interventions on mobile phones.</p>
<p>David Haddad Open mHealth</p>	<p>David is the director of Open mHealth, a nonprofit that envisions a world where disparate mobile health applications can be harnessed to tell a data-driven patient story.</p>
<p>Anna de Paula Hanika Open mHealth</p>	<p>Anna has recently joined Open mHealth to lead product strategy & marketing, following 6.5 years as a product marketing manager at Google, and is excited to be solving for how integrated data and design can transform the way we manage personal health.</p>
<p>Dawn Nafus Intel</p>	<p>Dawn is an anthropologist with Intel Labs, where she conducts ethnographic research to inform new product development and strategy. She holds a PhD from the University of Cambridge and has research interests in anthropology of measurement and experiences of time.</p>
<p>Sangita Sharma Intel</p>	<p>Sangita is a researcher at Intel Labs with a background in data processing and analysis. She holds a PhD in Electrical and Computer Engineering and her research interests include creating algorithms to analyze data and derive contextual information from multiple data sources.</p>

David Reeves Limeade	David is the Product director at Limeade, a well-being improvement service for employees and health plan members, and started the Seattle Quantified Self Meetup.
Ryan Beckland Validic	Ryan is the CEO and cofounder of Validic. Validic, an FDA Class I MDDS, delivers actionable, standardized, HIPAA-compliant consumer health data to healthcare companies.
Drew Schiller Validic	Drew is the CTO and cofounder of Validic. Validic, an FDA Class I MDDS, delivers actionable, standardized, HIPAA-compliant consumer health data to healthcare companies.
Anmol Madan Ginger.io	Anmol leads the overall vision and strategy at Ginger.io, and his past research has been in modeling large-scale human behavior data using statistical and pattern recognition methods. His work has been featured extensively in academic publications and popular media. He received his PhD from the MIT Media Lab.
Mike Polcari 23andMe	Mike Polcari is the director of software engineering at 23andMe. Mike works with a talented team of software engineers to bring personal genetics to the masses and advance humanity's understanding of genetics through 23andMe's Personal Genome Service, open API, and research platform.
Bill Davenhall Esri	Bill is the former global industry manager for health and human services at Esri for 15 years. He created the My Place History app (PC and iOS versions) in 2010, which links personal history of street address to toxic chemical data from the EPA.
Chris Hogg Practice Fusion	Chris leads the Data Science team at Practice Fusion, a leading free EMR service.
Beau Gunderson Practice Fusion	Beau is a senior software gardener at Practice Fusion. He is also a leading QS developer and open data advocate. Current Project: 23andMe Data Exporter
Carlos Olguin Autodesk, Inc.	Carlos heads the Bio/Nano/Programmable Matter Group at Autodesk Research. They are investigating the design spaces enabled by matter programming across domains and scales by collaborating with researchers around the world to co-envision the paradigms and tools needed to establish a robust scale-free body of knowledge of design.

<p>Doug Kanter Databetes</p>	<p>Doug is a type 1 diabetic, designer, and founder of Databetes. As a patient, he is interested in developing tools for improved self-management support. Doug is also a recent graduate of NYU's ITP, where he studied data visualization and interaction design.</p>
<p>Marco Della Torre Basis Science</p>	<p>Marco is a biomedical engineer, serial entrepreneur, and co-founder of Basis Science.</p>
<p>Khaled Hassounah MedHelp</p>	<p>Khaled focuses on bringing health quantification to the masses through a platform serving MedHelp's own website and a range of partners reaching tens of millions of active users.</p>
<p>David Albert AliveCor</p>	<p>David is a physician, scientist, inventor, and entrepreneur. He recently founded AliveCor, the creators of the first iPhone ECG.</p>
<p>Jason Bobe The Personal Genomes Project</p>	<p>Jason focuses on equitable access to biomedical technologies, expanding the definition of "researcher" to encompass broader audiences, creating public data resources and related issues like "open consent," and promoting research design that incorporates "sharing by default." Current Project: Open Humans Network</p>
<p>Marion Handler Qualcomm</p>	<p>Marion is a clinical health psychologist with extensive training and experience in behavioral medicine and health promotion as well as in high technology and wireless media to promote healthy behaviors. She is currently the senior manager of global employee health services at Qualcomm.</p>
<p>John Mattison Kaiser Permanente</p>	<p>John is deeply involved with person-centric care as well as federal and state policy on health information exchange. He is a member on the HHS/ONC Technical Expert Panel on Patient Generated Health Data and is interested in big data analytics.</p>
<p>Brian Foye Humana</p>	<p>Brian is a director in Human Resources for Humana, Inc., a \$40B U.S. healthcare company that offers a wide range of insurance products and health and wellness services that incorporate an integrated approach to lifelong well-being.</p>
<p>Amar Kendale MC10</p>	<p>Amar is passionate about bringing the needs of consumers, patients, providers, and payers into focus, in order to deliver enabling personal biosensing technology that will improve health and lives.</p>
<p>Roni Zeiger Smart Patients</p>	<p>Roni is the CEO of Smart Patients and the former chief health strategist at Google.</p>

Rajiv Mehta Unfrazzle	Rajiv’s work focuses on empowering people to better understand and manage their own and their family’s health. He consults on family-caregiver innovation, and is also the developer of the Unfrazzle app, a board member of the Family Caregiver Alliance, and co-organizer of the Bay Area Quantified Self meetup group.
Sean Ahrens Crohnology	Sean is a patient-centered health software designer and developer. He is currently focused on building Patient-Powered Research Networks.
Eric Blue TRAQS	Eric is a technologist, entrepreneur, and aspiring alpha geek with a passion for the quantified self, health, fitness, mind mapping, productivity, and learning.
Jon Hemphill Garmin	Jon is a product manager for the Health and Wellness product line at Garmin.
Adam Dole HHS/Mayo Clinic/Better	Adam is on sabbatical from the Mayo Clinic and Better as a Presidential Innovation Fellow, working with the White House and U.S. Dept. of Health and Human Services on the MyData Blue Button Initiative.
Ilene Klein Qualcomm	Ilene is a family physician leading Qualcomm’s health and wellness strategy for its global workforce through transformative connected health services and data. She is also the chair of Qualcomm Life’s Advisory Council.
Michael Nagle Lark Technologies	Michael is the former organizer of Boston’s QS meetup, advisor to Ginger.io, and R&D engineer at Lark Technologies.
Minna Ahn Samsung	Minna works in business development at Samsung’s new Open Innovation Center, based in Silicon Valley, NYC, and Korea, which invests in, acquires, and partners with entrepreneurs and startups with great ideas.
Mike Lee MyFitnessPal	Mike is the founder and CEO of MyFitnessPal. MyFitnessPal makes it easy to track the food that you eat and the exercise that you perform so that you can make healthier decisions.
Nick Crocker MyFitnessPal	Nick is a product manager at MyFitnessPal. He co-founded Sessions (recently acquired) and is excited about designing behavior change with humans and technology.

Walter De Brouwer Scanadu	Walter is the founder and CEO of Scanadu.
Richard Henrikson Scanadu	Rick is the director of mobile fluidics at Scanadu.
Jeroen Brouwer Philips DirectLife	Jeroen helped to launch the Philips DirectLife activity program as well as the ActiveLink program jointly with Weight Watchers.
Jessica Richman uBiome	Jessica is co-founder and CEO of uBiome, the first crowd science microbiome sequencing company. Jessica has spoken at TEDMED and SXSW and her work has been featured in the MIT Technology Review, Wired, Fast Company, and dozens of other media and conferences.
Matthew Diamond Misfit Wearables	Matthew is a rehabilitation and sports medicine physician and scientist and the medical lead at Misfit Wearables, makers of highly wearable computing products. A faculty member at NYU, Matthew is passionate about helping people improve their mobility and performance through innovative, integrative medical treatments and technology that promotes wellness.
Rob Rustad Open Health Cloud	Rob is a technology product development veteran focused on finding better ways to put health data to work for patients and research.
Emil Chiauzzi PatientsLikeMe	Emil is the research director at PatientsLikeMe. He is a research clinical psychologist specializing in behavioral health, addictive behavior, health behavior change and motivational strategies, outcomes research, and interactive health assessments/evaluation.
Esther Dyson Health Initiative Coordinating Council	Esther is the chairman of EDventure Holdings and an active angel investor in health companies.
Bryan Campen XCOR	Bryan is the director of Media and Public Relations for XCOR Aerospace, a spacecraft manufacturer and flight provider based in Mojave, California. He is also the founder of the QS Chicago chapter.
Jesse Slade Shantz OMSignal	Jesse is an orthopedic sports surgeon driven to redefine performance and precision outcomes through bio-sensing apparel.

