Four Billion Little Brothers?

Privacy, mobile phones, and ubiquitous data collection

Katie Shilton, University of California, Los Angeles

They place calls, surf the Internet, and there are close to 4 billion of them in the world. Their built-in microphones, cameras, and location awareness can collect images, sound, and GPS data. Beyond chatting and texting, these features could make phones ubiquitous, familiar tools for quantifying personal patterns and habits. They could also be platforms for thousands to document a neighborhood, gather evidence to make a case, or study mobility and health. This data could help you understand your daily carbon footprint, exposure to air pollution, exercise habits, and frequency of interactions with family and friends.

At the same time, however, this data reveals a lot about your regular locations, habits, and routines. Once such data is captured, acquaintances, friends, or authorities might coerce you to disclose it. Perhaps worse, it could be collected or reused without your knowledge or permission. At the extreme, mobile phones could become the most widespread embedded surveillance tools in history. Imagine carrying a location-aware bug, complete with a camera, accelerometer, and Bluetooth stumbling, everywhere you go. Your phone could document your comings and goings, infer your activities throughout the day, and record whom you pass on the street or who engaged you in conversation. Deployed by governments or compelled by employers, 4 billion "little brothers" could be watching you.

Whether phones engaged in sensing data are tools for self and community research, coercion, or surveillance depends on who collects the data, how it is handled, and what privacy protections users are given. As these new forms of data begin to flow over phone networks, application developers will be the first line of defense for protecting the sensitive data collected by always-present, always-on mobile phones.

I should mention that I'm *not* one of the developers on the front lines. I work in science and technology studies (STS; http://en.wikipedia.org/wiki/Science_and_technology_studies), a social science interested in the ways people, technologies, and data interact and affect each other. The developers I work with might say STS is about telling them what they *should* be doing—which I must admit is the goal of this article. I worry about the consequences of mobile phones as sensors, and have a number of opinions about what programmers, as well as social scientists, might do to make this sort of data collection work without slipping into coercion, surveillance, and control.

PARTICIPATORY SENSING

Research that uses mobile phones to collect data for personal or social projects is called mobile, urban, or participatory sensing.[2-4] Participatory sensing is meant to enable (and encourage) anyone to gather and investigate previously invisible data. It tries to avoid surveillance or coercive sensing by emphasizing individuals' participation in the sensing process. Applications designed to enable participatory sensing range from the very personal and self-reflective to shareable data meant to improve an individual's health or a community's experience. This article examines three applications from UCLA's CENS (Center for Embedded Networked Sensing) to illustrate the diversity of possibilities, as well as suggest data collection and sharing concerns.

PEIR (Personal Environmental Impact Report). Participants in PEIR (http://peir.cens.ucla.edu/) carry mobile phones throughout their day to calculate their carbon footprints and exposure to air pollution—both big concerns in smoggy Los Angeles where the project is based. By referencing GPS and cell towers, the phones upload participants' locations every few seconds. Based on these time-location 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 Southern California regional air quality and weather data to estimate individual carbon footprint and exposure to particulate matter. Sensing a participant's location throughout the day enables more accurate and previously unavailable information about environmental harms people face, as well as the harms they create. To participate, individuals need to record and submit a continuous location trace.

Biketastic. This project (http://biketastic.com) improves bike commuting in Los Angeles, a city notoriously unfriendly to cyclists. Bikers carry a GPS-enabled mobile phone during their commutes. The phone automatically uploads bikers' routes to a public Web site. The phone also uses its accelerometer to document the roughness of the road, and takes audio samples to analyze volume of noise along the route. Participants can log in to see their routes combined with existing data, including air quality, time-sensitive traffic conditions, and traffic accidents. They can also use the system to share information about their routes with other riders. By combining existing local conditions with biker-contributed data, Biketastic will enable area bikers to plan routes with the least probability of traffic accidents; with the best air quality; or according to personal preferences, such as road-surface quality or connections with public transportation. Biketastic shares location data through a public map, though individuals use pseudonymous user names.

AndWellness. Currently under development, AndWellness is a personal monitoring tool designed to encourage behavioral change. It helps clients work independently or with a coach to document places and times when they stray from a healthy eating or exercise plan. During an initial week of documentation, AndWellness prompts users to input personal assessments throughout the day. These assessments ask users when they last ate and whether they were on plan. After a week of tracking and data analysis, users can see places and times where they tend to stray from their plan, and plan interventions to combat unwanted variations. AndWellness collects not only location, but also sensitive data about diet and habits. Individuals might choose to share this data with a support group, coach, therapist, doctor, family, or friends.

