Participatory Sensing and New Challenges to U.S. Privacy Policy

Katie Shilton, Deborah Estrin, UCLA CENS

Mobile phones have created a radical new platform for data collection, discovery, and social analysis. At the same time, they fundamentally challenge our current understandings of privacy policy and data security. Mobile phones 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. Mobile phones are more often on and carried than any previous personal technology, and because they are connected to location services and the web, they can use a wealth of web-based data as context. Participatory sensing (PS) is a new practice which harnesses these tools to collect and analyze data for use in social science, environmental and health discovery.

Participatory sensing shapes phones into ubiquitous, familiar tools for quantifying personal patterns and habits. Phones become platforms for thousands to document a neighborhood, gather evidence to make a case, or study mobility and health (Anokwa, Borriello, Pering, & Want, 2007; Burke et al., 2006; Eisenman et al., 2006; Miluzzo, Lane, Eisenman, & Campbell, 2007; Pentland, Lazer, Brewer, & Heibeck, 2009). In addition, phones can be programmed for manual, automatic, and context-aware data capture. Because of the sheer ubiquity of mobile phones and associated communication infrastructure, it is possible to engage people of all backgrounds nearly everywhere in the world and collectively, provide unprecedented access to high resolution, real time, and scalable spatio-temporal data.

An infrastructure to collect, coordinate and analyze these data will enable researchers to conduct studies at individual, community, and population scales; research that addresses socially critical issues related to human behavior, sustainability, health, and education. However, a significant barrier to adoption of this technology is the need for careful attention to the privacy issues and data practices surrounding these very personal and intimate data. The issue is particularly challenging because of the multiple stakeholders involved in these systems: in particular, end-users (participants in studies), researchers designing and conducting studies, and mobile carriers and application developers who collect, warehouse, and distribute participatory sensing data.

Participatory sensing

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. As examples, we present three applications from UCLA's 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. By referencing GPS and cell towers, the phones upload participants' locations every few seconds. Based on these time-location traces, the PEIR system infers participant activities (indoors, walking, driving) 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. AndWellness is a personal monitoring tool designed to help individuals manage health conditions. AndWellness phones are programmed to prompt the user for quick input at 'appropriate times and places' during the course of their day, wherever they are. These "experience samples" are automatically time stamped, geocoded, uploaded, and stored in a database according to the prompt and the response details. Patients with conditions such as diabetes, who are struggling to stabilize their hypertension, can record frequent physiological measures (BP, BG, weight), and timing/dosage of medication. They can also document in-the-moment self-reports on physical symptoms and side effects such as dizziness and fatigue. Such data can help the clinician and patient build a picture over a week or two to inform personalization of the care plan. In addition to giving the clinician the information they need to optimize the patient's care plan, the same systems can be used to help patients with desired health-behavior changes--the notion of a personal-coach in your pocket--whether the behavior of interest is smoking, diet, prenatal care, or parenting. In order to fulfill this vision, 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?) (Kang, 1998; Nissenbaum, 2009; Palen & Dourish, 2003). 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. These traces also form living records that are pre-transactional: they are even less public than purchases from Amazon or web searches, or even an interaction with a doctor. And more often than not, location traces and associated data cannot be effectively anonymized.

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 individuals beg off a social engagement with a little white lie, or keep 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. And if authorities such as employers or local and federal governments collect or access location data, it's possible to imagine a chilling effect on legal, but stigmatized, activities. Would citizens be as likely to attend a political protest, or visit a plastic surgeon, if they knew their location was visible to others? Large databases of location data accessible by subpoena also could become evidence for minor disputes and civil court cases.

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 (*Personal Privacy in an Information Society: The Report of The Privacy Protection Study Commission*, 1977; U.S. Department of Health, Education, and Welfare, 1973).These codes are still considered a gold standard for privacy protection (Waldo, Lin, & Millett, 2007). 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. Individual developers or community groups might create participatory sensing applications and begin collecting personal data (Zittrain, 2008).

