Geospatial Information Services: Balancing Privacy and Innovation

Nathaniel Jurist Gleicher and Jong Sung Hwang

Korean National Information Society Agency
(100-170) NIA Bldg, Cheonggyecheonno 14, Jung-gu, Seoul, Korea
ngleicher@nia.or.kr; jshwang@nia.or.kr

Abstract
Geospatial Information (GSI) services are a new kind of technological platform that combine accurate mapping services with real-time user location and context data. GSI can power everything from navigation to social and political engagement. Although many GSI services are already in use, little attention has been paid to the implications of the technology and the privacy concerns that it raises. Restrictive regulation could kill the budding GSI industry in its early stages, but some initial steps should be taken to protect GSI users. This Paper defines GSI, identifies the aspects of the technology that contribute to its innovation potential and privacy risks, and surveys possible regulatory approaches. Finally, it recommends a combination of technological anonymity, data retention limits, and user access as a flexible first step toward industry regulation.

Introduction
In 1854, London was reeling from a particularly virulent cholera outbreak. At its epicenter, five hundred people died in the space of ten days. By mapping the location of the infections, Dr. John Snow convinced the local authorities that the epidemic was caused not by noxious vapors, but by the contaminated water of a local well. (Snow 1855) Once the well was closed, the contagion rapidly abated.

To make his point, Dr. Snow displayed data geographically, founding what would later become the science of Geographic Information Systems (GIS). Today, GIS is widely used in medicine, science, and urban planning. Google Flu plots flu-related search queries on a map of the United States, predicting flu outbreaks earlier than the Center for Disease Control. (Google Flu Trends 2009) In a different application, human rights groups such as United For Africa have layered live reports of abuses onto freely available online mapping services, providing visual evidence and testimony to war crimes and human rights violations. (Human Rights Map 2009)

These services use only an approximation of their users’ locations. Often, the user enters his location manually, without verification. Recent technological developments, however, have made it possible for new technologies to leverage a more accurate, automatic understanding of the user’s location and situation.

In the early 1970s, the U.S. Department of Defense developed the first Global Positioning System (GPS). Today, miniaturized GPS receivers in cars and cellular phones have made it easy for users to accurately track their location. Other technologies, such as Radio Frequency Identification (RFID) Tags (Soppera and Burbridge 2006, at 159), perform similar roles. By combining these Real Time Location Systems (RTLS) with GIS and ubiquitous sensor networks, modern innovators are developing new services that rely on Geospatial Information (GSI).

Many GSI applications are already in use. For instance, traffic-monitoring services provide real-time accident data and allow dynamic adjustment of traffic flows to maximize efficiency. (Siegler 2009) Google’s new Latitude service allows users to broadcast their location to their friends, facilitating chance meetings in unexpected places. (Google Latitude 2009) An emerging set of GSI services, commonly called Augmented Reality Services, overlay information onto the picture captured by a mobile phone’s camera, displaying restaurant reviews, real estate offerings, and Wikipedia articles that are relevant to what users see. (Layar 2009, Sorrel 2009) GSI is even being used to monitor the location of school children in Yokohama City, Japan. (Warner 2007, at 853)

Other GSI applications are still in early development. Researchers in government, academia, and business are working to develop “smart cities,” that will deploy ubiquitous computing services to allow municipal workers, emergency responders, and citizens to digitally monitor and interact with their physical and virtual environment. (Hollands 2008) GSI is a central component of the technological platform for these networked urban spaces. It will be fundamental to transportation management, e-government initiatives, and numerous other services.

GSI is also a great boon to emergency services, utilities, and law enforcement, allowing them to quickly reach situations and act more effectively when they arrive. For all of its potential, however, GSI carries very real privacy concerns. In fact, the very same factors that make GSI such an exciting new platform also endanger users.
Headlong development without consideration for GSI’s privacy implications could lead to rampant privacy violations, identity theft, harassment, and even physical harm. Over-regulation, however, could handicap the new industry before it gets off the ground.

Regulators must find a way to encourage innovation and industry development, while also protecting GSI users. This Paper defines GSI and seeks to support regulators with analysis and initial recommendations. First, it identifies the major factors that contribute to GSI’s innovation potential and privacy risks. It then identifies three main avenues for addressing the privacy concerns of GSI: technological anonymization of users, limitation of service data retention, and mandated user access and transparency. Finally, it closes with some initial recommendations for regulation. Rather than outlining a detailed regulatory scheme, this Paper is intended to begin a discussion on GSI and how best to encourage its development while still protecting its users’ privacy.

