Information Producers, Information Consumers: Location Data Privacy in Institutional Settings

Submission Date: 26 July 2013
Word Count: 7,403

Author:

Dr Caitlin D. Cottrill (Corresponding Author), Lecturer
University of Aberdeen
School of Geosciences
Elphinstone Road
Aberdeen, Scotland, UK AB24 3UF
Phone: +44 (0)1224 273694
c.cottrill@abdn.ac.uk

Abstract:

The move towards ubiquitous computing, particularly in the realm of transportation services, has resulted in the creation of vast data sets. Data sets such as those created through RFID-tagged goods and services, cellular and smartphone use, mobile and location-sensing applications (or "apps"), and GPS navigation devices may contain large stores of personally identifying information, or information that may be personally identifying in the presence of related data sets. Current practices regarding the collection, storage, sharing, and dissemination of these data sets may not adequately protect the privacy of persons on whom data have been collected. This paper endeavors to, first, outline the privacy concerns underlying the locational data sets increasingly becoming publicly available. Next, we review the privacy preserving practices that may be used by persons in the "data chain" to ensure protection of private data in the areas of data collection, storage, sharing and use. Finally, a process by which recommended practices reviewed here may be implemented in an institutional setting is presented in order to frame a pragmatic understanding of the responsibilities of various actors with respect to protecting the privacy of individuals. It is hoped that this review will provide a framework for enhanced understanding of the implications of increased locational and personally identifying information collection on the need to responsibly manage this sensitive data.

INTRODUCTION

In 2011, the LIVE! Singapore research group presented an exhibition at the Singapore Art Museum which provided a showcase of visualizations of data generated by people and activities in Singapore. According to the press release, "Employing real-time data recorded and captured by a vast system of communication devices, microcontrollers and sensors commonly found in our urban environment and mapping this information onto multi-dimensional maps of Singapore, this geographical information system merges cartography, statistical analysis and data platform technology (*1*)." The projects that were shown by the researchers included the following data:

- Movement of crowds, taxis and airline passengers;
- Microclimate conditions;
- Electricity consumption; and
- Shipping containers (*2*).

The diversity of data seen here provides an indicator of the types of information that are now available to researchers, practitioners, and, to some extent, the general public in the urban environment. The vast array of data collection agents positioned across the urban landscape have made it possible to bring together heretofore disparate concepts and data streams and combine or mine them to create information and discern patterns in space and time. The ability to collect and review these data in real-time adds an additional layer of value to these data streams, as the ability to watch events and activities as they occur increases the ability of planners, engineers, and policy-makers to respond to events as they transpire, rather than endeavoring to respond after the fact. The value that such data have added to the planning and policy framework is nearly incalculable, and will continue to grow as computer and sensor networks become ever more ubiquitous on the urban landscape.

Such technological advances and changes have undoubtedly produced a sea-change on the transport planning terrain. Whereas traditionally planners have been limited in their ability to collect data from a constrained number of users of the transport network via such methods as household travel surveys, inductive loop counters, or visual counts, the expansion of data creation methods has made it theoretically possible to collect data on nearly all travelers in the network via the use of data use records from cell phones, automated number plate recognition (ANPR) systems, traffic cameras, electronic transit fare cards, and various other sources. Implications of these innovative data collection methods for planning have been numerous. From travel diaries and four-step modeling, we have moved into activity- and agent- based models that rely upon large amounts of computing

power to model ever finer detail on the travel behaviors of individual agents, freight, and transit modes.

Such advances have enabled transportation planners and developers to make more accurate and considered estimations and decisions regarding transportation investments and projects; however, the amount of data collected and available in the public sphere have opened up a wide range of questions and concerns regarding the potential to use these data in intentionally malicious or unintentionally obtrusive manners. Paul Ohm (2010) has outlined the problem via the use of three examples of data reidentification, where "anonymized" data were either used to identify subjects on their own, or combined with publicly available datasets and de-anonymized: an AOL data release in which "anonymized" search queries were found to be traceable to individual users; an "anonymized" Massachusetts health record release which, when matched with voter rolls, was found to allow for identification of individual state employees; and a release of de-identified Netflix user ratings which, again, nevertheless allowed for identification of individual users (*3*). These three examples provide a clear picture of the difficulties faced when working with data sets in the public realm. The wide variety of data currently available in the public sphere, including Census data, property records, residential location data, and others, leaves open the potential for extensive data mining and reidentification of more sensitive data sets. As seen here, the mere exercise of "anonymizing" these data by removing personal identifiers may still leave open the potential for violations of privacy to occur.