<p>Kyler Eastman MapMyFitness</p>	<p>Kyler is an R&D engineer at MapMyFitness. He is looking for partnerships with academic researchers to leverage 250M GPS workout tracks worldwide for advancements in the understanding of public health, exercise physiology, and human behavior.</p>
<p>Dean Hovey Digifit</p>	<p>Dean, the CEO of Digifit, is a native of Silicon Valley, actively promoting innovation as a product designer, entrepreneur, and venture capitalist.</p>
<p>David Goldsmith Dossia</p>	<p>David is the executive director of the Dossia Consortium, a not-for-profit association focused on the use of a secure, patient-centric health platform to deliver actionable applications for patients and providers to improve the safety, quality, and efficiency of healthcare.</p>
<p>Margaret McKenna RunKeeper</p>	<p>Margaret is a software engineer and the head of Data & Analytics at RunKeeper. She is interested in the intersection of design, data, and behavior.</p>
<p>Steve Zadig Vital Connect</p>	<p>Steve is the co-founder of Vital Connect, a biosensor patch technology for very low cost continuous health monitoring.</p>
<p>Gil Blander Inside Tracker</p>	<p>Gil brings to Segterra more than 18 years of experience in systems biology, computational biology, aging, metabolism, and caloric restriction research. Gil previously worked at MIT (Guarente Lab), the Weizmann Institute (Israel), and several Systems Biology / Computational Biology companies (Genstruct, Compugen, and IBS).</p>
<p>Margie Ploch Inside Tracker</p>	<p>Margie is science project manager and researcher at Segterra, focusing on content and presentation of personal biomarker data.</p>
<p>Max Utter Jawbone</p>	<p>Max brings his experience building natural health practices to the research of health and behavior as the lead behind research partnerships for UP by Jawbone.</p>

Researchers

<p>Kevin Patrick University of California, San Diego</p>	<p>Kevin is a professor of Family and Preventive Medicine and director of the Center for Wireless and Population Health Systems at the Qualcomm Institute at UCSD. His research explores how to use mobile and social technologies to measure and improve health-related behaviors. Current Projects: Health Data Exploration Delphi</p>
<p>Patti Brennan University of Wisconsin</p>	<p>Patti designs, deploys, and evaluates people-facing technologies for health. In her work she also explores the use of a virtual reality CAVE to accelerate context-aware design. Current Project: Project HealthDesign</p>
<p>Audie Atienza National Cancer Institute</p>	<p>Audie is a senior researcher at the National Institutes of Health. He has been a health technology advisor for the NIH Office of the Director and the HHS Office of the Secretary, and has collaborated with ONC, the Surgeon General, and the White House.</p>
<p>Simon Marshall University of California, San Diego</p>	<p>Simon is professor of Family and Preventive Medicine, specializing in health behavior change and physical activity & public health. He serves as principal investigator and co-investigator on numerous NIH-funded grants examining technology-mediated behavior change in exercise and physical activity contexts.</p>
<p>Genevieve Dunton University of Southern California</p>	<p>Genevieve is the director of the USC REACH (Real-Time Eating Activity and Children’s Health) lab, which develops, tests, and applies real-time data capture methodologies to better understand the effects of environmental, social, and psychological factors on eating and physical activity. Current Project: REACH Lab</p>
<p>Colin Depp University of California, San Diego</p>	<p>Colin is a clinical psychologist and associate professor of Psychiatry interested in mobile health applications for mental health.</p>
<p>David Hickam PCORI</p>	<p>David is director of the PCORI Clinical Effectiveness Program. This program is charged with developing and managing an extramural portfolio of clinical comparative effectiveness research. The research projects address critical evidence gaps on the effectiveness of a wide range of clinical services.</p>

<p>Tim Hale Connected Health</p>	<p>Tim is a medical sociologist, studying how information and communication technologies are transforming healthcare.</p>
<p>Stephen Friend Sage Bionetworks</p>	<p>Stephen is president of Sage Bionetworks and a world leader in efforts to make large-scale, data-intensive biology more openly accessible in order to accelerate scientific progress. In June 2013, the White House recognized him as an Open Science Champion of Change for his efforts to run Big Data Health projects. Current Project: Sage Research Programs</p>
<p>Chris Pyke US Green Building Council</p>	<p>Chris is the vice president of research for the U.S. Green Building Council. He directs a diverse research portfolio leveraging green building project experience to gain practical insights into building performance, occupant experience, and market trends. Current Project: Green Building Information Gateway</p>
<p>Roozbeh Jafari University of Texas at Dallas</p>	<p>Roozbeh Jafari is an associate professor at UT-Dallas and director of the ESSP Lab. His research interest lies in the area of wearable computer design and signal processing. Current Project: Ultra-low Power Wearable Computers</p>
<p>Matthew Trowbridge University of Virginia</p>	<p>Matthew is a physician and public health researcher at the University of Virginia. His academic research focuses on the impact of architecture, urban design, and planning on public health. Current Project: Green Health Partnership</p>
<p>Nadir Weibel University of California, San Diego</p>	<p>Nadir is a research faculty member at UC San Diego, and a research health science specialist at the VA San Diego. He is a computer scientist investigating the deployment of innovative interactive multimodal devices in context, and an ethnographer studying the cognitive consequences of the introduction of this technology in the everyday life.</p>
<p>Jon Froehlich University of Maryland</p>	<p>Jon is an assistant professor at UMD in CS, founder of the Makeability Lab, and member of the Human-Computer Interaction Lab. Jon spends his time making, breaking, and studying stuff for the environment, health, and the city. Current Project: Social Fabric Fitness</p>
<p>Eugene Kolker Seattle Children's</p>	<p>Eugene is chief data officer at Seattle Children's and head of the Bioinformatics & High-throughput Analysis Laboratory at Seattle Children's Research Institute. He is also the co-founder of DELSA. Current Projects: Kolker Lab Data-Enabled Life Science Alliance</p>

<p>Donna Spruijt-Metz University of Southern California</p>	<p>Donna is director of the USC mHealth Collaboratory, and focuses her research on childhood obesity and development and use of mobile technologies to understand and change behavior in real time. Her team developed one of the first wireless body area network systems designed to intervene in real time to change physical activity behaviors in youth. Current Projects: Virtual Sprouts Robots Helping Kids</p>
<p>Ida Sim University of California, San Francisco, Open mHealth</p>	<p>Ida Sim, MD, is a professor of Medicine, co-director of biomedical informatics at UCSF's Clinical and Translational Sciences Institute and co-founder of Open mHealth, a nonprofit organization that is breaking down barriers to mobile health app and data integration through an open software architecture. Current Project: Open mHealth</p>
<p>Matthew Kay University of Washington</p>	<p>Matthew is a PhD student in computer science doing research in personal informatics. His particular focus is on understanding how people interpret complex, often noisy data, and what lessons we can learn to more systematically build tools that help people deal with complex personal data.</p>
<p>Leslie Saxon University of Southern California</p>	<p>Dr. Leslie Saxon is the executive director of the USC Center for Body Computing, a digital health innovation and research center. At the USC Center for Body Computing she brings together various USC schools, including Cinematic Arts, the Marshall School of Business, the Viterbi School of Engineering, and innovators to create the future of digital medicine.</p>
<p>Joyce Lee University of Michigan</p>	<p>Joyce is a pediatrician, diabetes specialist, researcher, and design thinker. She co-directs a mobile technology program at the University of Michigan and is interested in the use of mobile technology and data visualization for improving type 1 diabetes outcomes in adolescents and young adults.</p>
<p>Eric Hekler Arizona State University</p>	<p>Eric's research focuses on how to design and evaluate health behavior change technologies. Recent awards include a Best Paper Award at the 2013 ACM SIGCHI Conference and a Google Research Award. Current Projects: Designing Health Agile Science</p>
<p>Jose Gomez-Marquez MIT</p>	<p>Jose runs the Little Devices lab at MIT, where they make DIY medical technologies for affordable health and develop tools for the developing world. He is interested in empowerment technologies for patients and providers so they can come up with their own solutions. Current Project: Maker Nurse</p>

<p>Edison Thomaz Georgia Tech</p>	<p>Edison focuses his research on computational health, where he applies activity and lifestyle recognition approaches toward the domain of healthcare. He combines machine learning, ubiquitous computing, and HCI techniques to quantify health-related behaviors of individuals and populations.</p>
<p>James Fogarty University of Washington</p>	<p>James is an associate professor of Computer Science & Engineering at the University of Washington. His research is currently examining end user interactive machine learning as a tool for empowering end users to make sense of and act upon personal informatics data.</p>
<p>Priel Schmalbach University of California, Irvine</p>	<p>Priel is a physician scientist in training. His research interests include health promotion with mobile applications.</p>
<p>Jason Levine National Cancer Institute</p>	<p>Jason spends most of his time living at the interface of healthcare and IT at the National Cancer Institute, and in pursuit of a better understanding of fitness and health.</p>
<p>Hugo Campos Stanford, Medicine X</p>	<p>Hugo is a thought-leading crusader in the ePatient movement. Passionate about participatory medicine, connected health, and patient empowerment through the use of technology, he advocates for the rights of patients with pacemakers and implantable defibrillators to gain electronic access to the data collected by their devices.</p>
<p>Jerry Sheehan University of California, San Diego</p>	<p>Jerry serves as the chief of staff for the California Institute for Telecommunications and Information Technology (Calit2). In addition, he is co-director of the RWJ Support Health Data Exploration Project. Current Project: Health Data Exploration</p>
<p>Alexandra Hubenko University of California, San Diego</p>	<p>Alexandra is the program manager at the Qualcomm Institute and for the last 9 years has been working on a variety of projects including information technology for disaster response and disaster medicine; cultural heritage analytics and diagnostics; underwater imaging systems; and health data. Current Project: Health Data Exploration</p>
<p>Margie Morris Intel</p>	<p>Margie is a clinical psychologist who explores how technology can be emotionally helpful.</p>

