Designing the Personal Data Stream: Enabling Participatory Privacy in Mobile Personal Sensing

Katie Shilton, Jeffrey A. Burke, Deborah Estrin, Ramesh Govindan,* Mark Hansen, Jerry Kang, [+] and Min Mun

Center for Embedded Networked Sensing
University of California, Los Angeles
kshilton@ucla.edu, jburke@remap.ucla.edu, destrin@cs.ucla.edu, cocteau@stat.ucla.edu, bobbymun@cs.ucla.edu

*University of Southern California
ramesh@usc.edu

[+]UCLA School of Law
kang@law.ucla.edu

Abstract....1
I. Introduction....1
II. Expanding Codes of Fair Information Practice....2
Understanding privacy as a participatory process....3
Data types....4
III. Principles to Enable Participatory Privacy Regulation....4
Design for primacy of participants....4
An architecture for primacy: the Personal Data Stream....5
Design for data legibility....6
Design for longitudinal engagement....7
IV. Social and Policy Changes....8
Public discussion and debate....9
Transparency of services....9
Legal privilege for raw location data....9
Participatory privacy regulation in practice....10
V. Future Work....10
Documenting participant satisfaction....11
Documenting participant engagement....11
Considering cognitive outcomes....12
VI. Conclusion....12
VII. Acknowledgments....12
VIII. References....12

Abstract

For decades, the Codes of Fair Information Practice have served as a model for data privacy, protecting personal information collected by governments and corporations. But professional data management standards such as the Codes of Fair Information Practice do not take into account a world of distributed data collection, nor the realities of data mining and easy, almost uncontrolled, dissemination. Emerging models of information gathering create an environment where recording devices, deployed by individuals rather than organizations, disrupt expected flows of information in both public and private spaces. We suggest expanding the Codes of Fair Information Practice to protect privacy in this new data reality. An adapted understanding of the Codes of Fair Information Practice can promote individuals' engagement with their own data, and apply not only to governments and corporations, but software developers creating the data collection programs of the 21st century. To support user participation in regulating sharing and disclosure, we discuss three foundational design principles: *primacy* of participants, data *legibility*, and *engagement* of participants throughout the data life cycle. We also discuss social changes that will need to accompany these design principles, including engagement of groups and appeal to the public sphere, increasing transparency of services through voluntary or regulated labeling, and securing a legal privilege for raw location data.

I. Introduction

For decades, the Codes of Fair Information Practice have served as a model for data privacy, protecting personal information collected by governments and corporations (Waldo, Lin, & Millett, 2007). The principles originally set forward by the U.S. Department of Health, Education and Welfare in 1973 demand notice of data collection, choice and consent, access for data subjects, integrity and security, and enforcement and redress (U.S. Department of Health, Education, and Welfare, 1973). Governments and corporations through the United States and Europe have widely adopted and adapted these principles for privacy protection over the ensuing 35 years. New attempts to deal with the proliferation of mobile and web-based data often rely on finessing these principles and clarifying authority for data protection among corporate and government bodies (Chenok, 2009; Clarke, 2000; Connolly, 2008; Information Security Awareness Forum, 2009).

But professional data management standards such as the Codes of Fair Information Practice do not translate well to a world of distributed data collection, nor the realities of data mining and easy dispersal. Consider the case of *mobile personal sensing*: an emerging form of distributed personal data collection (Eagle, 2008; Eisenman et al., 2006; Goldman et al., 2009). Mobile personal sensing harnesses mobile phone capabilities, such as location awareness, image capture, motion sensitivity, and user input, for personalized discovery and community exploration. Current design initiatives enable individuals to use their phones to collect and respond to personal data about their habits, routines, and environment. Networks of phones could become technological platforms for individual analysis, medical research, or advocacy, helping a community make a case through distributed documentation of a problem or need.

To illustrate the data collection possibilities and challenges of mobile sensing, we use the example of the Personal Environmental Impact Report (PEIR), a mobile sensing application developed by our research center that tracks location data to give individuals feedback on their daily interaction with their environment.[1] Using GPS and cell towers, a small piece of software running on users' phones records and uploads their location every few seconds. Based on these time-location

[1] See http://peir.cens.ucla.edu/ to learn more about PEIR and experiment with demonstration data.

traces, the PEIR system can infer participant activities (walking, biking, driving, riding the bus) throughout the day. The system maps the combination of location, time, and activity to regional air quality data and weather data to estimate personal carbon footprint and exposure to particulate matter. In PEIR, recording a participant's location throughout the day enables more accurate and previously unavailable information about environmental harms a person faces, and the harms she creates (Min Y. Mun et al., 2009).

Individuals and groups harnessing phones as sensing systems pose a number of challenges to information privacy. While mobile sensing enables new forms of participatory research and discovery, such systems also create dispersed and massive databases of individuals' locations, movements, images, sound clips, text annotations, and even health data. Sometimes these data are collected by institutions (mobile carriers, health care providers, insurers, Google) following Codes of Fair Information Practice. But increasingly, this data is shared among networks of amateur enthusiasts.[2] In addition, mobile sensing data may be quite granular (e.g. thousands of GPS points or accelerometer readings), making these data more difficult for an individual to comprehend than traditional personal information such as addresses and social security numbers. Largely quantitative, these data are conversely easier for a machine to mine than the personal data the Codes of Fair Information Practice and their descendants were meant to address (Clarke, 2000).

We begin this paper with an argument for how new principles can expand the Codes of Fair Information Practice to support privacy in cases of distributed, granular data sharing. We support this argument with a new architecture for mobile sensing: the Personal Data Stream. We have designed this architecture to restructure how data are stored and shared in mobile sensing, allowing for privacy principles to exist even in a distributed sharing environment. We enumerate these alternative privacy principles, as well as complementary policy and social changes needed to make this vision successful. We close with proposed next steps to evaluate the outcomes of our privacy principles and the supporting Personal Data Stream architecture.