In this paper, we will first identify a number of emerging technologies that produce rich data collections in and from the mobile environment, including their uses and associated privacy concerns. Next, we will outline the various actors involved with the creation, dissemination and use of these data sets in the public realm, particularly insofar as their actions may lead to more or less privacy of the data with which they are concerned. The concerns identified here will then be discussed in the context of balancing the protection of data privacy with ensuring some degree of usefulness, by providing a pragmatic approach that will be used to identify concrete concerns and potential responses. Finally, recommendations will be made as to strategies that may be used by planning and research organizations to effectively address emerging privacy needs.

As Ohm (2010) states, "Data can be either useful or perfectly anonymous but never both (*3*)." This tension is acknowledged throughout this paper; however, we also maintain that there are some effective measures that may be taken by researchers and practitioners to protect against some of the more egregious violations of privacy that are possible in the absence of attention and effort.

ADVANCES IN UBIQUITOUS COMPUTING

In 1991, Mark Weiser, then chief scientist at Xerox PARC, wrote, "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it (*4*)." With this and other articles, Weiser introduced the idea of "ubiquitous" (or pervasive) computing ("ubicomp"), which posits that information processing will eventually be seamlessly integrated into everyday objects and activities, with users not necessarily aware that they are participating in the system. Weiser (1991) sums up the goal of ubiquitous computing with the following statement: "Machines that fit the human environment, instead of forcing humans to enter theirs, will make using a computer as refreshing as taking a walk in the woods (*4*)." Such a goal is laudable, in many ways, with ubiquitous computing technologies already taking shape in the form of such technologies as smartphones, Global Positioning Systems (GPS), and radio frequency identification (RFID) tags. These technologies add much value to our experience of the world around us; however, each also comes with concomitant issues regarding privacy and security. The following section will outline the benefits and concerns associated with each of these technologies.

Radio Frequency Identification (RFID)

RFID technologies are widely used in such areas as transportation (electronic toll passes and transit fare cards), supply chain monitoring (such as by Unilever, Chevrolet, Ford Motor Company, and the Port of Singapore (*5*)), passports (under the United States Visa Waiver Program, for example, which requires either a machine-readable or e-Passport), and inventory systems.

While these uses increase the convenience of travel, transport, and tracking, a number of concerns have also arisen. Meingast, *et al.* (2007) outline the following primary concerns associated with the growing use of RFID tags:

- Ubiquity: Transponders are permanently embedded into commonly-carried objects, thus making the transponder ever-present;
- Linkability: Data stored on transponders is static and may be linked to an individual;
- Unawareness: The user may be unaware of the presence of the transponder, or the transponder may not clearly indicate to the user by whom or when it is being read; and
- Security: Transponder-embedded objects are used in the public sphere, where they may be accessed by unauthorized entities without the knowledge of the individual (*6*).

These issues, in turn, raise the possibility for a wide variety of malicious security and privacy attacks, including:

- Rogue scanning and eavesdropping: Unauthorized access to data that has been stored on the tag or is sent via air interface;
- Replay attack: Repetition of an authentication sequence captured by an attacker for the purposes of identify usurpation
- Duplication: Duplication of a tag with the same functionality by an attacker;
- Malicious modification of the tag's memory content: Used to change reported tag attributes or to distribute malware; and
- General attacks: Blocking, frequency jamming, side channel attack, and backend attacks (*7)*.

In addition, privacy losses or leaks may also be caused by negligence or ignorance.

Smartphones and Applications

Cellular phones are some of the most ubiquitous technologies in the world today, and smartphones (defined as mobile phones with advanced computing capabilities) are quickly gaining a large share of the market. According to Gartner (2011), "Worldwide mobile device sales to end users totaled 1.6 billion units in 2010, a 31.8 percent increase from 2009... Smartphone sales to end users were up 72.1 percent from 2009 and accounted for 19 percent of total mobile communications device sales in 2010 (*8*)." According to some estimates, adoption rates for iOS and Android devices have surpassed that of any other historical consumer technology (*9*). The exploding market in smartphone-based applications (or "apps", generically defined as small programs, particularly those designed for mobile devices such as smartphones or tablet computers) is one of the predominant drivers of smartphone adoption. In a special report, The Economist (2011) reported that, "The appetite for apps appears insatiable: Gartner… estimates that almost 18 billion have been downloaded since the first app store was opened by Apple in 2008. By 2013, it thinks, the number will have risen to 49 billion (*10*)." App categories range from games (such as Angry Birds), to information finding (Google maps, Yelp), news and media (The Economist, the New York Times), social networking (Facebook and Foursquare), travel, sports and beyond. The proliferation of apps has been aided, in part, by the adoption of open source software for such platforms as Google's Android phone, as well as increasing numbers of programmers and application developers, shifts that indicate that such platforms will likely continue to thrive.