Taking participatory sensing from a possibility enabled by the mobile-phone network to a coordinated reality is rife with challenges. Among these challenges are the ethics of repurposing phones, now used as communication tools, for data collection and sharing. How can individuals determine when, where, and how they wish to participate? How much say do they get over what they wish to document and share?

PRIVACY IN PARTICIPATORY SENSING

Privacy—the ability to understand, choose, and control what personal information you share, with whom and for how long—is a huge challenge for participatory sensing. Privacy decisions have many components, including identity (who is asking for the data?), granularity (how much does the data reveal about me?), and time (how long will the data be retained?)[7,10,11] Location traces can document and quantify habits, routines, and personal associations. Your location might reveal your child's school, your regular trips to a therapist or doctor, and times when you arrived late or left early from work. These traces are easy to mine and difficult or impossible to retract once shared.

Sharing such granular and revealing digital data could have a number of risks or negative consequences. Safety and security threats are obvious: thieves, stalkers, etc. are possible dangers. Perhaps less obvious—and probably more likely—are other social consequences. Think about how frequently you beg off a social engagement with a little white lie, or keep your location and activities secret to surprise a friend. Much like Facebook's ill-fated Beacon service, participatory sensing could disrupt the social boundaries we have come to expect. What if someone with authority over you (your employer, the government) collects or accesses your location data? It's easy to imagine a chilling effect on legal, but stigmatized, activities. Would you be as likely to attend a political protest, or visit a plastic

surgeon, if you knew your location was visible to others? Large databases of location data accessible by subpoena also could become evidence for everything from minor civil disputes to messy divorce cases.

Maybe most importantly, privacy is an important part of your identity and self-presentation. Deciding what to reveal to whom is part of deciding who you are. I might want to track when and where I tend to overeat, but I see no reason to share that information with anyone but my doctor. Similarly, I might take part in a political data collection project on the weekend, but that doesn't mean my parents need to know. Respecting the many gradations between public and private, and giving people the ability to negotiate those gradations, are integral to respecting individual privacy.

In the United States and Europe, *fair information practices* are one standard for protecting the privacy of personal data. Originally codified in the 1970s, the Code of Fair Information Practices outlines data-management principles to help organizations protect personal data.[12,13] These codes are still considered a gold standard for privacy protection.[14] But the principles, designed for corporations or governments rather than many distributed data collectors, are no longer enough. Data gathered during participatory sensing is more granular than traditional personal data (name, Social Security number, etc.). It reveals much more information about an individual's habits and routines. Furthermore, data is no longer gathered solely by large organizations or governments with established data practices. Individuals or community groups might create participatory sensing applications and begin collecting personal data.[15]

ENABLING PARTICIPATION IN PRIVACY

This is where the responsibility of developers comes into the equation. How can developers help individuals or small groups launching participatory sensing projects implement appropriate data-protection standards? To create workable standards with data so granular and personal, systems must actively engage individuals in their own privacy decision making. At CENS, we call this *participatory privacy regulation*—the idea that systems can help users to negotiate disclosure decisions depending on context (who is asking, what is being asked for, etc.). We need to build systems that improve users' ability to make sense of, and thereby regulate, their privacy.

Building such systems is a major, unmet challenge.[6] As the first steps toward meeting this challenge, we propose three new principles for developers to consider and apply when building mobile data-gathering applications. These principles are purposefully broad, because "acceptable" data practices might vary across applications (a medical project might be justified in collecting much more personal data, with stringent protections, than a community documentation project). These principles are thinking tools to help developers adapt privacy protections to participatory sensing applications.

PARTICIPANT PRIMACY

The goal of participatory privacy regulation is to give participants as much control over their location data as possible. GPS traces or the secondary traces created by geotagged media should belong to individuals. Participants should be able to make and revoke decisions to share subsets of the data with third-party applications. Framed this way, participants are not just subjects of data collection, but take the role of investigators (when they collect data to participate in self-analytic applications) or co-investigators (when they contribute their data to larger research initiatives). As such, they should have input into how data is collected, processed, stored, and discarded.

Developers can enable participants to own and manage their data by tailoring access-control and data-management tools for use by individual participants. Users collecting revealing sensing data are going to need secure storage and intuitive interfaces to manage access and sharing. As an example, CENS researchers are developing a PDV (personal data vault) to give individuals private and robust storage for their sensing data. The PDV provides services such as authentication and access control, allowing participants not only to collect all of their sensing data in one place, but

also to specify which individuals and groups in their social network can see which datasets. Similar tools are in development in research labs at Stanford[8] and AT&T,[1] and are not unlike commercial applications such as Google Health[5] and Microsoft's HealthVault.[9]

As developers build data-management tools to put personal data control back in the hands of individuals, they will need to think about which controls users will need to make privacy and sharing decisions. At a very basic level, sharing decisions should take into account identity (who's asking?), time (send data only between 9 a.m. and 5 p.m.), location (send data only when I'm on campus), and data type (share only geotagged photos). More advanced techniques for developers to consider include access controls based upon activity (share only driving routes) or routine (don't share anomalous routes).