We need a nationwide discussion about when and how to share this new form of personal data. Currently, corporations such as mobile carriers as well as small-scale application developers are struggling with how best to provide privacy protections for participatory sensing data. One possible solution is encouraging personal tools and sensing architectures that support individual control over sensing data. Open and privacy preserving systems can create a level playing field in which public good and market innovation flourish, as we have seen in the development of the Internet. Several research labs are currently working on architectures which would provide essential cyberinfrastucture to accelerate participatory sensing while building in privacy from the outset. The commonality in these approaches individually-controlled secure data repositories we call Personal Data Vaults (PDVs). The PDV decouples the capture and archiving of personal data streams from the sharing of that information. Instead of individuals sharing their personal data streams directly with services, we propose the use of secure virtual vaults to which only the individual has complete access. The Personal Data Vault facilitates the selective sharing of subsets of this information with various services over time. Selective sharing may take the form of exporting filtered information from specific times of day or places in space, or may import service computations into the data vault and export resulting computational outputs. Tools for data owners to audit information flows are also essential to support meaningful usage, and are a critical part of vault functionality. These vaults, which could be made available to any interested individual as a public or private service, would provide secure archives of user-contributed data, and offer tools for managing and sharing subsets of that data for use by community groups, researchers, or health practitioners, according to specific filters approved by the individual on a per-service basis. The PDV construct is fundamentally a software function that (a) provides persistent, highly-available storage and management for spatiotemporally-tagged data, and (b) implements controlled sharing on behalf of the data owner.

But questions remain. Who will offer and manage data vaults? And will citizens adopt their use? Creating a business model for the data vault that does not rely on mining location data is a central unmet challenge. Regulations and mandates to encourage participatory sensing application providers to contract with vaults might be one way to support the adoption of such infrastructure. National or state financial incentives to develop and secure such vaults might be another. A second challenge is introducing greater transparency into the world of mobile services to which personal data vaults connect. A voluntary or regulated system of application labels could help sensing 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, 2002; Dodge & Kitchin, 2007). 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.

In addition, we need legal mechanisms to protect this data and encourage individuals to participate in sensing without fear for privacy or liability. For example, diaries – currently the pen-and-paper analogy for much of personal sensing data – are discoverable. How do we build a basis for automated, prompted self analytics to be treated with a stronger legal privilege? If raw location data and experience sampling 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. Wrapping the data stored in a PDV in an evidentiary privilege, similar to the non-commercial trade secret privilege, would mean that none of the data stored in the Vault could be subpoenaed or introduced into any legal proceeding. Some exceptions might apply, such as the "crime/fraud" exception to attorney-client privilege. But the protection would provide a currently unavailable promise that personal data would not harm a person's job prospects or civil liabilities. Such a privilege could be recognized by state judge application and extension of the common law. Some analogies can be found, for instance, in the recognition of a self-evaluation or self-critical analysis privilege in certain states. Alternatively, state legislatures could pass a statute creating the privilege, as some have done for medical committee reports. If this seems politically unlikely, recognize that we would need only one state to act as a first mover.

In closing, there is tremendous power in the secondary use of mobile phone and locative technologies for research, healthcare, and community building. But to recruit the participation necessary for these technologies to prosper, individuals must be persuaded that very sensitive data will be protected by both law and technology. The current privacy framework in the United States, emphasizing notice and consent and distributed, unregulated data collection, will not support such innovation. New protections to encourage participation and long-term engagement with data control are needed to encourage participatory sensing.

References

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

Anokwa, Y., Borriello, G., Pering, T., & Want, R. (2007). A User Interaction Model for NFC Enabled Applications. Presented at the PERTEC 2007 Workshop on Pervasive RFID/NFC Technology and Applications.

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

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

Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., & Srivastava, M. B. (2006). Participatory sensing. In *World Sensor Web Workshop, ACM Sensys 2006*. Presented at the World Sensor Web Workshop, ACM Sensys 2006, Boulder, CO: ACM.

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).

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.

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.

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 The 4th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, Philadelphia, PA: ACM.

Kang, J. (1998). Privacy in cyberspace transactions. *Stanford Law Review*, *50*, 1193-1294.

Miluzzo, E., Lane, N. D., Eisenman, S. B., & Campbell, A. T. (2007). CenceMe - Injecting Sensing Presence into Social Networking Applications. *Lecture Notes in Computer Science*, *4793*, 1-28.

Nissenbaum, H. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford Law Books.

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

Pentland, A., Lazer, D., Brewer, D., & Heibeck, T. (2009). Using reality mining to improve public health and medicine. *Studies in Health Technology and Informatics*, *149*, 93-102.

Personal Privacy in an Information Society: The Report of the Privacy Protection Study Commission. (1977). Retrieved from http://epic.org/privacy/ppsc1977report/

Shilton, K. (2009). Four billion little brothers?: privacy, mobile phones, and ubiquitous data collection. *Commun. ACM*, *52*(11), 48-53.

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.

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