### Defining GSI

Even more than GIS, GSI bridges the physical and virtual worlds, enabling new services by combining virtually generated data with physical data. While GIS gather static snapshots of data by surveying space, GSI is continuously generated by a network of GSI-enabled devices (GSIN). A GSIN is effectively a distributed platform of sensors that enable GSI services to crowd source real-time data gathering. For instance, a network of drivers whose cell-phones continuously broadcast their location allows Google maps to provide real-time traffic updates. (Berth 2009) While this data can be displayed as a static map, this is only one of many services made possible by a GSIN. The same network of driver locations and speeds, for instance, could be used to automatically route emergency vehicles around clogged highways, or integrated into a city’s traffic light system to encourage faster traffic flow.

GSI’s reliance on distributed networks of devices means that GSI data will frequently be consumed and processed by machines instead of by users. This can already be seen in Google’s traffic reporting service – instead of a single user viewing their location on a map, their data is combined with that of many other users to generate aggregate traffic information. As such, GSI will often be processed by and transported across multiple computer systems. To better understand GSINs, it makes sense to begin with GSI devices themselves.

GSI-enabled devices – be they cellular phones or other mobile or stationary sensors – require at least two kinds of information: their mapped surroundings, and their real-time location. In addition, GSI incorporates two kinds of supplemental contextual information: physical data about the GSI device or its surroundings (for example, ambient temperature and CO₂ level), and virtual descriptive data stored or retrieved from the Internet (such as restaurant reviews and historical information).

![Geospatial Information (GSI) Diagram]

Even GSI that combines only mapping and real-time location can provide useful services, from Google’s real-time traffic data to services that monitor the location of family pets for concerned owners. The increased richness of GSI devices’ contextual information, however, will empower an ever-wider range of services. Many of these will be easy to anticipate – for instance, a GSIN that provides carbon-dioxide data would obviously be useful for emissions monitoring. Others will be difficult to predict, however. For instance, the fact that a majority of GSI-enabled cars in a particular area have engaged their ABS breaks could indicate unsafe road conditions.

The fact that GSI devices continually broadcast GSI data also separates GSINs from many traditional location-based services, where a user broadcasts their information as part of an intentional query. The fact that GSI devices are always broadcasting information about themselves and their surroundings allows for crowdsourcing of data gathering and empowers remarkable new services, but also raises significant concerns about privacy. GSI analysis should consider how users will be able to control and restrict access to their devices’ data streams, and whether any groups (such as law enforcement) will be able to override the barriers that they set.

GSINs will be a critical element of the location-aware services of the 21st century. As these services expand throughout government, private industry, and social interactions, GSI will expand with them. One immediate question that should be considered is whether services should develop individual, proprietary GSINs, or establish shared platforms. We already have one such platform in mobile phones, which has made the first round of GSI services possible. Flexible technological platforms will allow new services to build on existing infrastructure, greatly increasing economic efficiency. At the same time,
they raise the stakes for privacy concerns by layering many services across a single framework.

Technology that is being developed today will power services for years to come, and policy positions (or their lack) will rapidly transform from tentative statements to hard-to-change aspects of the GSI industry. GSI’s incredibly rapid development pace, combined with its potential to expand into so many facets of society, and the privacy risks that it poses, make it important that we examine its policy implications while it is still in its early stages.

**Evaluating the Privacy Risks and Innovation Potential of GSI**

There are four main factors that contribute to the innovation potential and privacy risks of GSI services: The continuous, real-time nature of the data collected; the invisibility of data gathering; the interconnectedness of GSI services and devices; and the increasing ubiquity of GSI devices and the richness of the data that they provide. Not all of these elements are unique to GSI, but each raises important questions in the context of GSI.

**Continuous, Real-Time Data Flow**

GSI generates a continuous data-flow that relies upon a particularly sensitive piece of data: GSI devices’ real-time location. Mobile phones already provide some location data to telecommunications companies. Many U.S. academics have, in fact, noted the privacy implications of “refined location information” broadcast by devices that are “regularly carried into places where a person has a reasonable expectation of privacy.” (Dempsey 1997, Dalal 2006) These concerns grow even greater as location data becomes the lynchpin for services that process and distribute a wide range of contextual data.