<p>Puneet Kishor Creative Commons</p>	<p>Puneet likes to write programs that manipulate, analyze, and visualize information from large data sets, but he worries what might happen to those data and results 50 or 100 years from now. Puneet's main focus at Creative Commons is on science data policy.</p>
<p>Katherine Kim UC Davis</p>	<p>Kathy is a researcher focused on patient and stakeholder engagement in health technology, including mobile, social networking, and distributed research networks. She is a visiting assistant professor at UC Davis, Betty Irene Moore School of Nursing. Current Projects: Project Health Design pSCANNER Mobile Health and Tribal Food Security</p>
<p>Andrew Rajj Univeristy of South Florida</p>	<p>Andrew is an assistant professor of Electrical Engineering and the director of the Powerful Interactive Experiences (PIE) laboratory at the University of South Florida. The PIE lab is an interdisciplinary human-computer interaction lab with a mission to create and study new systems and experiences for learning, reflection, and behavior change.</p>
<p>David Mohr Northwestern</p>	<p>David is a professor at Northwestern University and director of the Center for Behavioral Intervention Technologies (CBITs). His work involves the development of mobile applications to treat and prevent depression, including context sensing methods relevant to the behavioral management of depression, and management of pharmacotherapy.</p>
<p>Stephen Schueller Northwestern</p>	<p>Stephen is a research assistant professor in the Department of Preventive Medicine and a member of the Center for Behavioral Intervention Technologies (CBITs) at Northwestern University. His research focuses on using Internet and mobile interventions for the treatment and prevention of depression and the promotion of happiness and well-being.</p>
<p>Chris Karr Northwestern</p>	<p>Chris is the developer of Purple Robot, and tech lead on native software development at CBITs.</p>
<p>Arno Klein Sage Bionetworks</p>	<p>Arno is the director of neuroimaging and principal scientist of systems biology at Sage Bionetworks (nonprofit) in Seattle. He is excited about brain imaging and visualization of complex information.</p>

<p>Julie Kientz University of Washington</p>	<p>Julie is an assistant professor in Human Centered Design & Engineering at the University of Washington. Her research focuses on lowering the user burden of technology for health and behavior change. She has designed and studied applications for sleep, children’s health, autism, smoking cessation, and physical fitness.</p>
<p>Misha Pavel Northeastern University</p>	<p>Misha Pavel is a professor at the College of Computer and Information Sciences and the Bouvé College of Health Sciences at Northeastern University. His research is focused on unobtrusive monitoring, neurobehavioral assessment, and computational modeling in support of healthcare, with a particular focus on chronic disease and elder care.</p>
<p>Margo Edmunds Academy Health</p>	<p>Margo is a medical psychologist and policy researcher by training. She oversees a variety of projects at the intersection of health policy, research, clinical practice, and informatics and teaches strategic communications at Johns Hopkins. Current Projects: EDM Forum</p>
<p>Scott Klemmer University of California, San Diego</p>	<p>Scott is an associate professor at UCSD and visiting associate professor at Stanford. His awards include the Katayanagi Emerging Leadership prize, Sloan Fellowship, and Microsoft New Faculty Fellowship. He advises organizations on design, and people worldwide use his curricula.</p>
<p>Rachael Fleurence PCORI</p>	<p>Rachael is the program director for PCORnet, a national research network that just launched in January 2014. PCORnet includes 18 patient-powered research networks comprised of activated groups of patients who will be exploring ways to share data for research. Current Projects: PCORnet</p>
<p>Kerry Anne McGeary Robert Wood Johnson Foundation</p>	<p>Kerry Anne joined the Robert Wood Johnson Foundation in 2013 as a senior program officer in the Research and Evaluation unit. Coming to RWJF after a distinguished career as a professor of health economics at Ball State University, Drexel University, and the University of Miami, she employs her background in health economics and health policy research to help the Foundation achieve its mission and to assist its researchers in promoting a culture of health.</p>

Quantified Self Labs

<p>Davis Masten The National Academies, Quantified Self Labs</p>	<p>Davis is the co-chair of the Presidents' Circle at The National Academies and a distinguished visiting scholar at Stanford MediaX. He is chair of the advisory board of Quantified Self Labs.</p>
<p>Ernesto Ramirez Quantified Self Labs</p>	<p>Ernesto is the program director for QS Labs.</p>
<p>Kate Farnady Quantified Self Labs</p>	<p>Kate is the development director for QS Labs.</p>
<p>Steven Jonas Quantified Self Labs</p>	<p>Steven is an organizer for the Quantified Self group in Portland and helps coordinate speakers for Quantified Self conferences.</p>
<p>Maarten den Braber Quantified Self Labs</p>	<p>Maarten is a digital health strategist, the co-founder of Quantified Self Europe / Amsterdam, and a co-organizer for Singularity University NL.</p>
<p>Sunny Bates Red Thread</p>	<p>Sunny holds leading roles in the technology community. She is on the board of directors at Kickstarter and the Creative Capital Foundation; on the advisory board of Endeavor, Jawbone, Maxwell Health, Acumen Fund, American Theater Wing, Girls Who Code, Startx, Thiel Fellowship Mentors, TechStars, TEDMED, and a founder of Summit Eden.</p>

Appendix B

Quantified Self APIs and Data Flows

Narrative Report and Sample Diagrams of QS Systems

Robin Barooah, Steven Jonas, and Gary Wolf

Introduction

We take for granted that there is a lot of self-collected data “out there.” What are the facts on the ground? What type of data is being collected, and by what instruments? How can it be accessed? What is the most useful way to represent the QS data ecosystem so that flows can be improved? The purpose of this draft report is to support work on the hard problems of access, privacy, research participation, and data quality with a comprehensible—if necessarily complex—picture of how QS data flows in and out of some typical self-tracking systems.

In our conversations with public health researchers who have an interest in self-collected data, we’ve been frequently reminded that scientific investigators care intensely about the quality of their data. Researchers want data produced by standard instruments, collected using well-defined protocols, reflecting measures that have a known relevance to public health.

Self-trackers also care about their data. Self-trackers want data based on common measurements; collectable without too much effort; expressed in everyday units; and easily downloadable, archived, and graphed.

While there is no necessary antagonism between scientific and personal data practices, neither is there easy alignment. Therefore, the question of how to support access to self-collected data by public health researchers cannot simply be a matter of turning on the tap or getting permission. Instead, an active collaboration is necessary, in which researchers and self-trackers either mutually benefit in achieving different goals, develop shared goals, or both. We’re optimistic that the answer can be “both.” But, the first step to tackling the problems involved is recognizing that such problems exist, and exploring them in some detail. We hope these diagrams help. They were conceived and designed by Robin Barooah, researched and drawn by Steven Jonas; Gary Wolf wrote the narrative.

Our Point of View

Although our goal in making the diagrams and writing the supporting material is to aid collaboration between self-trackers, toolmakers, and researchers, it’s important to note that we have a stake in the discussion, and we’ve drawn the diagrams from our side; that is, from a self-tracker’s perspective. By self-tracker, we do not mean “typical self-tracker.” Obviously, some domain knowledge and technical judgment were used in making the graphs. However, they are drawn from the self-tracker’s perspective in the sense that the data flows from one personally relevant context to another: from one consumer device, app, or service where we can see it and use it, to another where we can see it and use it. There are other ways self-collected data could flow. For instance, a commercial service may sell data access to partners in a private transaction, or donate it to health researchers. We did not consider these types of data flows. These diagrams also represent a self-tracker’s point-of-view in that all the information represented in the graphs was gathered either by looking at publicly available web pages or by tinkering. (Not all user-relevant data flows are officially documented.)

How to Use the Diagrams

The diagrams are meant to be helpful in two ways. First, they contain reported detail about data access procedures in a sample set of self-tracking systems, selected to expose a range of approaches, and organized by commercial brand. Second, and more importantly, they offer a conceptual model of data flows that can be used by others, including toolmakers themselves, to understand and express how self-trackers and researchers may be able to access self-collected data. The best way to begin reading the diagrams is by following the data from left to right as it flows through three technological “zones”: Devices, Applications, and Services.

A **device** is an application-specific computer, such as an activity tracker. These include the bands, straps, scales, and other gizmos familiar to all of us. It’s important to remember that even little toy-like trackers in a colorful plastic coating are in fact computers running software, and the measurements they make are influenced both by the electromechanical systems they contain and by the software they run. We consider something a device if the software it runs is “baked into” the instrument, perhaps updatable by the vendor remotely but normally invisible to the user.

An **application** is software running on a general purpose personal computer, where personal computer includes smart phone, tablet, or traditional PC. Increasingly, self-tracking applications can use the native sensors installed in the platform, as iOS applications can use Apple’s M7 chip. Thus self-tracking data flows may not originate in external devices, but instead originate in the application zone.

A **service** is software running on a vendor’s infrastructure. Most opportunities, though not all, for pulling data out of a commercial ecosystem occur in the service zone. Services are normally accessed by users through websites, and by 3rd party developers through APIs. Importantly, self-trackers can also add data to a system via a website, so that new measurements can enter the system from the service zone.

Data types available through a vendor’s API are represented on the diagram as boxed text. (For instance, see the Fitbit diagram, page 48). How temptingly simple these boxes appear! But by following the lines of the diagram backward, the underlying complexity of the data becomes visible.