II. Expanding Codes of Fair Information Practice

The Codes of Fair Information Practice offer a firm ethical and technical starting point for dealing with the personal data collected by mobile personal sensing. Participants in mobile sensing certainly deserve notice and awareness, choice and consent, access and participation, integrity and security, and enforcement and redress. But the distribution of these data beyond governments and large corporations complicates application of fair information principles. The Codes assume organizations to be the data collectors, and individuals to be the data subjects. This may not be the case in many mobile sensing applications. Consider the individual tracking his own weight loss statistics using an online application, or the community group that bands together to document pollution released by a chemical plant in the neighborhood. Individuals may intend their data collection for their own purposes and use. The data collected by community groups might be cooperatively analyzed and widely shared. By enabling dispersed data collection and sharing, mobile sensing collapses the role of data collectors and data subjects. Fair data practices begin to lose their coherency when the roles of data subjects and collectors become blurred. In these examples, which parties are responsible for ensuring notice, access, security and redress?

One example of the blurring of institutions and individuals is the environmental application Ecorio, (http://www.ecorio.org/ecorio.htm). Much like PEIR, Ecorio collects activity and location

[2] Examples of amateur and enthusiast data analysis and sharing are diverse. A sampling include "My Tracks," http://mytracks.appspot.com/, Daytum, http://daytum.com/, Mycrocosm, http://mycro.media.mit.edu/, Moodstats, http://www.moodstats.com/, and Your.Flowingdata, http://your.flowingdata.com/.

data from users. The creators of this application describe themselves as "five guys from Ontario." Ecorio offers little information as to what privacy measures it takes, or whether it adheres to professional data management standards such as the Codes of Fair Information Practice. This may be the result of disinterest in, or ignorance of, fair information practices. But it also could result from a lack of resources to devote to professional data management practices. The dispersed development of mobile sensing is an example of the deprofessionalization of information collection and protection (Braman, 2006) and the challenge that Zittrain (2008) refers to as "Privacy 2.0." Data management codes developed for organizations, dependent upon best-practice security, and supervised by privacy officers become tenuous in a future where single developers or small teams create prolific data sharing applications.

We define privacy as the process of controlling the flow of personal information (Westin, 1970). In situations where individuals cannot be certain if their data are subject to Codes of Fair Information Practice, they lose control. Data they collect themselves, or data collected by small-scale applications, may not be protected by voluntary requirements for notice, consent, access, security and redress. Acquaintances, friends, employers or authorities might coerce disclosure of these sensitive data. Could a spouse or a boss compel an individual to turn over her mobile sensing data? Could these data be subpoenaed in a traffic accident or civil case? Even voluntary sharing can have unseen repercussions. Data shared with an acquaintance might reveal minor indiscretions, exposing little white lies about plans or social obligations. Or databases of locations and routines could be used to further segment and sort consumers, encouraging controversial forms of economic and social discrimination based on new demographic categories (Curry, Phillips, & Regan, 2004). At the extreme, mobile phones could collect data without participant consent, becoming the most widespread, embedded surveillance system in history. Consider the 2007 case of a New York City employee, fired after his boss tracked his GPS traces recorded by his work phone without his knowledge (Seifman, 2007). A second example: a pending case in the 3rd Circuit U.S. Court of Appeals questions whether the government can request phone location information from mobile providers without a warrant (Freiwald & Swire, 2009). Consumers who wish to use services like PEIR and Ecorio need a new set of privacy protections.

Understanding privacy as a participatory process

Loss of control over one's personal data is heightened in mobile sensing because recording devices embedded in phones disrupt expected flows of information in both public and private spaces (Nissenbaum, 2004). In increasingly mediated environments, scholars such as Palen and Dourish (2003) suggest that privacy negotiation becomes a dynamic and ongoing process that relies heavily on user engagement with data and ongoing sense-making: which data am I sharing now, with whom, and what do they say about me? In mobile sensing, understanding privacy as engagement and control of data specifies that privacy decisions take place throughout the sensing process, from deciding to turn on a sensor to making post-facto decisions to delete data (Shilton, Burke, Estrin, Hansen, & Srivastava, 2008). Mobile sensing participants invested in their data will have reason to explore and make privacy decisions. PEIR participants, for example, can alter their carbon impact by varying their commuting routes, giving them reason to observe, play with, and interpret their data. We call this principle *participant primacy*. Participants should also be able to understand what the data mean and reveal about them. PEIR participants sharing commute data with coworkers might be concerned that the data document the time of day they enter and leave the workplace. We call this principle *data legibility*. Finally, participants should have the ability to make and revoke data sharing and withholding decisions over time, as the context of their privacy needs change. A PEIR participant may decide to stop tracking her location traces when a regular stop for medical treatment

becomes part of her commute. We call this principle *longitudinal engagement*. Each principle has roots in an individual right to manage one's image and identity, and consequently the data that increasingly are part of that identity (Phillips, 2005).

The Codes of Fair Information Practice do not support these principles well, because they do not adequately promote data subjects' engagement with their own data. Notice is not enough to spark investment. Access is not enough to promote understanding. And redress is not enough to support long-term changes in context and subsequent privacy needs. We propose that participant primacy, data legibility and longitudinal engagement can expand the Codes of Fair Information Practice to support systems that improve users' ability to make sense of, and regulate decisions to share or withhold, data. Individual decision-making about privacy is a complicated challenge, because individuals often suffer from incomplete information about data collection and limited understanding of consequences (Acquisti & Grossklags, 2008). But asymmetries in information and understanding can be at least partially eliminated by design choices that help illuminate data and support user decision-making. In this way, mobile personal sensing can support *participatory privacy regulation*: a privacy paradigm in which individuals gain back the ability to control their own data (Shilton et al., 2008). In a participatory approach to privacy, data management, discretion and sharing are integral parts of participation in sensing projects. To facilitate this engagement, we describe an architecture for enabling user decision-making about data collection and sharing, and the principles that undergird that architecture.