While increasingly popular and often useful, mobile phones, smartphones, and apps bring with them new concerns in the realm of privacy protection. For example, Gralla, *et al.* (2011) report that:

> Apps can trace your Web habits, look into your contact list, make phone calls without your knowledge, track your location, examine your files and more. They can also automatically send information such as location data to mobile ad networks. In addition, apps can gather the phone number and the unique ID number of each type of phone…Personal information that apps gather about you can be matched to these IDs. That means that ad networks can easily combine various pieces of information collected by multiple apps, build a sophisticated profile about your – and then legally sell that data to other marketing companies (*11*).

In addition, Efrati, *et al.* (2011) report that federal prosecutors in New Jersey issued a number of subpoenas to various app makers related to a federal grand-jury investigation of app information-sharing practices. According to the article, "The Journal tested 101 apps and found that 56 transmitted the phone's unique device identifier to other companies without users' awareness or consent. Forty-seven apps transmitted the phone's location in some way. Five sent a user's age, gender and other personal details to outsiders. At the time they were tested, 45 apps didn't provide privacy policies on their websites or inside the apps (*12*)." Even for those apps that provide privacy policies, however, the likelihood that the consumer will read them is uncertain. The US FTC has reported that, "Although many companies use privacy policies to explain their information practices, the policies have become long, legalistic disclosures that consumers usually don't read and don't understand if they do. Current privacy policies force consumers to bear too much burden in protecting their privacy (*13*)." Such a statement is supported by findings presented in (*14*).

On the other hand, data collected via use of these apps may potentially be of great benefit to transportation and other planners and engineers. Both by direct development of apps by local governments, planning agencies, and transportation service providers, as well as by collaborating with developers of other apps of interest, increasing amounts of real-time and crowdsourced data may be collected, providing an enhanced toolkit for maximizing the potential efficacy of planned urban improvements. Access to more detailed and accurate data streams has the potential to ensure that planners and others are responding to the needs of a greater variety of persons within the urban environment, as well as allowing for more targeted and timely responses to immediate needs. With these possibilities, it is likely that government and other civic agencies worldwide will soon begin to focus more time, staff and resources on developing smartphone-based technologies that will provide continuing and comprehensive data streams on the actions and events of persons and places.

The growing adoption of smartphones, wide dispersal of potentially privacy-invasive apps, and lack of consumer knowledge of privacy procedures, combined with likely future investment in and use of these technologies by government and public agencies highlights the need to ensure that privacy concerns are adequately and comprehensively taken into account. Ensuring the protection of the data of citizens, residents, and visitors will require a combination of technology- and policy-based means of protection, and will also require that associated staff are given adequate training and support in privacy protection matters.

GPS and Location-Based Services

The advent and proliferation of global positioning systems (GPS)-enabled devices (such as cellular phones, GPS receivers, in-vehicle navigation systems, and others) has greatly expanded the availability of location-based services (LBS), defined as "information services accessible with mobile devices through the mobile network and utilizing the ability to make use of the location of the mobile device (*15*)." A variety of types of such services have been developed, including orientation and localization, navigation, search, identification, and event check (*16*). One particular area of note where GPS services have begun to make strong inroads is in the realm of household travel surveys, which generally consist of demographic data on the person and his or her household, as well as a travel "diary" that asks participants to record their daily travel information, including origins and destinations, modes, and times. Traditionally conducted via paper and pen, many recent implementations have incorporated a GPS component in order to collect data that are more accurate in terms of route information, trip timing, and actual trips.

GPS data and data sources, however, hold the potential for severe privacy issues. The low cost associated with many GPS devices has made them accessible both for government agencies as well as private concerns (be they individuals or firms), and there is little overall consistency or knowledge of how their use should be treated. For example, the recent Supreme Court finding in *United States v. Jones* (2012) directly addressed the following questions: "(1) Whether the warrantless use of a tracking device on respondent's vehicle to monitor its movements on public streets violated the Fourth Amendment; and (2) whether the government violated respondent's Fourth Amendment rights by installing the GPS tracking device on his vehicle without a valid warrant and without his consent (*17*)." Such questions have also arisen over the placement of a GPS device on a subject's car by a private actor, for example, for stalking purposes (*18*). The collection of location information via these methods allows for a detailed portrayal of an actor's habits, proclivities, and activities, and the storage of such data over time may paint a more vivid picture of a person's overall being than would commonly be thought of as privacy preserving.

Conclusions on Overview

As shown in the overview above, advances in ubiquitous computing have led to the creation and potential creation of vast stores of real-time, location-tagged and personally identifying data. While such data sources may be used to great advantage for transport planning and decision-making, they may also open up users and data managers to a wide variety of potential privacy invasions and data management concerns. In response to these concerns, a number of data management and privacy policy approaches have been suggested. The next section will provide a concise overview of privacy preserving actions and approaches that may be used by various actors in the data chain. Approaches recommended here will then be reviewed to demonstrate a pragmatic approach to privacy preservation that provides an overall balance between ensuring usefulness of location data and keeping personal details protected to a reasonable extent.