GSI also relies on very accurate location information – the more accurate the better. An Augmented Reality service, for instance, will be of little value unless it knows the user’s location exactly, as well as the direction they are facing. Transacting in such accurate location data raises a range of privacy concerns. First and foremost, unauthorized access could endanger users’ safety. (Goldacre 2006, Donovan 2009) Harrassers, stalkers, thieves, a snooping government, and even terrorists could put such information to a wide range of unsavory uses. Although user location data is generated in real-time, it need not be accessed in real time to pose a significant privacy threat. Real-time monitoring raises immediate safety concerns, but a user’s location history could still reveal his habits, as well as his social and professional connections. It could even be used to predict his current location.

Managing law enforcement’s access to GSI data will be particularly delicate. Although networks that generate real-time streams of location data, such as GSINs, are extremely useful for legitimate investigations, they can also be used to rapidly develop activity and social network dossiers on virtually any user. They could even become a tool for clever criminals to disguise themselves, faking virtual “locations” to throw off pursuit. So far, U.S. courts have split on whether law enforcement requires a warrant to access real-time location data. (Wisconsin v. Sveum 2009, Murphy v. Commonwealth 2009, U.S. v. Garcia 2007) Some authors have already expanded upon these decisions, considering the challenges of RTLS-surveillance in detail. (Shah 2009, Penney 2007, Smith 2007) Their work provides some guidance in the GSI context, but an appropriate balance, and a framework for managing GSI surveillance has not yet been clearly delineated.

However it is used, GSI-enabled devices’ constant stream of real-time location and context data exposes their users to privacy violations that are both more severe and more frequent than ever before. At the same time, reducing the quantity, quality, or frequency of data gathered could fatally handicap GSI services. Thus, despite the obvious risks, regulations that constrain these factors should be approached with care, as they risk eliminating the very characteristics of GSI that make it so valuable. Instead, regulations should focus on how this information is exchanged and retained, and how much control users have over it.

**Invisibility**

Giving users control over their location and context data is made more difficult because the data gathering often happens invisibly. Many devices that we carry, from mobile phones to credit cards with RFID chips, already broadcast streams of data, even though we may not realize it. (Kang and Cuff 2005) Other devices leverage wireless sensor networks to invisibly gather data about the world around them, and public video cameras trained at traffic intersections and ATMs are becoming more prevalent by the day. As these devices become increasingly connected and increasingly commonplace, users will pay less and less attention to the information that is broadcast and gathered about them. Privacy analysts are most concerned when data is gathered without users’ knowledge or consent, which could easily be the case with GSI-enabled devices.

Although it has troubling implications for privacy, the ease with which data can be gathered and exchanged is also an important element of GSI’s potential. If users must jump through too many hoops to access GSI services, they will be unlikely to adopt the new technology. Rather than limiting data gathering, targeted regulation could simply require clear notice and consent by users. This would be only a minimal burden, and would eliminate the problem of invisible data exchange. This would not eliminate the risk of users feeling compelled to consent because of social pressure, or undervaluing their privacy and consenting...
Interconnected GSI services that rely on a small number of open platforms will be able to interact more efficiently, and may be better at providing relevant information to their users. Interconnected networks also increase the possibility of privacy violations – in an interconnected network, even a single service with a poor privacy policy could expose the data of many users. Further, indirect services connected to users only through third-party platforms may experience less user pressure to properly secure user data. Thus, interconnectedness may pose significant risks for user privacy.

Although eliminating data exchange between services would provide strong user protection, however, it would also greatly inhibit the deployment of GSI services. Rather than directly limiting GSI interconnectedness, steps should be taken to ensure more user control and to limit data retention, while still allowing the economic benefits of an open GSI platform.

Ubiquity of Devices and Richness of Data
The final factor – ubiquity and richness – acts as a multiplier for the factors identified above. Many cellular phones already have GPS, and an increasing number of GSI-ready platforms, from the iPhone to Google Maps, are widely available. Other location tracking services – most prominently RFID – are being built into numerous consumer devices. In addition, as GSI becomes more common, the richness of each device – the range of contextual data that the device provides – will likely expand.

Increased ubiquity and richness will increase the quality of GSI services, but also increase the privacy risks that they pose. (Kang and Cuff 2005) Ubiquity will train users to hardly notice their devices’ data gathering. Ubiquity will expand the network of GSI devices, and make it ever easier for user data to travel between services. Increased ubiquity and richness will increase the quantity and quality of GSI data. New devices and new data streams will interact in unexpected ways, generating new services, and posing new risks to privacy. In Japan, researchers are already using real-time electricity usage to monitor the health of elderly citizens who live alone. (Yukio Nakano 2006) More frightening, analysts could aggregate users’ daily routines, matching them to key locations, such as churches, mosques, or the meeting places of substance-abuse support groups, to reveal groups for whom anonymity might be a shield to persecution. These services, and many others, will be made possible by the wide availability of context-rich GSINs.