Data types

Mobile personal sensing is not a fixed set of practices, and it may be impossible to plan for all of the diverse data types such research might capture. But our experience designing applications and architectures for mobile sensing, and our interest in privacy issues, has led us to concentrate on three canonical, but not exclusive, types of data. These data types are location traces (created through tracking GPS points or cell tower triangulation over time), geotagged images, and geotagged text contributed by users. We concentrate on these three data types for several reasons. First, they are useful in current mobile sensing applications such as PEIR. Second, these emerging databases of geotagged data will likely grow increasingly popular and prevalent.[3] Third, these types of data, when linked together or shared in large quantities, reveal individuals' identity, routines, and preferences (Barkhuus & Dey, 2003; Curry et al., 2004; Kang & Cuff, 2005).

III. Principles to Enable Participatory Privacy Regulation

Design for primacy of participants

We encourage the professional and amateur developers building location-aware applications to take a stronger stance than consent to data collection. Mobile sensing systems should enable participants to retain control over their raw data. Participants own their raw location data and any annotations to that data (photos, sound clips, co-location data, etc.) and are responsible for making and revoking decisions to share subsets of the data.[4] Framed this way, participants are not just

[3] Telecommunications providers increasingly have such databases, but the data is protected (with varying efficacy) by both U.S. law and the proprietary interests of the companies that have gathered the data (Mitrou, 2008). This data is, so far, not widely sharable or accessible, unlike mobile sensing data.

[4] Ownership of data has both legal and psychological dimensions. Although we support legal ownership of personal data, this is an emerging and unsettled policy question that we address in other work. Instead, we use "ownership" here in its psychological sense, to refer to the investment felt by individuals when they store, manage, and control their data. Participant primacy is meant to separate control of personal data from the interest of application developers.

subjects of data collection, but take the role of investigators (when they use self-analytic services) or co-investigators (when they contribute their data to larger research initiatives). This is the understanding of data collection developed and tested by Community-Based Participatory Research (CBPR) (Cargo & Mercer, 2008; Shallwani & Mohammed, 2007). CBPR successes in health and environmental research have not only increased the validity of research data, but also improved the ability of marginalized or underserved groups to act on the results of the data they have helped collect and analyze. CBPR traditions provide a model for distributed, bottom-up mobile sensing research. CBPR applied to mobile sensing gives participants an interest in how data are collected, processed, stored, and discarded. When users care about their data, and see them as theirs to manage and distribute, they redefine their relationship with data. Research in behavioral economics suggests that ownership may be a powerful motivator for individuals to engage in privacy decision-making (Acquisti & Grossklags, 2008). Data were once a commodity collected by corporations and governments. Now they are also an asset collected by individuals, community groups, or cooperative research projects. Like credit records, resumes, or social networking profiles, individuals who feel responsible for ownership of their data will have incentive to periodically review and manage those data.

An architecture for primacy: the Personal Data Stream

A new architecture for data collection and sharing can help to encourage participant primacy in mobile sensing. To facilitate ownership of data, individual participants need some of the same tools for storage and access control currently used by corporate data collectors. We have designed the Personal Data Stream (PDS) to give users new data management tools and enable them to cope with data privacy. In the current instantiation of PEIR (and many other location-aware applications), location data flow directly from a user's phone to the application servers. If users do not wish to share data with PEIR, they must remember to turn off the application on their phones, giving users only coarse and error-prone control of data capture. A Personal Data Stream-enabled PEIR will put more control of this sensitive data back in the hands of participants.

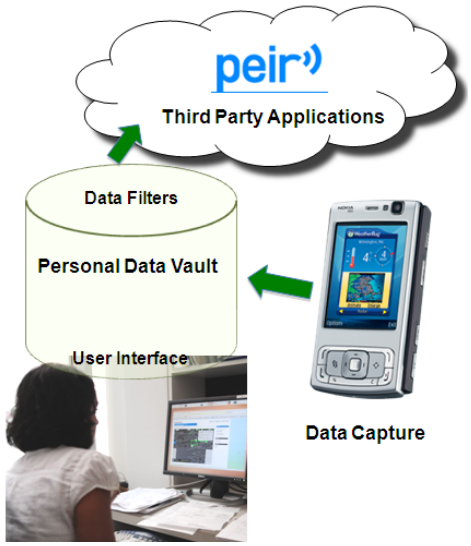

Figure 1: The PDS imposes a data management tool between capture and third-party analysis. *Photo credit: Cat Deakins.*

The Personal Data Stream architecture consists of four parts. Specialized data collection software runs on mobile phones. It communicates with a Personal Data Vault (PDV), which is inserted between the user's phone and third party applications. A set of filters in turn controls flow of the data from the vault to third party applications. Finally, third party application software accepts data from the vault, and communicates information such as data requirements and usage logs back to the vault.

The vault lies at the center of the Personal Data Stream. The vault is a system of specialized software and decentralized hardware that provides storage, authentication, access control mechanisms, and a user interface. It resides between a user and an application, to give individuals an initial set of controls over when, and with whom, their location traces are shared. All data uploaded from a participant's phone first enter their vault. Functioning like a bank account or personal file storage, the vault becomes the place to check on and make decisions about data. Vaults could also work together or within an online social network to allow data sharing and joint analysis among

informal groups (Hao et al., 2009). A number of research labs and commercial developers are building data gathering architectures that rely on a vault-like entity to provide trusted storage and processing. Complimentary designs include Stanford's PRPL architecture (Lam, 2009) and AT&T Lab's virtual individual servers (Cáceres, Cox, Lim, Shakimov, & Varshavsky, 2009), both of which are designed to support privacy in mobile sensing. The vault also has similarities to commercial developments for personal health records, such as Microsoft's HealthVault ("Microsoft HealthVault," n.d.) and Google Health ("Google Health," n.d.).