We hope these diagrams will support critical reflection on the common belief that improving access to self-collected data through common APIs, aggregation schemes, and measurement standards will satisfy research needs. Simplicity of access is a dangerous simplicity when it silently erases the provenance of the data. Our diagrams aim to illustrate some of the places where provenance is relevant. Questions that these diagrams can help answer include:

- Can I upload data from a particular device directly into a website without going through a smart phone?
- Can a single data type (steps, for instance) accessed via this API represent measurements made by different devices?
- Can data accessed via the API include data from more than one service?

Highlights

Provenance in an Aggregator: The diagrams should inspire anybody who intends to use APIs to support research to consider the provenance of their data. For a example, consider the RunKeeper diagram.

RunKeeper has a well-maintained API that serves data aggregated from other self-tracking systems, called the Health Graph API. A glance at the diagram shows that there are many devices and services that can flow data into the RunKeeper cloud and therefore will show up in data accessed via the API. What do calls to the API reveal about the sources of this data? In the case of RunKeeper, we discovered through reading the documentation and tinkering that in the Health Graph API “source” refers to the application that most recently handled the data. RunKeeper has validation procedures in place to distinguish between manually entered data and data automatically gathered by applications, as well as which applications gathered the data, but it appears that device identity is lost in the flow of data through an application into the API. Our diagrams cannot answer provenance questions about self-tracking ecosystems in detail. However, the diagrams can be helpful in identifying circumstances where establishing provenance is likely to be difficult.

Much ongoing discussion in the QS community about best practices for data flows has been focused on provenance. These graphs may provide some background for the discussion of the proposed Open mHealth architecture. One key innovation of Open mHealth lies in its recommendations for moving the discussion of data standards away from a reference list of standard data types and toward a reference system that makes it possible to track provenance.

Device Compatibility within a Smart Phone Ecosystem: The Samsung / S Health diagram provides a look at a different model of data access via an API to measurements from heterogeneous devices. The five Galaxy phones currently compatible with S Health connect via various wireless communication protocols with an immense range of devices, and Samsung provides affordances for application developers to collect data via S Health and flow it outside the Samsung ecosystem. The Samsung system aims to integrate a wide range of automatically gathered health data at the application level, with “openness” as a philosophy on both ends. This approach maximizes device compatibility at the cost of restricting flows in both the application and the service zone to dedicated Samsung Galaxy developers. While little is known about the much rumored Apple self-tracking architecture, the Samsung diagram of device compatibility within a single smart phone ecosystem provides an existing architecture against which to speculate.

Two Single Service APIs: Foursquare and Nike+ are similar in almost no respects, but the flow of data in these two systems makes an interesting comparison. Both offer single service APIs. Data downloaded from Foursquare or Nike+ unambiguously originates in the service of the API owner. Foursquare data travels freely outside the Foursquare ecosystem, available via an API to registered 3rd party apps and, via feeds, to any user or user-authorized app. Nike+ data, on the other hand, is closely guarded within the Nike+ ecosystem. It originates on a Nike+ device, is viewable on a Nike+ website, and can be used only within registered 3rd party apps. Some of these registered 3rd party apps may initiate data flows of their own; perhaps allowing download or aggregation of Nike+ data via the secondary source, but as far as we’ve been able to determine these flows are not publicly visible.

Simple Data Flow without an API: Sleep Cycle is a sleep tracking app that offers users free download of their data in .csv format. But there is another way data flows from the Sleep Cycle app. That is, into the RunKeeper ecosystem via RunKeeper’s Health Graph API. From here, Sleep Cycle data can flow into any of RunKeeper’s registered 3rd party apps. Data that begins in an application represented by the simplest of our diagrams is, in fact, only one node away from the full complexity of these interlinked ecosystems. RunKeeper reports that over 81,000 RunKeeper users also use Sleep Cycle.⁴⁴

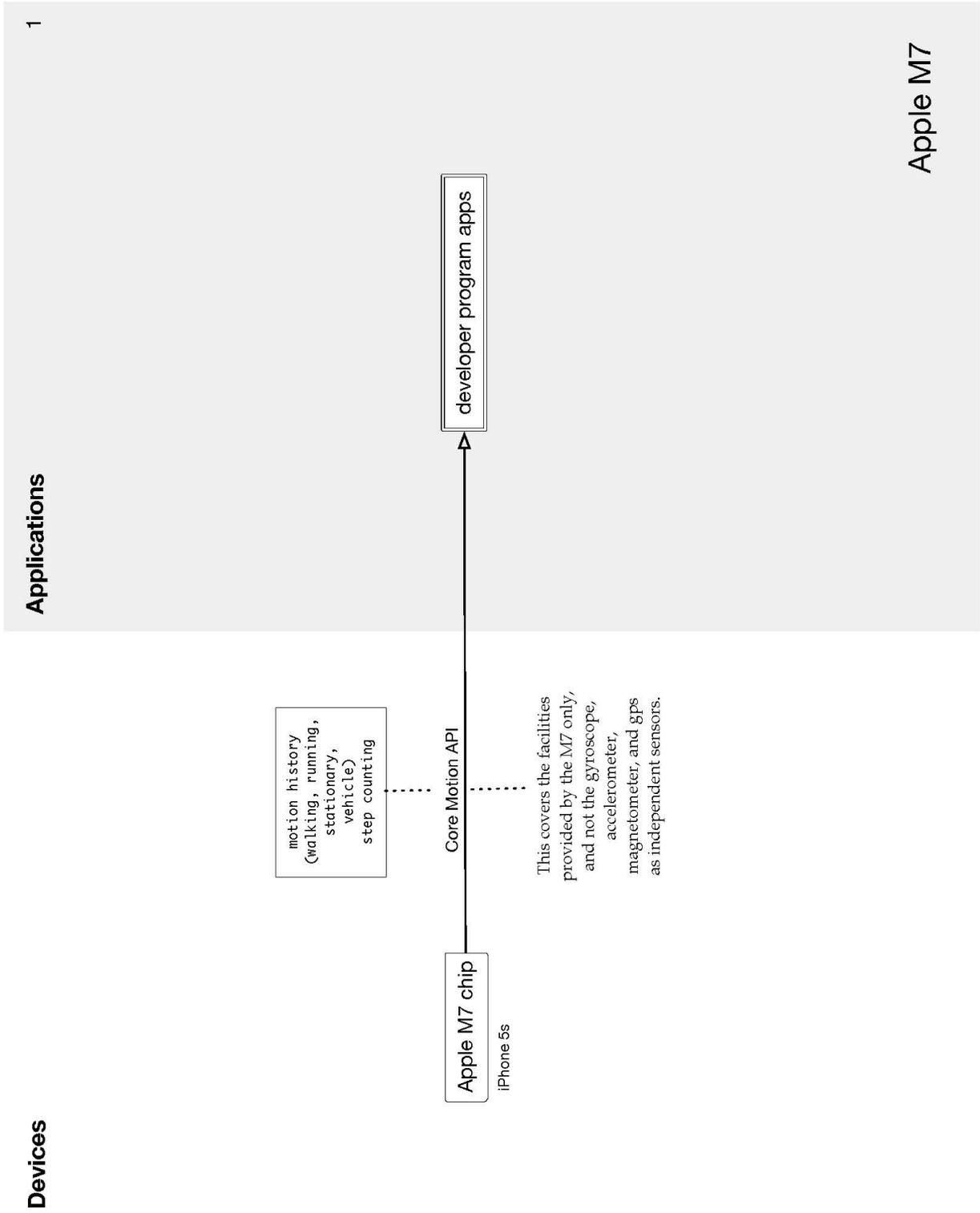
⁴⁴ <http://runkeeper.com/apps/sleep-cycle/>¹¹²⁴³⁹