APPROACHES TO PRIVACY PRESERVATION

Advances in ubiquitous computing, as well as increasing amounts of attention being given to issues of privacy management in the location environment, have generated increasing amounts of research interest in the area of technical approaches to privacy and confidentiality preservation. Research centers such as the Data Privacy Lab and the Electronic Privacy Information Center (epic.org), academic groups such as Carnegie Mellon University CyLab's Usable Privacy and Security Laboratory (CUPS), and industry divisions such as IBM's Privacy Research Institute and Microsoft's research group in Security and Privacy have worked towards developing solutions for striking the balance between privacy and usability such that data may be mined or applied usefully while not impinging on the privacy of persons on whom data have been collected. Such research has taken place at several functional levels, including that of data creators, data collectors, data users, and data sharers. While it is obvious that different entities may serve different functions at different times, it is necessary to recognize that privacy preservation should be approached via different methods depending upon both where privacy is to be addressed and the intended uses of collected data. Acknowledging the role(s) played by the entity of interest will greatly impact the decisions made relevant to privacy protection and preservation. This section will present an overview of recommended approaches to protection and management of private data from each of the viewpoints (creator, collector, user, and sharer) identified above.

Data Creators

At the most basic level, no data management is necessary unless there is data to manage; thus, the first agent of concern in the data chain is the data creator, or the subject of the private information.

The primary mechanisms given to data creation agents in the realm of privacy protection are those of "opting out" and "opting in" to participation in services or applications. These approaches, however, bring with them some controversy, as there is disagreement over whether "opt-in" or "opt-out" provisions are preferable for balancing usability and anonymity. According to Bouckaert and Degryse (2005), "Opt-in permits use of personal information within the organization but requires the opt-in consent before personal information could be disclosed to third parties outside the organization (*19*)." Opt-out requirements, on the other hand, as characterized by Lacker (2002), require that an organization, "…that intends to share non-public customer information with third parties (companies not related by ownership ties) must give customers an opportunity to deny them permission to do so, or opt out (*20)*." There is little consistency in the use of opt-in or opt-out methods by businesses, governments, or industry groups, though a consumer's use of a product or service often includes a *de facto* assumption that he or she has "opted-in" to the applicable privacy policy (if one is provided) and practices. According to Rapp, *et al.* (2009), "Privacy advocates prefer opt-in methods, under the belief that explicit permission should be the norm so that potential customers feel less vulnerable to and alienated from firms with which they transact business (*21*)." Included here is the implicit belief that consumers also tend to prefer opt-in regimes, as they must actively and explicitly choose to participate in data collection activities. Opt-out and the noted d*e facto* opt-in practices, on the other hand, tend to be preferred by product and service providers, perhaps because they default to the sharing of potentially valuable information.

Privacy policies are the most common method by which consumers are notified about their privacy rights and responsibilities pursuant to use of a particular product or service, but they are not always required from providers, particularly those in the private realm. For those that do provide notices, many polices explicitly state that the consumer is responsible for staying informed about privacy practices and changes to privacy policies (*22*). Given the relatively low percentage of persons who report reading privacy policies at all (a February 2009 study by the UK Information Commissioner's Office found that 71% of people do not read or understand privacy policies (*23*)), the likelihood that persons will follow up on potential changes to policies is fairly low. A further concern is that privacy policies are generally written at a reading level that is above that of the average consumer (*22*). Given these constraints, it may be argued that privacy policies are not the most effective way by which data collectors may communicate their privacy practices to data creators.

Data Collectors

While privacy policies are the primary way in which a firm or organization's data protection practices are conveyed to the user, they are not generally written primarily for the consumer. Rather, "Privacy

statements generally serve two major purposes: to mollify consumers wary of conducting transactions online for fear of privacy violations; and to convince regulators that further legislative initiatives to guarantee consumer privacy are unnecessary, since the industry self-policing efforts sufficiently protect citizen rights (*24*)." In short, privacy policies are often written as much to provide coverage and protection to companies and organizations, which may be reflected in vague or overly-inclusive language within policies, as they are to protect the consumer. While this is understandable from the point of view of data collectors, it is less useful from the point of view of consumers. In this context, privacy policies are useful for describing the general approach to privacy preservation that an organization may take, but internal documents and procedures may provide a more accurate picture of steps actually taken by collectors.