One set of vault-enabled decisions focus on whether, and with whom, to share data. Though the vault does not prevent the privacy problems associated with transmission or reuse by third parties, this new architecture does give users a first cut at managing data sharing. The vault provides access control, allowing users to select and limit who can see what kinds of data. The PDV could default to keeping all data private as they arrive. Access control tables allow users to adjust these default policies, setting new sharing policies for particular third parties (individuals, groups, or applications such as PEIR). Users might also set policies according to time (e.g. only send PEIR data from 9am-5pm), geography (send only zipcode level data to PEIR), activity (share only driving routes with PEIR), or data type (share only geotagged photos with PEIR) (Hao et al., 2009). By giving users a variety of options for sharing their data, the PDV gives participants the ability to avoid total accountability for their locations and actions. For example, PEIR participants might share location data to take part in an office carbon reduction challenge. But to protect the details of their working hours, employees might send location traces stripped of time of day from their vaults to their employer.

Beyond basic access control, the PDV could incorporate customized filters to help users manage the logistical burdens of selective sharing. For example, an individual may wish to share only her usual routine with the (less secure) PEIR service. Filters within her PDV might detect anomalies in her routine (an unusual route, or an unprecedented trip late at night) and contact the user to ask if she wants to share the deviant trip with PEIR. The PDV could further protect data by incorporating location privacy measures. For applications that need less granular or accurate data than a GPS trace provides, the PDV could send coarser data, such as cell tower locations (Min Y. Mun, Estrin, Burke, & Hansen, 2008). The vault might take an even greater role in protecting individual's data by supporting processing of the most sensitive data inside the vault. Instead of exporting a daily location trace to PEIR, for example, the PDV might perform the transformations PEIR requires directly on a user's data. The PDV could then export only the results of those calculations, avoiding the security and privacy risks of sending location data to a third party application. Through authentication, access control, and privacy processing, the PDV is designed to keep users informed of, and in control over, who is using their data. It also minimizes the amount of data released to less-secure third party applications. Through these measures, the PDV will enable users to maintain a sense of ownership and control over their data.

Design for data legibility

Personal sensing systems can help participants make sense of, and decisions about, their data by visualizing granular, copious data in ways individuals can understand. Methods to improve data legibility include visualization tools such as maps, charts, icons, pictures, or scales. Data legibility also includes showing users who has accessed their data and how frequently, and showing participants where their data go and how long they remain accessible. Such system features increase participants' understanding of complex risk and help them make better decisions about data capture, sharing, and retention. Legibility is a stronger interpretation of data access, and can fortify participants to be better data stewards.

Data visualization techniques have a rich history and are an ongoing topic of research and innovation in statistics, computer science, and design. Both PDVs and third-party applications will need to incorporate creative visualizations to help people understand their long-term data store. Natural language could also be a key to helping people understand what their data say about them. For instance, in the PEIR project, the algorithms that transform GPS data into impact and exposure numbers are translated into plain text, to help users understand how PEIR transforms their data (Figure 2).

The PDV's user interface will also play an important role in helping people understand and manage their data. A graphical interface that allows users to drag and drop data sets into specific privacy categories might ease the logistical burden of selecting new policies for every data set. In addition, the interface can not only help users set policies, but help them see the results of those policies. Illustrating who can see what data will go a long way towards helping users understand the consequences of data sharing.

A more detailed way to help users interpret their data and make privacy decisions would be to include a set of inference agents within the Personal Data Vault. Basic agents might keep track of who had access to what data from the vault, and give the user easy-to-read reports about what each friend or third-party application learns about her. An advanced agent could scour the web or third-party applications for personal data about an individual, and make a series of inferences from that data. By checking on the inferences the agent is able to make, individuals would have an enhanced idea of what the data they share reveal about them. Algorithms to support such inferences are still under development, and remain a challenge for PDV developers (Hao et al., 2009).

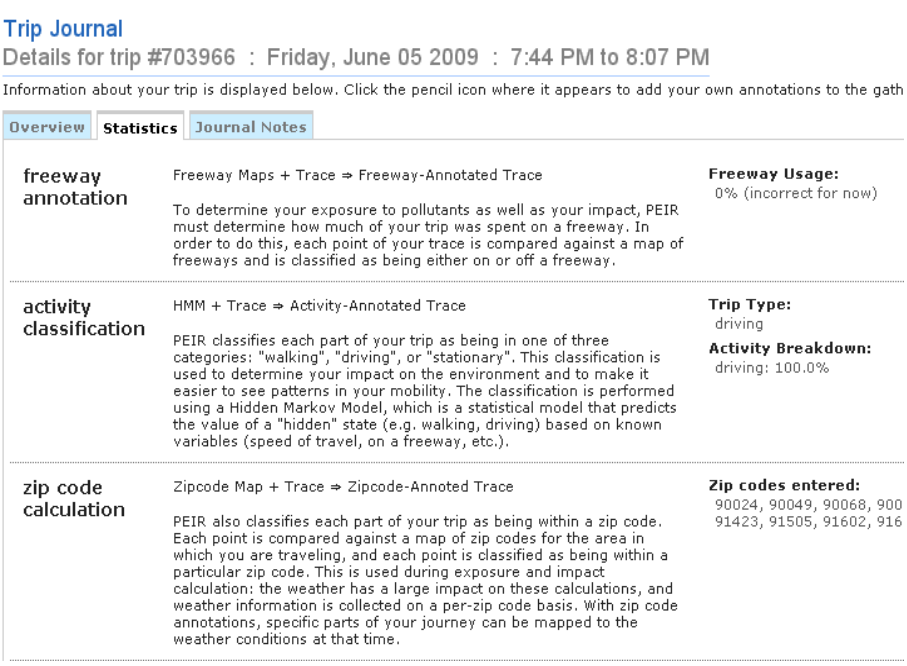

Figure 2: Natural language explanations of algorithmic processes

Design for longitudinal engagement

The interfaces of both the PDV and mobile sensing applications should encourage participants to engage with their data from the point of collection through analysis, long-term retention, or deletion. Privacy decisions about sharing and retaining data are part of the sensing process, and can occur at many points in a sensing project. System features to encourage the continued engagement of participants can allow them to change their data practices as their context changes. Stronger than notice, the crux of engaging individuals with decisions about their data is

refusing to put data in a black box. Instead, collecting high-quality data, learning from the data, and making ongoing choices about the data throughout their lifecycle become the goals of sensing.