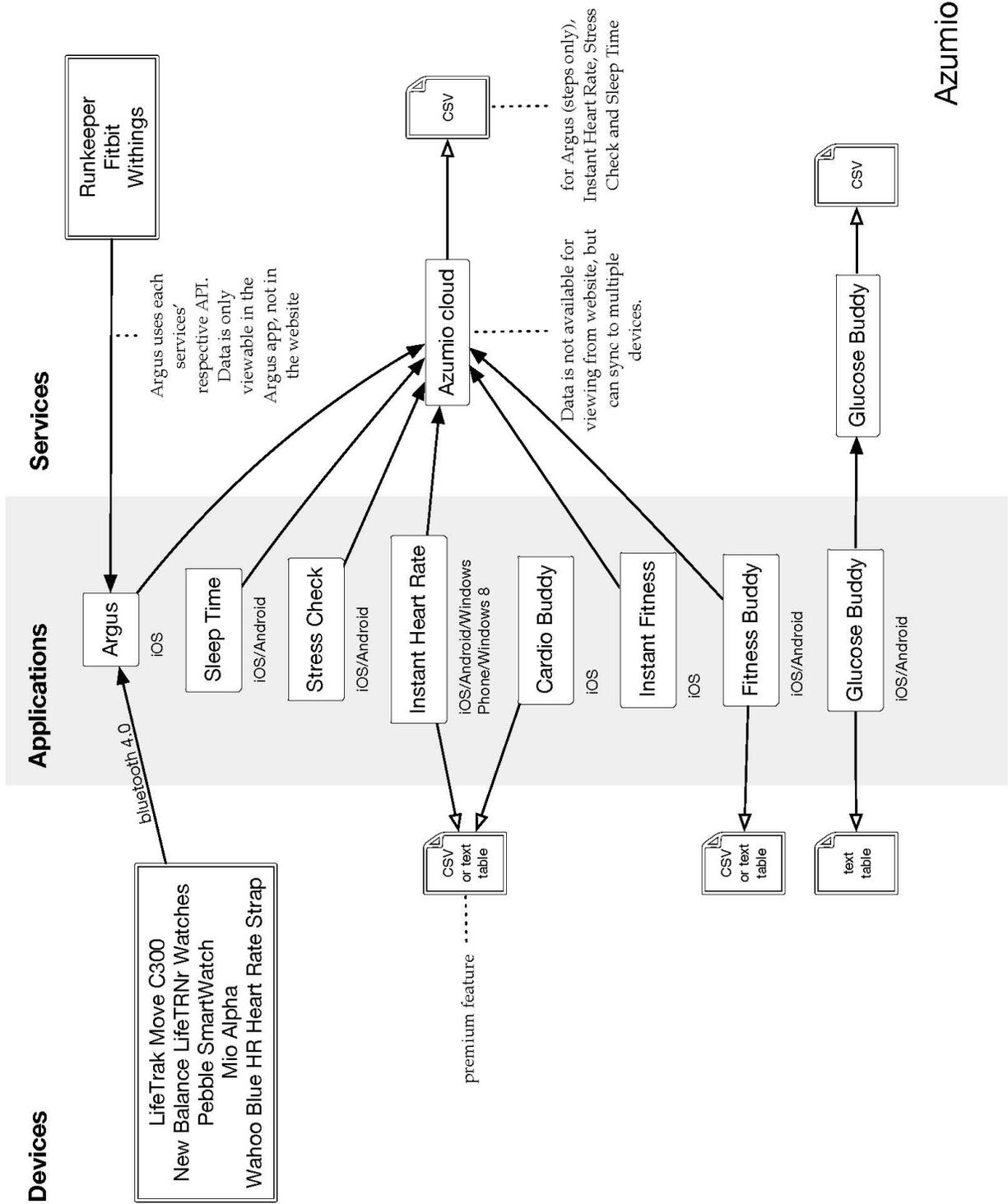
Lability: The reported detail in these diagrams is a snapshot of a changing state. While some toolmakers have formal notification procedures in place to help 3rd parties understand when data access procedures or data types available have changed, many do not. The lability of data flows in self-tracking ecosystems is itself a barrier to research progress, but one that is unlikely to diminish.

⁴⁵ Twitter is often praised by QS toolmakers for exemplary API practices. Twitter specifically recommends the following methods for tracking changes to the API:

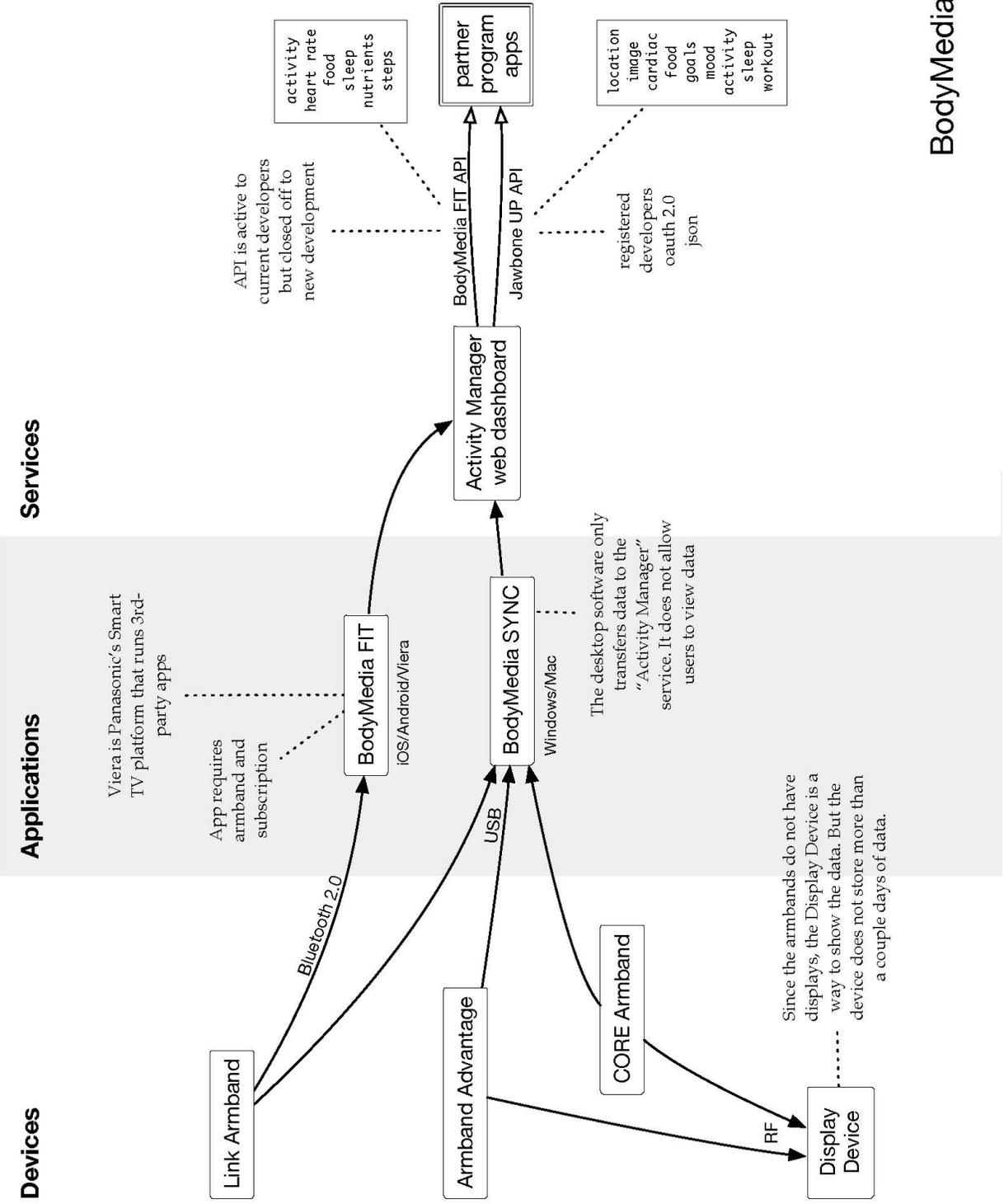
- Follow @twitterapi.
- Keep track of our Developer Blog and Discussions.
- See the recently updated documentation.
- Consult the Platform Calendar.
- Notice the lability implied in this advice.