Beyond the privacy policy, a number of internal procedures, both technological and administrative, may be used for the protection of private data by collecting agencies. The US Department of Commerce's National Institute of Standards and Technology's (NIST) draft document, "Security and Privacy Controls for Federal Information Systems and Organizations," provides a comprehensive overview of "privacy controls" (defined as, "…the administrative, technical and physical safeguards employed within an organization to protect [personally identifiable information (PII)] (*25*)."). Such controls may include documents, such as privacy notices and policies, as well as administrative procedures such as determining the legal authority to collect data, minimizing the collection and retention of PII, limiting access to PII, and conducting privacy audits. Such methods may be viewed as supporting the more general privacy policy, and providing more direction to agencies charged with their fulfillment. These procedures, in turn, are supported by technological processes – such as the use of cryptography, password protection, and data encryption – that are used to provide physical safeguards for the protection of data. Of note is that many privacy policies aim to be "technology neutral," or non-prescriptive regarding the technologies used to meet their constraints. The use of this approach allows for flexibility as new privacy-preserving technologies are developed and deployed.

The profusion of methods that data collectors may use to manage the privacy protection of personal data is a double-edged sword. On the one hand, collecting agencies are given wide scope to determine which procedures most make sense for protection under their management rubrics; for instance, an agency that collects and stores large quantities of personally identifying information will need stronger policies related to technological and policy-based methods of privacy protection than an agency that collects predominantly non-personal information, which may simply need documents indicating that appropriate measures will be taken if personally identifying information is collected.

There is, however, a concern that emerges; namely, that the plethora of options may lead to confusion and inconsistency. For example, if access management is addressed in more than one document and does not consistently identify who will have access to collected data under what circumstances, an agency might find itself liable for legal action if data are incorrectly released. Issues might also occur if technological methods of protection, such as perturbation, masking, or cryptography, are used without providing accurate and appropriate metadata. In this case, it is possible that the accuracy of the data could be compromised, particularly if it is intended to be combined or mined with other data sets. These and other issues make a strong case for coordinated planning of technological and policy-related privacy protections to take place within agencies that collect original data.

Data Sharers

A number of privacy concerns are raised when data are shared between agencies or organizations, a practice that is becoming increasingly common with the advent of cloud computing, the rise in public-private partnerships, and increasingly large numbers of private location-sensing application developers. In essence, a large number of agencies and organizations are gathering data that may be leveraged by other agencies and organizations for purposes ranging from planning to marketing to law enforcement, and technological developments have greatly increased the ease of sharing of these datasets. While beneficial to agencies and organizations, such a shift has, once again, revealed additional issues related to the protection of private information that may be contained in these datasets. For example, policies related to tertiary uses of data shared between agencies may not be clear, and may result in accidental or malicious release of sensitive data. While such a potential occurrence may be avoided by judicious use of non-disclosure agreements (NDAs), the development and application of such documents may be insufficient for protection purposes if they are not properly applied. In general, it will be necessary for those entities sharing data to ensure that policies and procedures guiding their treatment of data are consistent and, preferably, meet the needs of the more restrictive entity. In this situation, clear practices regarding data access may be the most critical component of mutual agreements between data sharers. Ensuring that appropriate records are kept regarding who has accessed data, for what purposes, and with what results, will provide appropriate traces for review in the event that data are misused or mishandled.

Data Users

To some extent, all of the actors identified above (creators, collectors and sharers) may be regarded as data users; however, the requirements for privacy-preserving use of data are significantly

different from those of its collection, management, or storage. The use of data implies that either it or its derivatives will be available in the public sphere, at which point management and protection aspects take on new degrees of concern. Here, technological considerations may be the most pressing concerns in two areas; namely, the ability to re-identify data (as outlined in the introduction), and the need to protect such data via technological methods such as consistently secure storage. Here, one may look to the US Census Bureau for a very comprehensive approach which relies on a trio of protections, including:

- Federal regulation (specifically, Title 13 of the U.S. Code)
- Internal privacy principles
- Statistical safeguards

The Census Bureau's use of these protections immediately reduces the privacy concerns of users of their data sets, as accidental or intentional disclosure of personally identifying information is difficult due to the required protections. It is evident, however, that in the absence of such precautions, data users must be cognizant of the potential for infringement of privacy policies of those agencies and organizations from which data have been obtained.

Depending upon the level of detail of the data sets obtained by outside agencies (for example, raw versus processed data), the procedures needed for appropriate handling will vary. For example, if raw data traces collected via RFID-enabled electronic transit cards are shared with an outside agency, it is critical that the third-party user should ensure that these data are, first, treated in accordance with the policies of the sharing agency; and, second, not mined or combined with other data sets such that individual identification would be possible. Some precautions that may be taken include removal of unique identifiers and aggregation of trips and travelers into more generic flows; such actions, and the public presentation and use of data, will greatly lower the possibility of violating privacy expectations. In contrast, if more generalized data are to be used, there may be less concern encountered, as the potential for accidental or malicious privacy invasion may be smaller, though (as noted above) still of concern. In either case, where data, whether generic or detailed, are to be mined or combined with other data sets, it is critical that the resulting data sets be evaluated for potential privacy infringement.