The Personal Data Stream architecture can encourage long-term interaction with data in a variety of ways. The mobile phone represents the capture point and therefore the beginning of the data life cycle. Data collection clients designed for the phone should therefore have clear signifiers of data capture, and give the user easy options for turning capture on and off (Bellotti & Sellen, 1993). Once data have reached the PDV, design features can encourage users to visit the vault regularly to check on their data. Features employed by commercial sites, such as weekly email updates with data sharing summaries, or monthly reminders (or requirements) that users review their sharing policies, might be useful to encourage regular data check-ups. And as users allow vault access to new friends and applications, or join groups that pool data, they will need to revisit and update their sharing policies, creating an additional incentive for longitudinal engagement.

Feedback has also been demonstrated as critical to privacy decision-making (Tsai et al., 2009). Users who receive feedback on who has viewed their data, and how frequently, can update their sharing decisions should they become uncomfortable with an individual's level of access. They can similarly make decisions and alter sharing if they see that an application has passed their data on to another party. PEIR users, for example, may see notifications when their zip code-level data are shared with regional Air Quality Management Districts to obtain pollution exposure estimates. In this way, auditing can provide users with accountability for how their data are used (Weitzner et al., 2008). The Personal Data Vault can host audit mechanisms to enable users to monitor and adjust data sharing as their interests and context changes. This could be as simple as logs that reflect what data the vault has sent to third party applications. A PDV inference agent could also check and report on the vault data an outside party already has access to, helping users make a more informed decision about sharing with that party.

A more complete form of feedback could take the form of a "TraceAudit": a trail of data use and sharing outside the vault, modeled after an internet traceroute ("Traceroute," 2009).[5] Such a mechanism would begin to address participants' lack of control once their data have left the Personal Data Vault (Krishnamurthy & Wills, 2009). This would require third party applications to log sharing or transformation of data back to the PDV, according to terms agreed upon when an individual (or the PDV on her behalf) contracts with the application. Audit mechanisms like the TraceAudit might go so far as to require accountability for data use from third party applications. A third party auditor or periodic certification of third parties is perhaps an even stricter implementation of accountability for longitudinal engagement.

IV. Social and Policy Changes

Designing for participant primacy, legibility, and longitudinal engagement will shape the Personal Data Stream architecture, but technical decisions will not be enough to ensure the success of participatory privacy regulation. Focus on participant investment and legible, changeable systems will help people make more informed decisions about withholding and sharing information. But informing participants is not enough to ensure sufficient data privacy. Individuals' privacy decisions are often less than rational (Acquisti & Grossklags, 2008). Mobile sensing participants may make disclosure decisions based on a host of structural and contextual factors, ranging from whether making privacy changes is easy and intuitive, to knowing what privacy decisions their friends and colleagues are making. To enlarge the privacy protection discussion beyond individual decision-making, supporting social structures can fortify participant engagement with the Personal Data

[5] The traceroute is a tool used to map the path of packets as they move through a network.

Stream and the protection of data privacy. These include public discussion and debate to improve data literacy, increasing the transparency of third party data practices, and securing a trade secrets-like legal privilege for raw sensing data.

Public discussion and debate

Developing a lay public understanding of mobile sensing privacy, security and risk is critical to a vision of participatory privacy regulation. But casual technology users often underestimate or misunderstand data sharing and security risks (Camp, Asgharpour, Liu, & Bloomington, 2007). Data literacy will become an increasingly important knowledge set as applications for collecting and sharing data proliferate.

Developers can engage traditional media, new media, educators, and civic groups to get citizens interested in, and talking about, mobile sensing. This will help to move discussion of privacy decision-making into the public sphere. Public discussion and debate of social issues engendered in mobile sensing technologies can fortify both individual understanding and democratic decision-making. It can also subject sensing systems to both academic and lay reflection and critique (Calhoun, 2000). This debate might take place in the popular media, or increasingly within online settings and communities of interest (Kelty, 2008). This is a space of debate, discourse, and consensus building, and makes privacy about more than just individual choice. The recent furor over Facebook's terms of use (and before that, the site's Beacon and news feed features) are examples of ways the public sphere can discuss, comment on, and affect privacy issues (Stone & Stelter, 2009). By hosting discussion boards on mobile sensing services such as PEIR, by commenting on blogs and joining debates in city councils, by reading and critiquing popular press accounts, both sensing designers and participants can contribute to discussion about protecting and sharing location data.

Transparency of services

Helping individuals understand the data practices of mobile sensing services to which they subscribe will help participants make better sharing decisions. A voluntary or regulated system of application labels would help participants understand levels of risk inherent in location-aware services. If an application has "best practice" data practices, it might be certified as a 'fair data' application. In much the same way that voluntary and regulated labels such as 'fair trade' and 'organic' increase the transparency of food products for consumers, labeling can help individuals contract with trusted service providers. Best practices might start with the Codes of Fair Information Practice, and grow to include anonymizing data when possible (Cheng & Prabhakar, 2004; Horey, Groat, Forrest, & Esponda, 2007), collecting minimal information (Agre, 1994), visualizing and explaining data analysis and aggregation procedures, and supporting audit trails (Weitzner et al., 2008) and data retention limits (Bannon, 2006; Blanchette & Johnson, 2002; Dodge & Kitchin, 2007).

Legal privilege for raw location data

Finally, a legal evidentiary privilege similar to the one commonly recognized for trade secrets must be created to protect mobile sensing data. If individuals and groups are to explore their world with new sensing tools, they should not be threatened by the potential negative consequences of capturing accurate information about themselves. If raw location data is too easily discoverable in civil litigation, individuals or entire demographics might be dissuaded from participation in this new

form of investigation. A qualified privilege modeled after the trade secrets privilege strikes a good balance of protecting this sensitive data from casual and unnecessary disclosure.[6]

Participatory privacy regulation in practice

What does the Personal Data Stream architecture, bolstered by these principles and structures, look like in practice? Consider the example below.