Diagrams

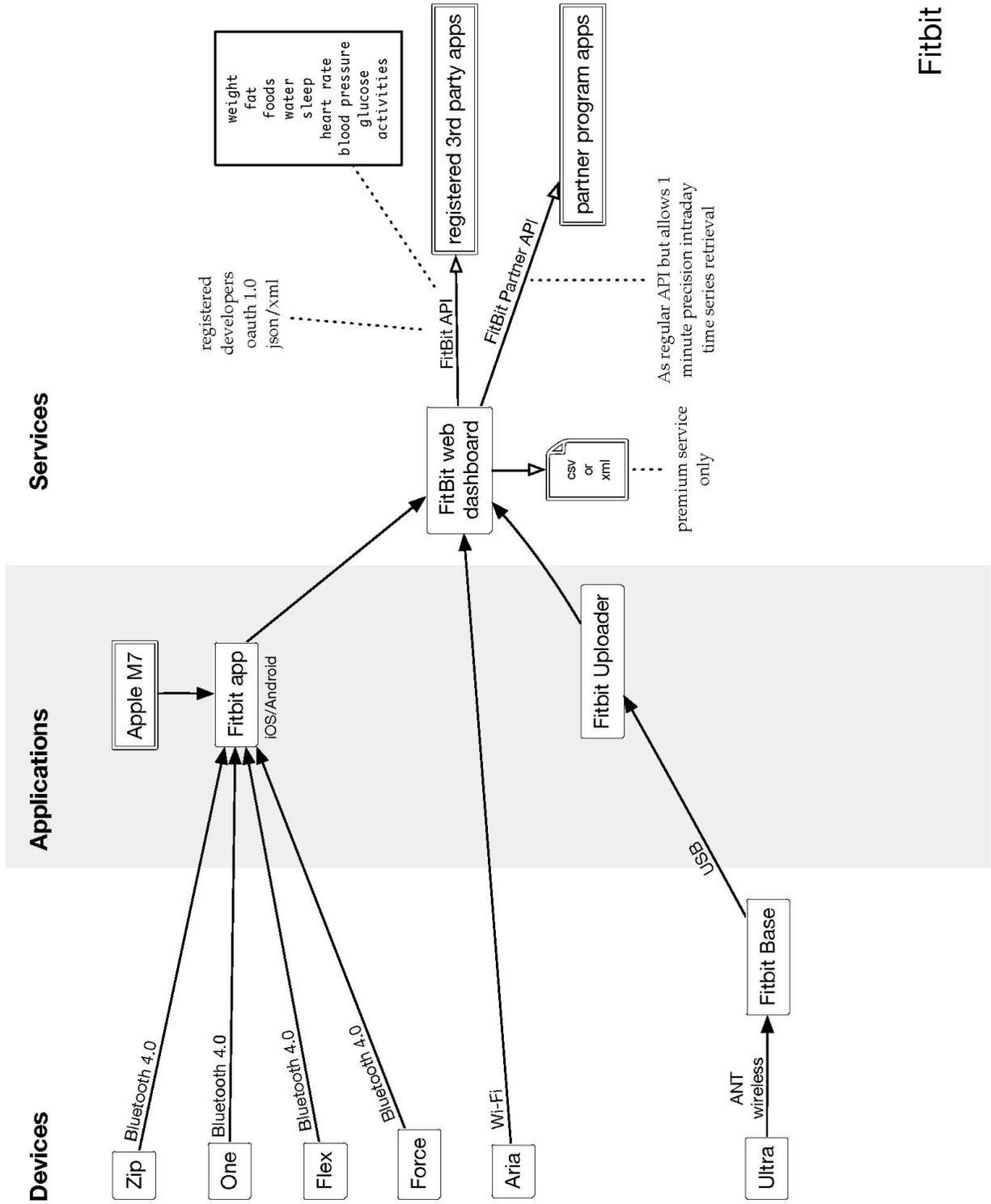




Azumio



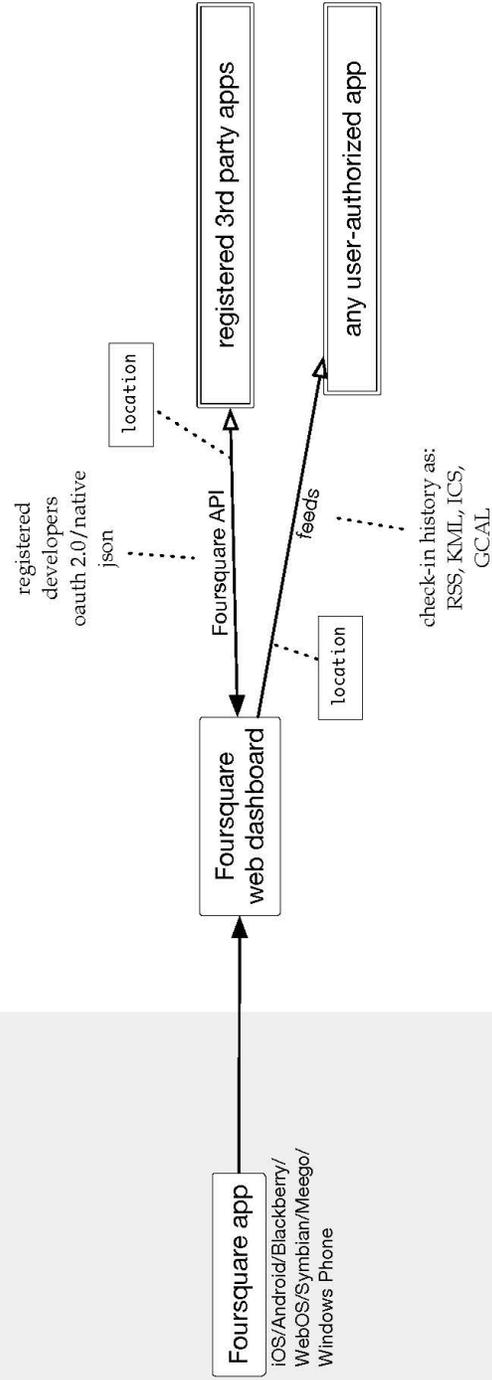
BodyMedia



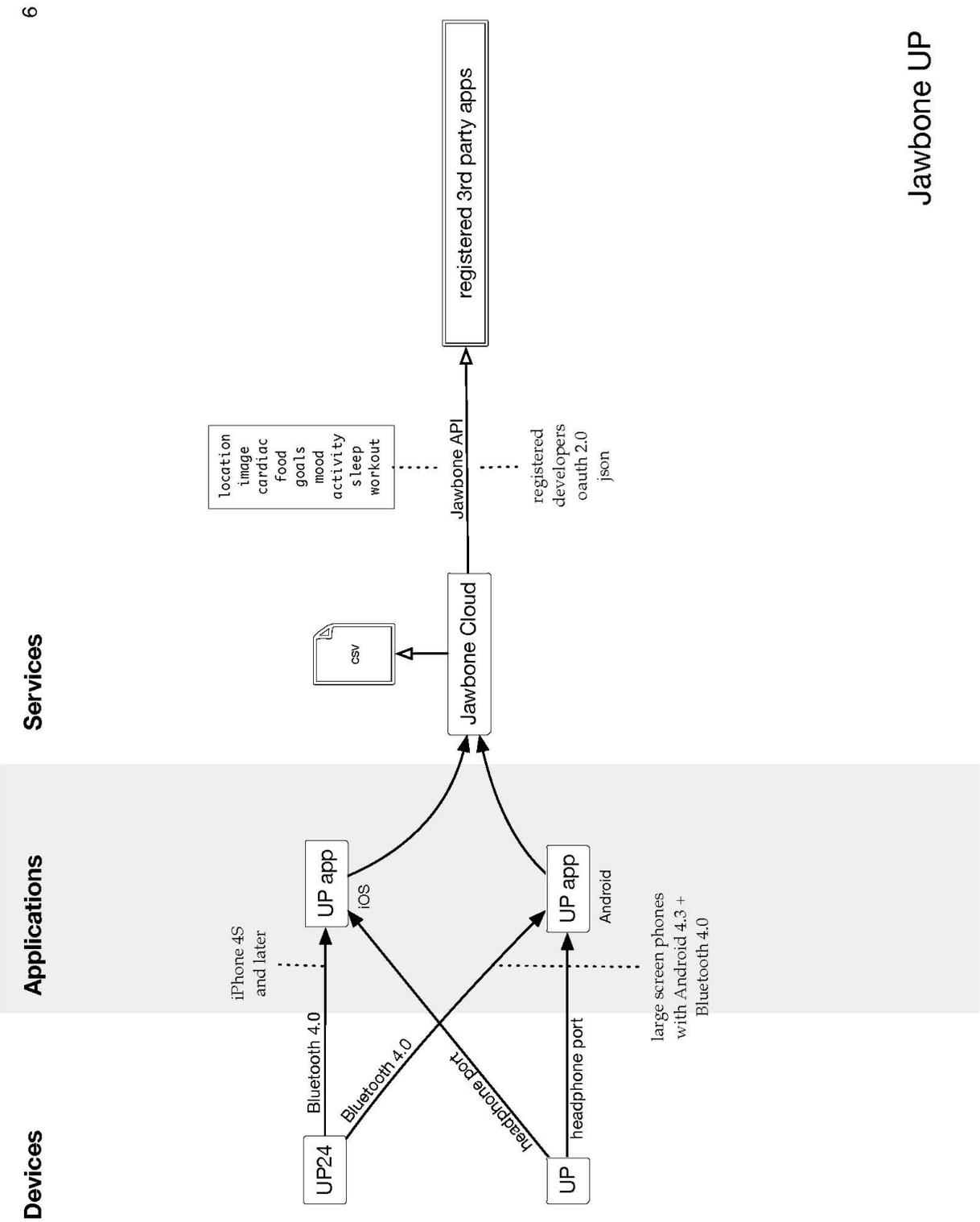
Fitbit

Services