Conclusions on Privacy Preserving Practices

In general, there are two primary concerns in play when discussing how data should be treated: on the one hand, protecting the privacy of the data creators is of utmost importance, on the other hand, it is also critical to ensure that the privacy policies of the sharing agencies are not being

violated, as this may cause negative repercussions both legally and in terms of agency trust. While there are distinctions in approaches that should be taken by various actors in the data chain, there are also a number of actions that should be taken by all actors, including:

- Creating appropriate metadata indicating how the data set was created or obtained, information on any personally identifying information, and any other information that may prove germane to how the data are treated in the future.
- Obtaining copies and following the guidance of privacy policies, non-disclosure agreements, or other relevant documents for all agencies or organizations associated with a particular data set.
- Developing data access plans to formalize direction regarding who does or does not have access to what data sets, for what purposes, and whether they may disclose these data to others.

Generally following these guidelines will provide a first step in ensuring that the privacy of data is not compromised.

A PRAGMATIC APPROACH

The combination of extensive location data collection technologies and a growing concern related to data privacy as outlined above indicates that the time is ripe to formalize pragmatic approaches to privacy preservation within the transportation arena. Here, we use the background provided above to recommend a process by which transport agencies may better institutionalize pragmatic privacy practices within their organizations given access to data from governmental organizations, private industry, and self-collection. Here, the primary focus will be on developing a framework by which both technical and policy aspects of privacy protection of spatiotemporal data may be addressed. Balancing these needs in a manner by which adequate privacy protections are achieved while not overly restricting data usefulness requires a holistic view of the data structure of an organization, in addition to a thorough understanding of the competing interests of associated agencies.

There are often close association between public and private agencies and research organizations, and it is critical to ensure that each organization type is familiar and comfortable with the necessary data protection regulations that are applicable to gathered, shared, and utilized data sets. One key component of this approach is the use of non-disclosure agreements (NDAs - legal contracts between parties that provide an outline of those shared materials that should be treated as confidential and are restricted from sharing with third parties). The use of this tool is effective in a number of ways, two in particular being that they provide legal recourse in the case of accidental or

malicious release or use of data and they provide an opportunity for collaborating agencies and organizations to negotiate their expectations regarding both the degree of collaboration and related responsibilities. Common elements can include:

- Expectations and understanding of what will and will not constitute confidential information under what circumstances and how this information should be treated;
- Obligations of each party subject to the NDA;
- Duration of the agreement and data availability; and
- Any other relevant information not covered in the above.

While the development and enforcement of NDAs can be time-consuming, they do provide a context for clearly defining the expectations regarding data privacy and confidentiality between partnering organizations, and may help more clearly set forth expectations regarding data management structures.

The development of an appropriate access plan is a key element in effectively managing the flow of data and information with respect to implemented NDAs. Close connections between public, private, and research organizations make this a critical point, as determinations of how data should be treated under established NDAs will rest in large part upon who has access to data, in what form, and for what purposes. Such concerns will become especially relevant as, for example, electronic transit fare payment cards are made interoperable with other systems (for example, the Singapore EZ-Link card or the Chicago Transit Authority's Ventra card). While basic transit use records are generally fairly innocuous, the potential linking of additional data – such as parking fees paid, road tolls, or other purchases – to these records opens up the potential for myriad privacy invasions. If benefits are to be gained from such data sources, it is necessary to pay close attention to what entities are involved, what portions of the overall data may be useful to these entities, and how collected data may be segmented and protected in order to ensure that privacy is not compromised.

The first issue to address in this example is that of what data will be necessary for what uses. Assuming here that electronic fare payment information will be used for transportation research purposes, it is unnecessary to obtain records on such card uses as retail purchases or the balance of the card's value; however, records pertaining to transit boarding and alighting and, potentially, payment of taxi fares will be quite relevant. By identifying fields relevant to the purposes for which data have been requested and removing unrequired data, the amount of unneeded and potentially invasive data that are shared by agencies and organizations is minimized, and may also mollify consumer concerns about privacy by indicating a respect for the context of data use. We assume

here that data obtained from external parties will be used in conjunction with data that will be collected by the agency of interest itself, which will also impact the amount and type of data needed. When mining or combining datasets from multiple sources, it is particularly necessary to evaluate the potential for identification of participants. In this case, with adequate protection of data from release (as described further below), the combination of multiple data sets should not prove detrimental to the privacy of data creators; however, when preparing the data for public release, records should be carefully evaluated to ensure that they are adequately protected from revealing personally identifying information. This is particularly critical for location data, as it has been shown that it is nearly impossible to anonymize location traces (*26*)