> Maya is a participant in PEIR. PEIR uses Maya's time-location series to infer how much she drives and whether she spends time near polluted highways. Maya heard about PEIR from a friend, and checked out the PEIR website. She was happy to discover that PEIR is a 'fair data' service provider. She knows this means they will not keep her location data for more than a year and they will not share them with third parties. She also notices that PEIR hosts lively discussion forums about everything from reducing carbon impact to privacy issues. PEIR gathers participant suggestions and iterates on the service based on user input to these forums. Maya decides she feels comfortable trying PEIR, and agrees to let her Personal Data Vault release her location information to PEIR every day between 7:30 am (when she typically leaves home) and 6:30 pm (when she arrives back home).
>
> When Maya logs onto PEIR at the end of her day, she can see a map illustrating all of her driving trips and periods of inactivity throughout the day. She can see what time she arrived at work, her trip to the doctor's office, and her quick stop at the liquor store. The PEIR map encourages her to click on each trip to learn more about conditions (such as time spent idling) that raised her carbon impact. As she clicks on each trip, she is reminded that she is sharing these data with PEIR. She also thinks about whether she is comfortable sharing each trip with a group of friends. She decides to share her commute, but not her trip to the doctor, with her 'Friend Map.' She has already added her partner and a few coworkers to her Friend Map, and can see how impact and exposure compare to those of people she knows. As she reviews her maps and daily impact, she decides to delete that trip to the liquor store from PEIR's database.
>
> The next morning, one of Maya's favorite blogs posts an article on location privacy and the dangers of creating an archive of personal location data. Maya reads the article and is inspired to comment on the blog about her experience using PEIR. Maya knows the data in her Personal Data Vault is secure and free from the risk of subpoena. But PEIR is less secure. She weighs the blog's arguments and chooses to continue using PEIR for the near future. She decides, however, that she may delete all of the data in her PEIR account after a few months of use.

V. Future Work

Technical innovations in designing and securing the Personal Data Vault, improving data interpretation through user interfaces, and constructing auditing methods will be important to the success of participatory privacy regulation. We have also identified several areas with unsolved

[6] The Federal Rules of Evidence did not adopt an explicit trade secrete privilege rule. However, in practice, some such privilege has been recognized in federal and state courts as a matter of common law, or in certain states by its rules of evidence. See, e.g., Cal. Evid. Code § 1060 (2009) ("If he or his agent or employee claims the privilege, the owner of a trade secret has a privilege to refuse to disclose the secret, and to prevent another from disclosing it, if the allowance of the privilege will not tend to conceal fraud or otherwise work injustice.") For a more detailed argument for a legal privilege for mobile sensing data, see Kang, Burke, Estrin, Hansen, & Shilton, under development.

technical, legal and social challenges. Addressing these areas will be necessary to ensure that individuals can fully participate in privacy decision-making.

Creating a business model for the data vault that does not rely on mining location data is a central unmet challenge. Perhaps data vaults could be hosted by universities or nonprofits, or alternately tied to existing enterprises such as banks. Developing a financial infrastructure to support vigorous development and use of PDVs is necessary before a vision of participatory privacy regulation can be realized. Another challenge will be defining 'fair data' best practices for third party applications. Best practices for ubiquitous computing suggested by scholars include design to draw awareness to, and increase accountability of, capture systems (Bellotti & Sellen, 1993; Langheinrich, 2002), methods to enable ubiquitous computing systems to allow for multiple identities and pseudonomy (Phillips, 2002, 2005), techniques and recommendations for privacy-sensitive design (Camp & Connelly, 2008; Iachello, Smith, Consolvo, Chen, & Abowd, 2005; Kang & Cuff, 2005), and trusted architectures to protect personal information (Cáceres et al., 2009; Hong & Landay, 2004). Much as the process convened to establish the Codes of Fair Information Practice took negotiation between diverse experts (Waldo et al., 2007), discussion and debate will determine appropriate definitions for 'fair data' requirements. A final challenge is the social measures necessary to encourage broad data and location privacy literacy. Helping individuals or groups understand the importance of data privacy, as well as the possibilities of sharing, may become as important as financial literacy. Individuals may also need guidance in seeking redress for breaches of data sharing licenses and contracts with third party applications.

Outcomes of participatory privacy regulation and the Personal Data Stream architecture, and indicators of its success, will be important to understanding the ability of the principles we have described to tackle information privacy in a distributed datascape. There are numerous ways researchers could document improvements in the environment in which personal sensing participants make privacy decisions. We describe a few possibilities below.

Documenting participant satisfaction

Our ongoing research will begin analyzing the success of participatory privacy regulation measures by undertaking interviews with, and observations of, participants in mobile sensing. These qualitative data will gauge participants' level of engagement in, and satisfaction with, sensing research and privacy decision-making. Qualitative data can suggest answers to contextual questions about when and why participants make decisions to share or withhold data. Interviews with participants can elicit how participants feel while interacting with sensing systems and how much participants trust the systems. Interviews can also establish whether participants engage in a broader debate about privacy with people they know or in public spaces. Direct observation of participants during the course of data collection and data analysis can also provide clues to when and why participants feel boundary or identity sensitivities, and whether participatory privacy regulation principles adequately address these sensitivities. Explicit participant critique of our design methods and software can help us understand whether our systems make data legible and fortify participants in the ways our principles intend. But because interviews are time consuming to perform and analyze, the sample of mobile sensing participants interviewed will necessarily be small.

Documenting participant engagement

To bolster the small sample that we can effectively interview, we can add quantitative use statistics drawn from the Personal Data Vault and third party applications such as PEIR. How frequently do users take advantage of privacy features from data deletion to 'fair data' certification labels? With participant permission and even participation, we can analyze logs detailing use of

existing software to evaluate how frequently and under what conditions participants engage in existing privacy processes such as changing data resolution, sharing selectively, and deleting. We can compare individual's privacy actions to their degree of involvement, measured according to amount of data gathered, and length and frequency of involvement in data gathering. These quantitative data will indicate whether and how engagement correlates with sharing and discretion decisions.