Applications



Foursquare

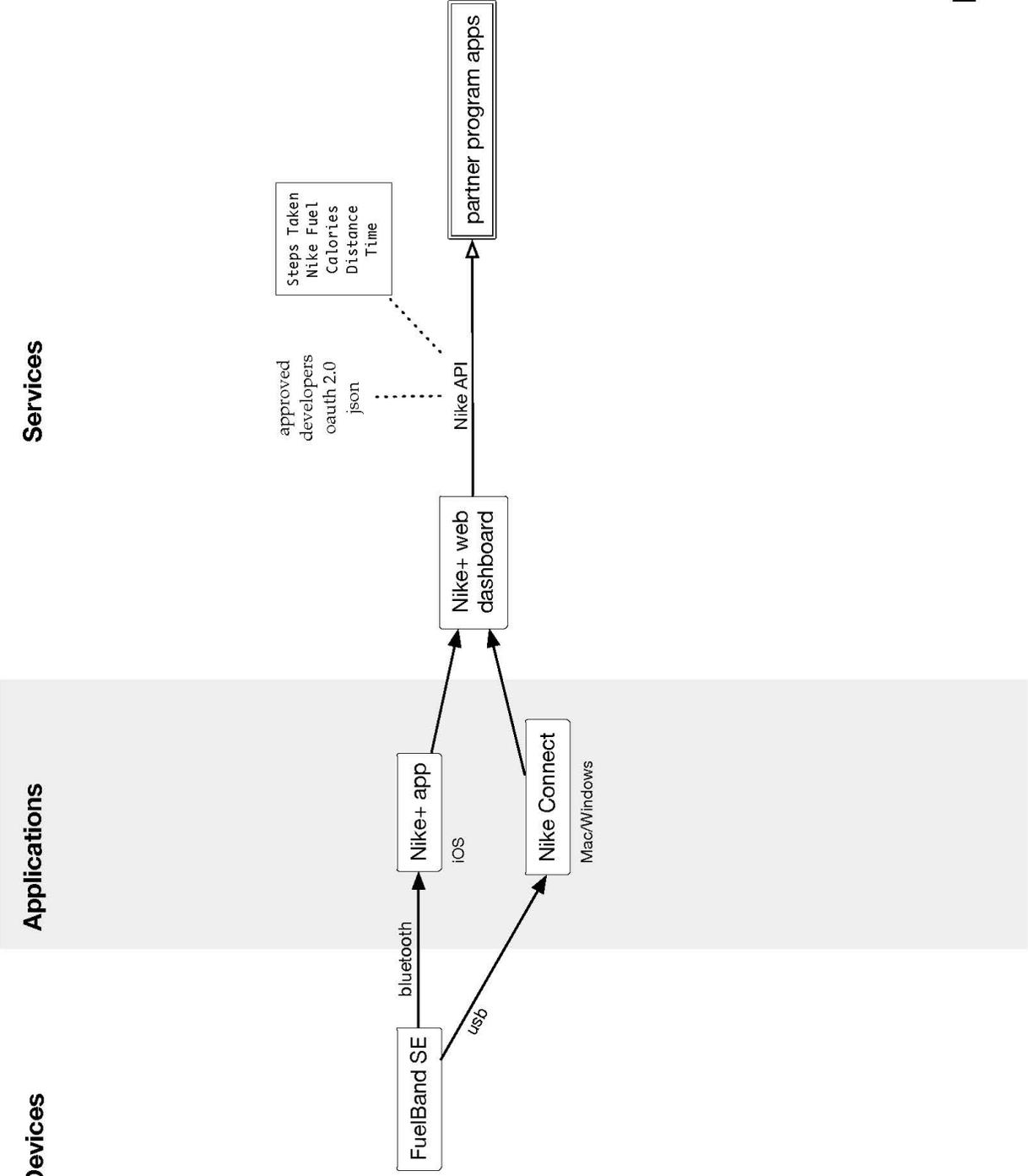


Jawbone UP

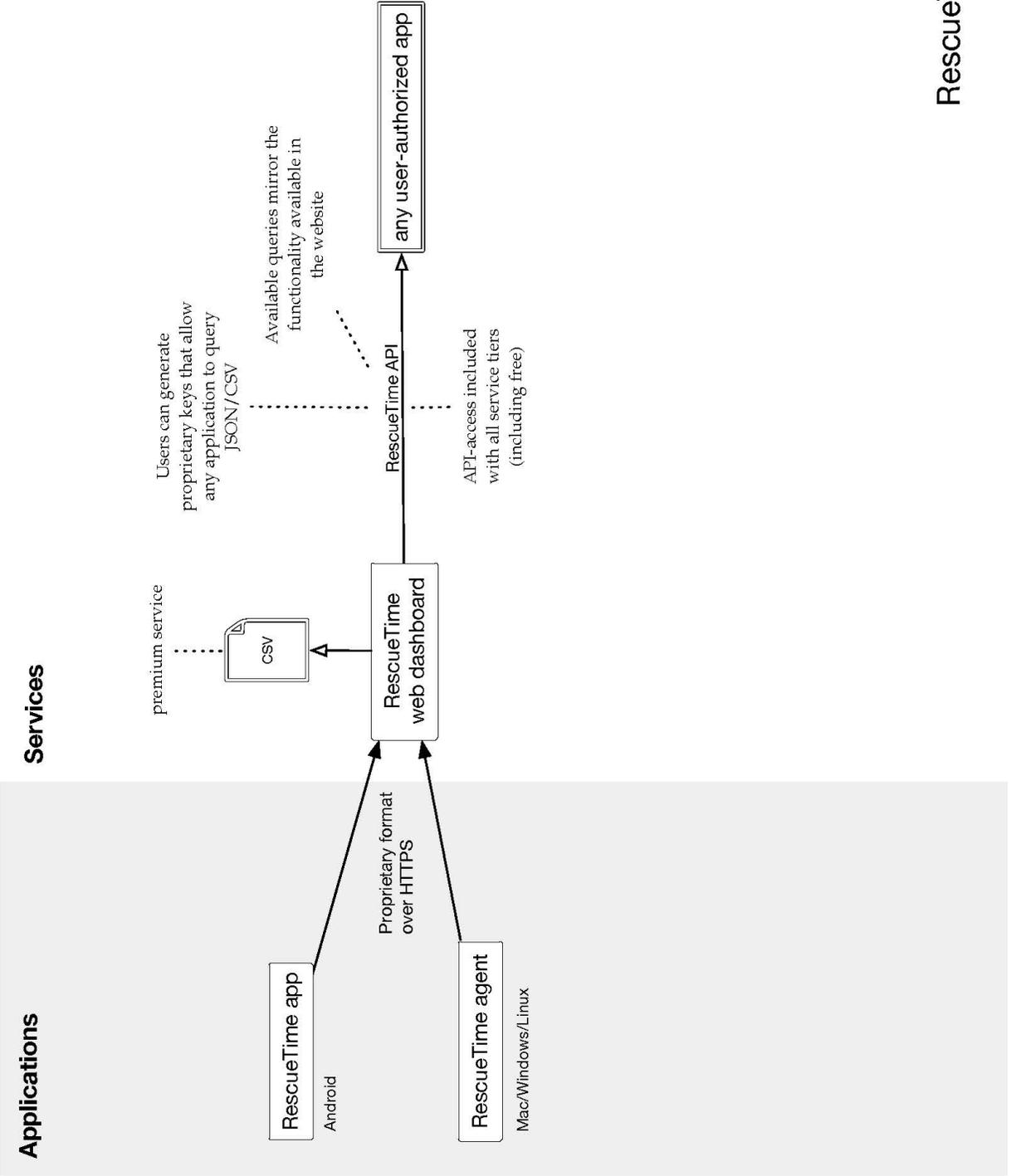
Services

Applications

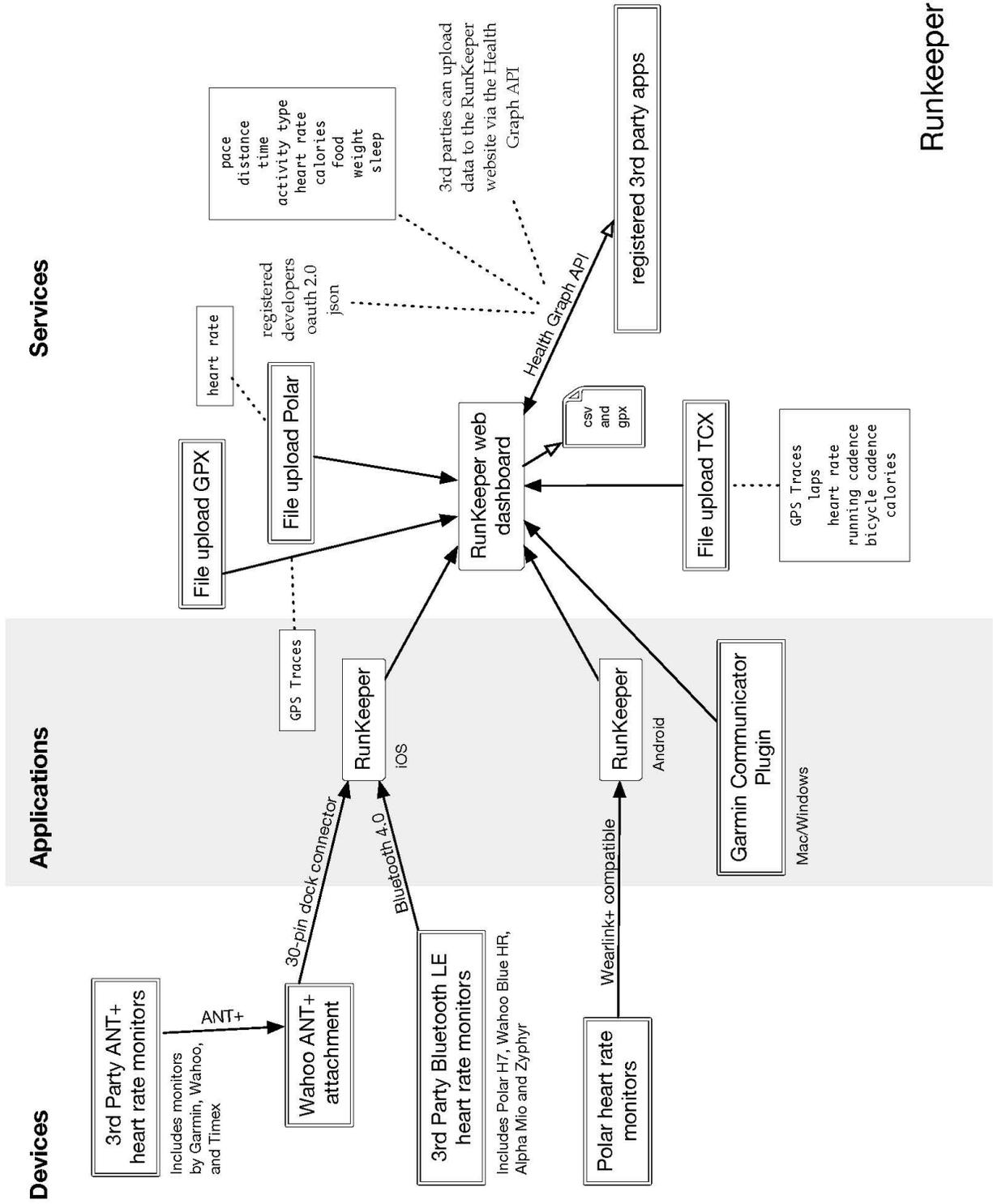