Application developers can further protect participant primacy by limiting the amount of raw data a participant is required to share outside of the vault. When privacy is at stake, more data is not always better. For example, participants in Biketastic may collect their location data 24/7 to the PDV, but share data with Biketastic only during days and times when they regularly commute by bike. Biketastic doesn't need to know where the participants are during working hours, when they take their lunch breaks, or what they do during their evenings. A different example of collecting minimal data is requesting processed, rather than raw, data. Developers could build applications such as PEIR to request only inferred activity data (time spent driving, walking, and indoors) and ZIP code, rather than granular location data. PEIR doesn't need to know what street a participant was on—only what carbon-generating activity they were engaged in. By collecting the minimum amount of information needed for a service, application developers can help participants maintain control over their raw data.

DATA LEGIBILITY

Participatory 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 using 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 goes and how long it remains accessible. System features should increase participants' understanding of complex risks and help them make better decisions about data capture, sharing, and retention.

Developers should get creative about what legibility might mean. An application's user interface, for example, could help users not only set data-sharing policies, but also see the results of their policies. Imagine a Facebook pop-up that asks, "Do you really want to share the album 'Party Pics' with your father?" Developing features either for data vaults or for sensing applications that illuminate who can see what data will help users better understand the consequences of data sharing.

Another approach is to show multiple interpretations of collected data. The AndWellness interface, for example, uses both maps and timelines to help users draw conclusions about when and where their eating habits strayed from their plans. Developers might also experiment with natural language, helping translate numerical data or complex algorithms into something easier to understand. Natural language might make inferences from data points (e.g., this bike route has a few hills in the middle, most of them easy, and one difficult hill at the end); or plain text descriptions can explain how calculation and processing works (e.g., clicking on a route in PEIR takes the participant to a "Trip Journal" with a step-by-step breakdown of how the system calculated the impact and exposure for that route.

LONGITUDINAL ENGAGEMENT

Finally, developers will need to consider time as a factor that affects privacy in participatory sensing. You may end participation in a carbon footprint calculator when you start taking public transit to work, but enroll in a new health

program after receiving a surprising diagnosis. Personal habits and routines change over time, altering the data collected into personal data vaults.

Because time is such a critical factor, application interfaces should encourage participants to engage with the data from the point of collection through analysis, long-term retention, or deletion. Systems should enable continued engagement to allow participants to change their data practices as their context changes. The crux of engaging individuals with decisions about their data is refusing to put that data in a black box. Instead, analyzing, learning from the data, and making ongoing choices about the data become the goals of sensing.

We offer several suggestions for how developers can encourage long-term engagement. Policies that require users to check back in with a vault or application on a regular basis can remind them to update their sharing preferences as their needs change. A data vault could remind users to update their sharing preferences every time they add new contacts or applications. Building adaptive filters can also enable participants to change their data sharing as their preferences change. Such filters could learn from user behavior to respond to privacy preferences. For example, the vault could learn never to share a certain route or could learn to check with users before sharing any routes recorded after 9 p.m.

A TraceAudit is another idea that helps users engage with their data over time. The TraceAudit builds on the idea of an Internet traceroute and relies on careful logging procedures. An interface that allows users access to logs can let them trace how their data is used by an application, where the data has been shared, and who has had access to it. For example, a TraceAudit of data use in PEIR can show participants exactly how their raw location traces become calculations of impact and exposure, and how data is shared during that process. A log could show users that their PDV sent PEIR raw data on weekdays between the hours of 7 a.m. and 8 p.m. PEIR performs activity classification based on this raw data (minutes spent walking, driving, etc.) and sends a summary of the activities and the ZIP codes in which they occurred to the California Air Resources Board. PEIR receives back PM2.5 (fine particle) pollution exposure and CO2 emission values to correspond with these activities and ZIP codes. PEIR then saves and displays these total calculations for users. The TraceAudit provides transparency and accountability, helping individuals to see how PEIR has used and shared their data.

CHALLENGES BEYOND TECHNOLOGY

System design that pays attention to participant primacy, longitudinal engagement, and data legibility will help users make data-sharing decisions and protect their privacy in participatory sensing. Technical decisions, however, won't be enough to ensure privacy for sensing participants. Participant engagement in privacy decision making needs to be fortified by supporting social structures, as well.