Considering cognitive outcomes

Future research could also take a psychological approach to measuring participants' understanding of, and engagement in, privacy regulation. Do participants understand the flow of their data through the vault architecture and third party systems? Do they understand privacy risks? Can legibility measures such as data visualization or inference agents improve individual cognition of risk? Psychological measurements of understanding of these concepts might help us understand whether our efforts have fortified the decision-making skills of sensing participants.

VI. Conclusion

Mobile sensing provides the ability to bring individuals and groups into research on a massive scale, opening up data collection and participation in data analysis by taking advantage of mobile phones, tools widely adopted across the world. Applying this broad notion of participation directly to privacy can help to mitigate some of the invasive aspects of this vision. This paper has described what participating in privacy might entail, and described a system architecture and set of social structures that might enable such participation. Through this process, we hope to build mobile sensing technologies that reflect a vision for data privacy as a continual, engaged process of discretion and disclosure according to context. By creating systems that enable interaction and data flow control, mobile sensing can enable meaningful, workable information privacy.

VII. Acknowledgments

The authors would like to thank the CENS data practices team (Dr. Christine Borgman, David Fearon, Andrew Lau, Matt Mayernik, Alberto Pepe and Jillian Wallis) and the USC Embedded Networks Laboratory (Shuai Hao, Nilesh Mishra and Sumit Rangwala) for review and helpful comments on this work. This material is based upon work supported by the National Science Foundation under grant no. 0832873.

VIII. References

Acquisti, A., & Grossklags, J. (2008). What can behavioral economics teach us about privacy? In *Digital Privacy: Theory, Technologies, and Practices* (pp. 363-377). New York and London: Auerbach Publications.

Agre, P. E. (1994). Surveillance and capture: two models of privacy. *The Information Society*, *10*(2), 101-127.

Bannon, L. (2006). Forgetting as a feature, not a bug: the duality of memory and implications for ubiquitous computing. *CoDesign*, *2*(1), 3-15.

Barkhuus, L., & Dey, A. (2003). Location-Based Services for Mobile Telephony: a study of users' privacy concerns. In *Proceedings of the INTERACT 2003, 9TH IFIP TC13 International Conference on Human-Computer Interaction* (Vol. 2003, pp. 709-712).

Bellotti, V., & Sellen, A. (1993). Design for privacy in ubiquitous computing environments. In *Third European Conf. Computer-Supported Cooperative Work ECSCW'93* (pp. 77-92). Milano, Italy: Dordrecht Kluwer.

Blanchette, J., & Johnson, D. G. (2002). Data retention and the panoptic society: the social benefits of forgetfulness. *The Information Society*, *18*(33-45).

Braman, S. (2006). *Change of state: information, policy, and power*. Cambridge, MA and London: The MIT Press.

Cáceres, R., Cox, L., Lim, H., Shakimov, A., & Varshavsky, A. (2009). Virtual Individual Servers as Privacy-Preserving Proxies for Mobile Devices. In *Proc. of 1st ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds (MobiHeld)*. Presented at the 1st ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds (MobiHeld), Barcelona, Spain: ACM.

Calhoun, C. (2000). Social theory and the public sphere. In *The Blackwell Companion to Social Theory* (Second., pp. 505-544). Malden, MA: Blackwell Publishing.

Camp, L. J., Asgharpour, F., Liu, D., & Bloomington, I. N. (2007). Experimental Evaluations of Expert and Non-expert Computer Users' Mental Models of Security Risks. In *Proceedings of WEIS 2007*. Presented at the WEIS 2007, Pittsburgh, PA.

Camp, L. J., & Connelly, K. (2008). Beyond consent: privacy in ubiquitous computing (Ubicomp). In *Digital Privacy: Theory, Technologies, and Practices* (pp. 327-343). New York and London: Auerbach Publications.

Cargo, M., & Mercer, S. L. (2008). The Value and Challenges of Participatory Research: Strengthening Its Practice. *Annual Review of Public Health*, *29*, 325-350.

Cheng, R., & Prabhakar, S. (2004). Using uncertainty to provide privacy-preserving and high-quality location-based services. In *Workshop on Location Systems Privacy and Control, MobileHCI* (Vol. 4).

Chenok, D. J. (2009). *Toward A 21st Century Framework for Federal Government Privacy Policy*. Washington, D.C.: Information Security and Privacy Advisory Board.

Clarke, R. (2000). *Beyond the OECD Guidelines: Privacy Protection for the 21st Century*. Canberra, Australia: Xamax Consultancy Pty Ltd. Retrieved August 5, 2009, from http://www.rogerclarke.com/DV/PP21C.html.

Connolly, C. (2008). The US Safe Harbor - Fact or Fiction? *Privacy Laws and Business International*, (96). Retrieved July 7, 2009, from http://www.galexia.com/public/research/assets/safe_harbor_fact_or_fiction_2008/.

Curry, M. R., Phillips, D. J., & Regan, P. M. (2004). Emergency response systems and the creeping legibility of people and places. *The Information Society*, *20*, 357-369.

Dodge, M., & Kitchin, R. (2007). 'Outlines of a world coming into existence': pervasive computing and the ethics of forgetting. *Environment and Planning B: Planning and Design*, *34*(3), 431-445.

Eagle, N. (2008). Behavioral Inference across Cultures: Using Telephones as a Cultural Lens. *Intelligent Systems, IEEE*, *23*(4), 62-64.

Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R. A., Ahn, G. S., & Campbell, A. T. (2006). MetroSense Project: People-Centric Sensing at Scale. In *Proceedings of the ACM Sensys World*

Sensor Web Workshop. Presented at the ACM Sensys World Sensor Web Workshop, Boulder, CO: ACM.

Freiwald, S., & Swire, P. (2009, April 17). Phone Tracking Should Require a Warrant. *ACSBlog: The Blog of the American Constitution Society*. Retrieved April 21, 2009, from http://www.acsblog.org/ip-and-tech-law-phone-tracking-should-require-a-warrant.html.