Devices

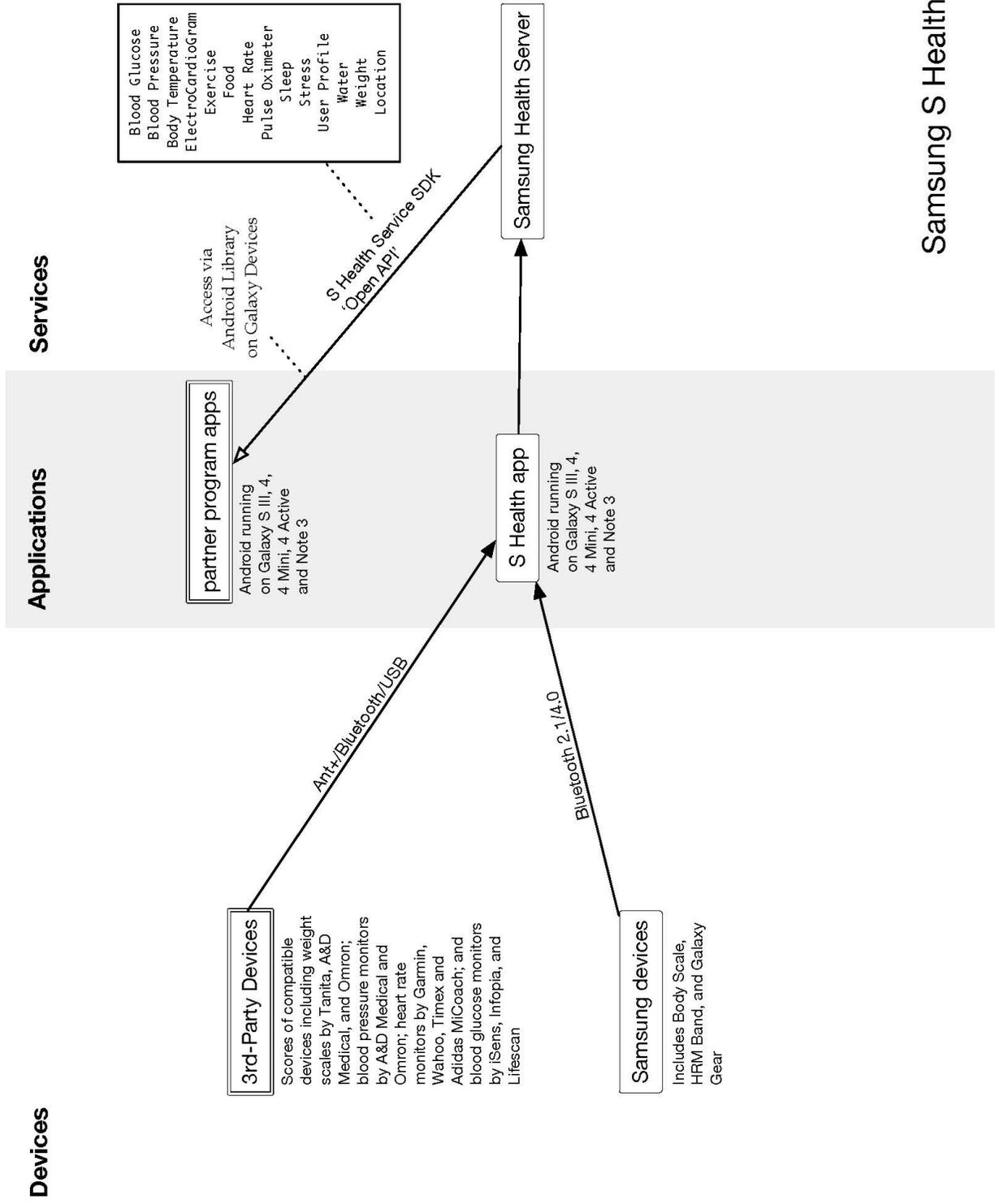


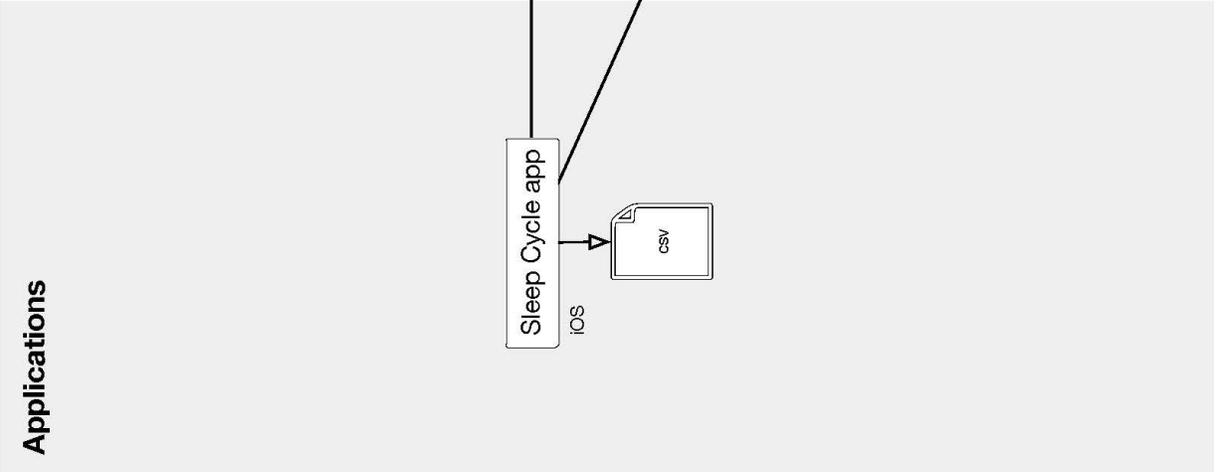
Nike



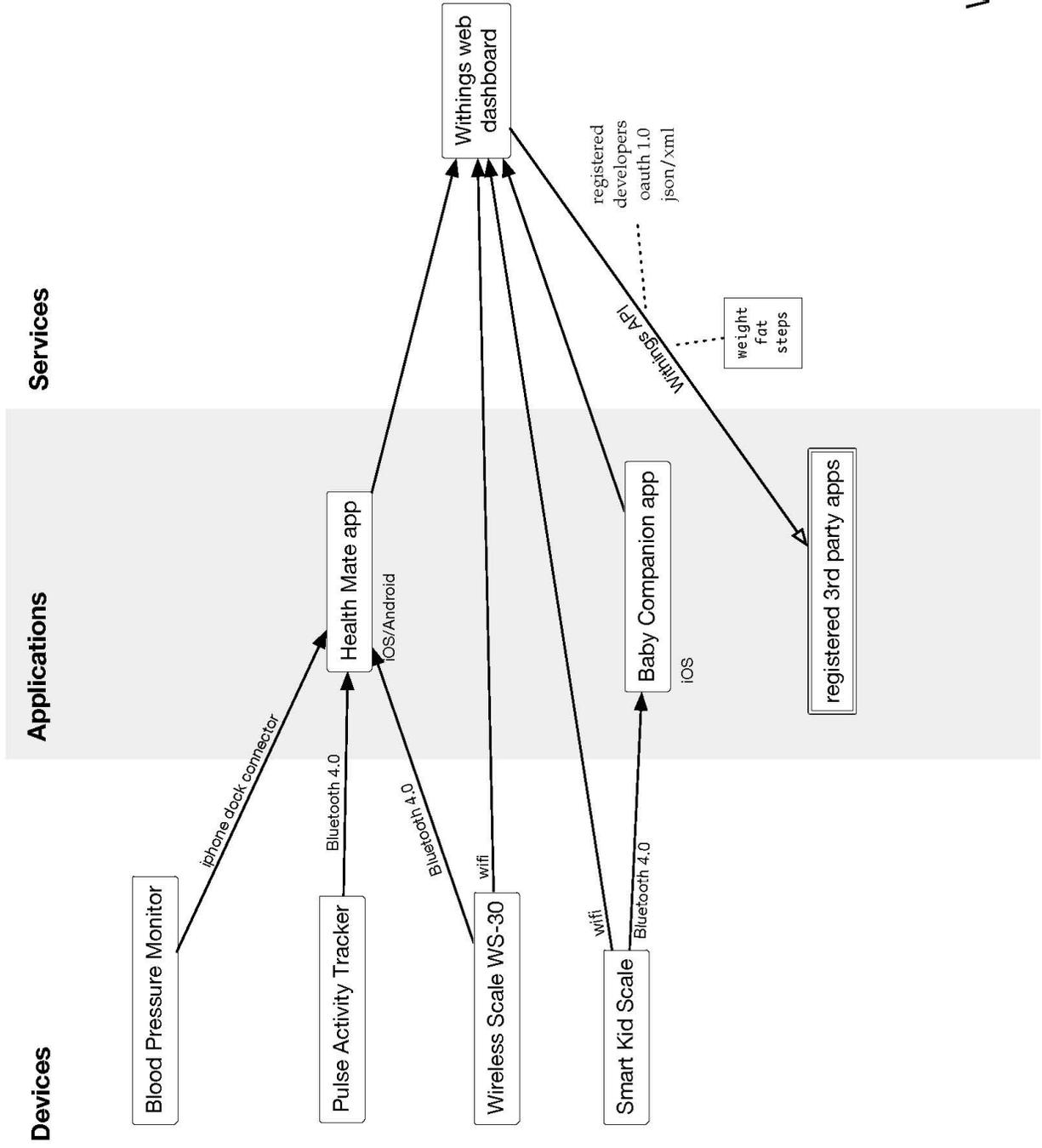
RescueTime







Sleep Cycle



Withings