Participatory sensing opens the door to entirely new forms of granular and pervasive data collection. The risks of this sort of data collection are not always self-evident. Even if we give people options for managing their data, they may not understand the benefits of doing so. Data literacy must be acquired over time through many avenues. Public discussion and debate about participatory sensing will be critical to educating participants about the risks and possibilities of sensing data. Discussion forums and blogs play an important role, as do traditional media and even community groups.

Further, participants in participatory sensing are going to need to understand what happens with their data once it leaves their personal vault and is used by third-party applications. Diverse and plentiful applications for participatory sensing data can help to achieve the potential usefulness of participatory sensing, but will also make it difficult for participants to understand which applications are trustworthy and abide by acceptable data practices. Participants need to know what they are signing up for—and cryptic, fine-print EULAs (end-user license agreements) aren't the answer. Users should know how long an application will retain their data and whether it will pass the data on to other parties.

A voluntary labeling system, much like “Fair Trade” labels on food, could help consumers distinguish applications that abide by a minimum set of responsible data practices. These might include logging data use and keeping audit trails, and discarding location data after a specified period of time. Such measures could help to increase transparency of participatory sensing applications.

Finally, enhanced legal protections for unshared vault data can encourage participation in participatory sensing. Ongoing work is investigating the possibility of a legal privilege for personal-sensing data. Such a privilege could be enabled by statute and modeled on attorney-client or doctor-patient privilege.

CONCLUSION

While lawyers and social scientists work on structural changes to help ensure privacy in participatory sensing, many of the initial and critically important steps toward privacy protection will be up to application developers. By innovating to put participants first, we can create systems that respect individuals’ needs to control sensitive data. We can also augment people’s ability to make sense of such granular data, and engage participants in making decisions about that data over the long term. Through attention to such principles, developers will help to ensure that 4 billion little brothers are not watching us. Instead, participatory sensing can have a future of secure, willing, and engaged participation. Q

ACKNOWLEDGMENTS

Many thanks to collaborators Jeffrey Burke, Deborah Estrin, and Mark Hansen, whose ideas and contributions have shaped this material. This article is based upon work supported by the National Science Foundation under Grant No. 0832873.

REFERENCES

1. Cáceres, R., Cox, L., Lim, H., Shakimov, A., Varshavsky, A. 2009. Virtual individual servers as privacy-preserving proxies for mobile devices. Proceedings of First ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds (MobiHeld), Barcelona, Spain.
2. Cuff, D., Hansen, M., Kang, J. 2008. Urban sensing: out of the woods. *Communications of the ACM* 51: 24-33.
3. Eagle, N. 2008. Behavioral inference across cultures: using telephones as a cultural lens. *IEEE Intelligent Systems* 23: 62-64.
4. 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. *Proceedings of the ACM Sensys World Sensor Web Workshop,* Boulder, Colorado.
5. Google Health; https://www.google.com/health.
6. Iachello, G., Hong, J. 2007. End-user privacy in human-computer interaction. *Foundations and Trends in Human-Computer Interaction* 1: 1-137.
7. Kang, J. 1998. Privacy in cyberspace transactions. *Stanford Law Review* 50: 1193-1294.
8. Lam, M. 2009. Building a social networking future without Big Brother;. http://suif.stanford.edu/%7Elam/lam-pomi-ws09.pdf.
9. Microsoft HealthVault; http://www.healthvault.com/.
10. Nissenbaum, H. 2004. Privacy as contextual integrity. *Washington Law Review* 79: 119–158.
11. Palen, L., Dourish, P. 2003. Unpacking “privacy” for a networked world. CHI 2003, Ft. Lauderdale, FL: 129-136.
12. Personal Privacy in an Information Society: The Report of The Privacy Protection Study Commission. 1977; http://epic.org/privacy/ppsc1977report/.

13. U.S. Department of Health, Education, and Welfare. 1973. Records, Computers, and the Rights of Citizens. Cambridge, MA: MIT Press.
14. 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.
15. Zittrain, J. 2008. The future of the Internet—and how to stop it. New Haven and London: Yale University Press.

LOVE IT, HATE IT? LET US KNOW
FEEDBACK@QUEUE.ACM.ORG

KATIE SHILTON is a doctoral student in information studies at the UCLA. Her research explores privacy and ethical challenges raised by ubiquitous sensing technologies, and she coordinates a research project at the Center for Embedded Networked Sensing focused on these questions. She received a B.A. from Oberlin College in 2003 and a masters of library and information science from UCLA in 2007.