Goldman, J., Shilton, K., Burke, J., Estrin, D., Hansen, M., Ramanathan, N., et al. (2009). *Participatory Sensing: A citizen-powered approach to illuminating the patterns that shape our world*. Washington, DC: Woodrow Wilson International Center for Scholars.

Google Health. (n.d.). . Retrieved June 28, 2009, from https://www.google.com/health.

Hao, S., Mishra, N., Mun, M., Rangwala, S., Shilton, K., Burke, J. A., et al. (2009). Controlled data sharing in an online world. Under review.

Hong, J. I., & Landay, J. A. (2004). An architecture for privacy-sensitive ubiquitous computing. In *Proceedings of the 2nd international conference on Mobile systems, applications, and services* (pp. 177-189). Boston, MA, USA: ACM.

Horey, J., Groat, M. M., Forrest, S., & Esponda, F. (2007). Anonymous data collection in sensor networks. In *Proceedings of the 4th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. Presented at the 4th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, Philadelphia, PA: ACM.

Iachello, G., Smith, I., Consolvo, S., Chen, M., & Abowd, G. D. (2005). Developing privacy guidelines for social location disclosure applications and services. In *Proceedings of the 2005 symposium on Usable privacy and security* (pp. 65-76). Pittsburgh, Pennsylvania: ACM.

Information Security Awareness Forum. (2009). *Personal Data Guardianship Code*. Swindon, UK: The British Computer Society. Retrieved from http://www.bcs.org/server.php?show=nav.10666.

Kang, J., Burke, J. A., Estrin, D., Hansen, M., & Shilton, K. (2009). Self-Analytic Privacy. Under development.

Kang, J., & Cuff, D. (2005). Pervasive Computing: Embedding the Public Sphere. *Washington and Lee Law Review*, *65*, 93-146.

Kelty, C. M. (2008). *Two Bits: The Cultural Significance of Free Software*. Durham, NC: Duke University Press.

Krishnamurthy, B., & Wills, C. (2009). On the leakage of personally identifiable information via online social networks. In *Proceedings of the ACM SIGCOMM Workshop on Online Social Networks*. Presented at the ACM SIGCOMM Workshop on Online Social Networks, Barcelona, Spain: ACM.

Lam, M. (2009, April 14). Building a Social Networking Future Without Big Brother. Presented at the POMI 2020 Workshop, Palo Alto, CA. Retrieved June 3, 2009, from http://suif.stanford.edu/%7Elam/lam-pomi-ws09.pdf.

Langheinrich, M. (2002). A Privacy Awareness System for Ubiquitous Computing Environments. In *UbiComp 2002: Ubiquitous Computing : 4th International Conference* (pp. 315-320). Göteborg, Sweden: Springer.

Microsoft HealthVault. (n.d.). . Retrieved June 28, 2009, from http://www.healthvault.com/.

Mitrou, L. (2008). Communications data retention: a Pandora's box for rights and liberties? In *Digital Privacy: Theory, Technologies, and Practices* (pp. 409-433). New York and London: Auerbach Publications.

Mun, M. Y., Estrin, D., Burke, J., & Hansen, M. (2008). Parsimonious mobility classification using GSM and Wi-Fi traces. In *Proceedings of the Fifth Workshop on Embedded Networked Sensors (HotEmNets)*. Presented at the Fifth Workshop on Embedded Networked Sensors (HotEmNets), Charlottesville, VA.

Mun, M. Y., Reddy, S., Shilton, K., Yau, N., Boda, P., Burke, J. A., et al. (2009). PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services*. Presented at the International Conference on Mobile Systems, Applications, and Services, Krakow, Poland.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, *79*(1), 119–158.

Palen, L., & Dourish, P. (2003). Unpacking "privacy" for a networked world. In *CHI 2003* (Vol. 5, pp. 129-136). Ft. Lauderdale, FL: ACM.

Phillips, D. J. (2005). From privacy to visibility: context, identity, and power in ubiquitous computing environments. *Social Text*, *23*(2), 95-108.

Phillips, D. J. (2002). Negotiation the digital closet: online pseudonyms and the politics of sexual identity. *Information, Communication & Society*, *5*(3).

Seifman, D. (2007, August 31). 'Track' Man is Sacked. *New York Post*. Retrieved June 27, 2009, from http://www.nypost.com/seven/08312007/news/regionalnews/track_man_is_sacked.htm.

Shallwani, S., & Mohammed, S. (2007). *Community-Based Participatory Research: A Training Manual for Community-Based Researchers*. Aga Khan University – Human Development Programme.

Shilton, K., Burke, J., Estrin, D., Hansen, M., & Srivastava, M. (2008). Participatory privacy in urban sensing. Presented at the MODUS 2008, St. Louis, Missouri.

Stone, B., & Stelter, B. (2009, February 19). Facebook Withdraws Changes in Data Use. *The New York Times*. Retrieved June 11, 2009, from http://www.nytimes.com/2009/02/19/technology/internet/19facebook.html.

Traceroute. (2009). In *Wikipedia, the free encyclopedia*. Retrieved July 26, 2009, from http://en.wikipedia.org/wiki/Traceroute.

Tsai, J. Y., Kelley, P., Drielsma, P., Cranor, L. F., Hong, J., & Sadeh, N. (2009). Who's viewed you?: the impact of feedback in a mobile location-sharing application. In *Proceedings of the 27th international conference on Human factors in computing systems* (pp. 2003-2012). Boston, MA, USA: ACM.

U.S. Department of Health, Education, and Welfare, U. D. O. H. (1973). *Records, Computers, and the Rights of Citizens*. Cambridge, MA: MIT Press.

Waldo, J., Lin, H. S., & Millett, L. I. (2007). *Engaging privacy and information technology in a digital age*. Washington, D.C.: The National Academies Press.

Weitzner, D. J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., & Sussman, G. J. (2008). Information accountability. *Communications of the ACM*, *51*(6), 82-87.

Westin, A. F. (1970). *Privacy and freedom*. New York: Atheneum.

Zittrain, J. (2008). *The Future of the Internet--And How to Stop It*. New Haven & London: Yale University Press.