The second step in this process, related to the above, is that of access management. Establishing an access plan that reflects the structure of data need allows useful data sets to be obtained by those entities for which they will add value, while limiting the possibility for accidental or malicious use or release of sensitive data. Data management plans should generally provide the first cut for such access determinations, but these should be updated regularly in recognition of changes in technologies, staff, projects, and data sources in order to ensure that practices agreed upon are being followed and updated to reflect actual circumstances. In addition, it may also be necessary to establish additional internal guidelines regarding technological protections and specifics regarding who will control access to the data sets within teams and between the team and the public. For technological protections in the example here, it will be necessary, first, to process the data to ensure that no personal identifier (such as name or unencrypted ID number) is associated with a given record. One key point is that, for maximum usefulness, it may be necessary to maintain the relationship of individual points across each individual entity; however, such data can be privacy invasive. Access management combined with technological methods of protection, however, provides a way in which both needs may be addressed. In this instance, certain persons (those most heavily involved with modeling needs, for example) may have access to raw data with encrypted identifiers maintained across all records. These persons, in turn, may either assist persons who do not have data access privileges with those projects requiring access, or use masking, aggregation, or perturbation of data to allow some internal access to useful data at the required level of detail. It is here that data protection needs come into play. Depending upon the sensitivity of the data, a number of different methods may be used for protection, including simple password protections, data encryption, and secure servers. The combination of securing sensitive data with technological protection and limiting access may do much in the way of promoting privacy preservation.

With appropriate data access and security plans in place, public visualization of data in a privacy-preserving manner should be fairly straightforward. For public display purposes, data should be aggregated such that it is not possible to identify individual agents across time and space, a process which should have begun with those persons not provided with access to the full dataset. In some cases, and as identified under the terms of the NDA, those agencies which have provided data should be given the ability to review some publicly displays of shared data to ensure that they do not run counter to the purposes for which they have been shared. In addition, any public display of data created from human agents should be reviewed to ensure that it will not cause discomfort on the part of those who have been the witting or unwitting subjects of data collection.

This last issue that we will address here in privacy preservation ties into the responsibilities of agencies that use personal data and information to ensure that the data creators (those persons on whom data have been collected) are both aware of and comfortable with that information that has been gathered and is being used. While many data creators will pay little attention to the fact that data are being collected, if they are made aware of such collection due to media attention or data misuse or mishandling some may reject the use of those technologies or practices that lead to its availability. A number of methods may be used to mitigate these concerns, including following the guidelines outlined above to minimize the possibility that data will be improperly accessed or used. A second practice is to have in place policies that will be implemented in the case that data are improperly released or used in order to mitigate the negative impacts of such releases. Finally, as outlined in many fair information practices, it is necessary to have in place policies that allow consumers (or data creators) to review and correct, amend, or remove data that they feel are incorrect or do not wish to be made available for research or other purposes. Having such policies in place and making them available to consumers both provides a sense of security and trust to data creators, as well as protecting those agencies that collect and use data from the negative ramifications of not having policies in place whereby consumers may take an active role in the protection of their privacy.

Conclusions on Approaches

The approach outlined above presents a pragmatic method for addressing privacy concerns within an organization that uses both internally and externally created location data. It addresses concerns outlined in the description and overview of emerging technologies and associated data collection, as well as those described in the review of associated actors. The objective here has been to the practical issues that arise when attempting to consider privacy protection in the use of extensive location datasets. By reviewing both the types of data collected and the different actors involved, it

is possible to more clearly define the process by which data management for privacy preservation may be approached in a manner that is effective while not overly burdening any one agency.

CONCLUSION

In this paper, we have attempted to draw attention to the privacy risks presented by the multitude of emerging data sources in the mobile environment, as well as identifying the implications to and concerns of various affiliated actors and agency types. In essence, data-generating technology uses have developed at a rate that far outpaces the development of policies to respond to potential concerns. While a number of approaches have been proposed to begin to address this issue, it is critical at this point to provide information and suggestions in order to encourage early adoption of recommended methods, both to promote the immediate protection of sensitive data, as well as to set in place an approach that will dovetail smoothly with regulations once enacted. The key objective is to maintain a large degree of usefulness and functionality of data while ensuring that procedures are in place to protect its privacy to the maximum extent possible. Here, we recommend a hybrid mixture of inter-agency agreements, technological protections, access management, and statistical protection in order to allow for maximum use at the agency level while providing consumer with essential protection. It is hoped that such an approach would prove practical to a number of identified agency types.

References

1. SENSEable City Lab. "The Real-Time City is Now Real." SENSEable, 2011. Available at http://senseable.mit.edu,lLast accessed October 2012.
2. SMART, "LIVE Singapore! - The pulse of the city in real-time," Press release. April 6, 2011. Available at http://smart.mit.edu.
3. Ohm, P., "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." *UCLA Law Review*, 2010, Vol. 57, p. 1701-1777; 2010; U of Colorado Law Legal Studies Research Paper No. 9-12.
4. Weiser, M. "The Computer for the Twenty-First Century," *Scientific American*, September 1991, pp. 94-10.
5. Angeles, R., "RFID Technologies: Supply-Chain Applications and Implementation Issues," *Information Systems Management*, 2005, 22:1, pp. 51-65.
6. Meingast, M., J. King and D. Mulligan, "Embedded RFID and Everyday Things: A Case Study of the Security and Privacy Risks of the U.S. e-Passport," *IEEE International Conference on RFID*, March 2007, pp. 7-14.
7. Rotter, P., "Security and privacy in RFID applications" in C. Turcu (ed.) *Development and Implementation of RFID Technology*, I-Tech 2009, pp. 237-260.
8. Gartner, "Gartner Says Worldwide Mobile Device Sales to End Users Reached 1.6 Billion Units in 2010; Smartphone Sales Grew 72 Percent in 2010," press release, February 9, 2011. http://www.gartner.com/it/page.jsp?id=1543014.
9. Farago, P., “iOS and Android Adoption Explodes Internationally.” The Flurry Blog, Aug. 27, 2012. Accessed 25 July 2013, http://blog.flurry.com/bid/88867/iOS-and-Android-Adoption-Explodes-Internationally.
10. The Economist, "Apps on tap: The beauty of bite-sized software," Oct. 8th, 2011. Available at http://www.economist.com.
11. Gralla, P., A. Sacco and R. Faas, "Smartphone apps: Is your privacy protected?" *Computerworld*, July 7, 2011. Available at http://www.computerworld.com, last accessed 15 March 2013.
12. Efrati, E., S. Thurm and D. Searcey, "Mobile-App Makers Face U.S. Privacy Investigation," *The Wall Street Journal*, April 5th, 2011.
13. Kaplan, P. and C. Olsen, "FTC Staff Issues Privacy Report, Offers Framework for Consumers, Businesses, and Policymakers," Dec. 1st, 2010. Available at http://www.ftc.gov, last accessed 15 March 2013.

14. Cottrill, C. and P. Thakuriah, “Protecting Location Privacy: Policy Evaluation,” *Transportation Research Record: Journal of the Transportation Research Board,* 2011, 2215, pp. 67-74.
15. Steiniger, S., M. Neun and A. Edwardes, "Foundations of Location Based Services," Lecture Notes on LBS. University of Zurich, 2003.
16. Reichenbacher, T., Mobile Cartography - Adaptive Visualisation of Geographic Information on Mobile Devices. (PhD). 2004.
17. U.S. Supreme Court, "United States v. Jones," Docket No. 10-1259, January 23rd, 2012.
18. Southworth, C. and S. Tucker, "Technology, Stalking and Domestic Violence Victims" *Mississippi Law Journal*, 2007, 76:3, pp. 667-676.
19. Degryse, H. and J. Bouckaert, "Opt In versus Opt Out: A Free-Entry Analysis of Privacy Policies." *CESifo Working Paper Series No. 1831*; CentER Discussion Paper No. 2006-96, Sept. 2006.
20. Lacker, J, "The economics of financial privacy: To opt out or opt in*?" Economic Quarterly*, 2002, 88:3, pp. 1-16.
21. Rapp, J., R.P. Hill, J. Gaines, and R.M. Wilson, "Advertising and consumer privacy: old practices and new challenges," *Journal of Advertising*, December 22, 2009, 38:4, pp. 51-61.
22. Jensen, C. and C. Potts, "Privacy policies as decision-making tools: An evaluation of online privacy notices," *CHI 2004*, 2004, pp. 471-478.
23. Out-Law.com. "Regulators demand clearer privacy policies," Out-law.com, 16 Feb 2009. Available at http://www.out-law.com, last accessed 15 March 2013.
24. Fernback, J. and Z. Papacharissi, "Online privacy as legal safeguard: the relationship among consumer,online portal and privacy policies," *New Media Society*, 2007, Vol. 9, Iss. 5, pp. 715–734.
25. National Institute of Standards and Technology, "Security and Privacy Controls for Federal Information Systems and Organizations," *NIST Special Publication 800-53*, February 2012.
26. Zang, H. and J. Bolot, “Anonymization of Location Data Does not Work: A Large-Scale Measurement Study,” *MobiCom ’11*, Las Vegas, NV.