For all the risks that they pose, however, neither ubiquity nor richness lend themselves to direct regulatory solutions, simply because their growth is also fundamental to the success of GSI. Instead, regulators should address the other factors, while recognizing that increasing ubiquity and richness will render each set of risk factors more potent.

Regulating GSI
Although GSI poses serious risks to privacy that could become institutionalized if not regulated, government intrusion upon the burgeoning GSI market could constrain innovation and hobble the development of GSI services. Thus, regulation should be approached cautiously and with an eye toward flexibility. The final section of this Paper outlines three main regulatory schemes. First, regulation could mandate that GSI users be anonymized through technological and cryptographic techniques. Second, regulation could limit or forbid GSI services from storing user data. Third, regulation could empower users to police GSI services themselves. Each of these approaches has advantages and disadvantages, and none of them should be viewed as a complete solution on their own. By combining elements from all three areas, however, it should be possible to develop regulations that will be both effective and tailored to the emerging nature of GSI.
Anonymizing GSI Users

One solution would be to enable GSI devices to operate anonymously. This might seem counter-intuitive – how can a GSI service operate without broadcasting its location? There are cryptographic techniques, however, that would allow two users to connect through a third-party GSI service without the service (or any other in-range users) knowing either of their locations. (Zhong, Goldberg and Hengartner 2007) Similarly, encrypted, trusted computing systems would allow users to query a service for location-based information without revealing their location. (Hengartner 2007) Finally, devices could generate single-use identifiers for each information request, appearing to services as a string of disconnected devices, rather than a single, traceable object. (Blumberg and Eckersley 2009)

In certain situations, these techniques could provide a web of technological anonymity for GSI users. Services that use location data only in queries, such as navigation services, could be prevented from learning their users’ location. At the same time, services that do rely on location – such as traffic management systems and social networks – could have their users anonymized through single-use identifiers. If effective, this solution would provide strong privacy protection.

Unfortunately, technological anonymization faces several challenges. First, truly impenetrable user anonymity could raise concerns from law enforcement, which would be unable to make use of user data for legitimate investigations. Second, implementing the complex, multi-layered computing processes necessary for user-anonymization would place a heavy burden on GSI services. For more complex services, especially those that seek to aggregate broad sets of user GSI, ensuring users’ identity and location anonymity could rapidly become unwieldy, or even impossible. For fixed-location devices, or mobile devices used along a routine route, anonymizing users would not be sufficient – their location alone could effectively reveal their true identity.

This situation raises one of the fundamental problems of relying on anonymization – defining when users are considered sufficiently “anonymous.” De-anonymization techniques – methods to re-identify users based on supposedly anonymous data – are increasing in sophistication every year, and even data that seems anonymous may in fact hold clues to user identities. (Narayanan and Shmatikov 2008) True technological anonymization is almost certainly impossible for some GSI services, and for many others, it may be an error-prone question of definition.

Finally, in some cases, users may choose to waive their anonymity, as GSI services may be able to significantly improve their users’ experiences if they can track their history and preferences. Unless a regulation were to mandatorily impose user-anonymity for all GSI services (politically unlikely, technologically difficult, and innovation-constraining), at least some users will be left outside technological anonymity’s protection.

Thus, although user anonymization may be technologically achievable and desirable in some cases, it cannot provide robust privacy protection for all GSI users. Where technological anonymity is insufficient, we must turn to regulatory methods of privacy protection.

Limiting Data Retention

Instead of trying to anonymize data, regulators could limit the length of time that services may retain GSI. If properly enforced, this would reduce user risk by ensuring that only a small segment of data would be susceptible to illicit uses. A carefully tailored regulation might even prevent GSI services themselves from building user profiles – another significant privacy concern.

Data retention poses problems, however. First, like technological anonymity, limits on retention could limit user customization and the range of GSI services. Second, it would also place a significant compliance burden on GSI services, compelling them to develop systems to monitor and discard data appropriately. Third, enforcement would be extremely difficult, as identifying illicit data retention, much less finding evidence of that retention, would require significant law enforcement resources. (Blumberg and Eckersley 2009) Finally, given the uncertainty of how the GSI industry might develop or what innovations might drive new services, it would be difficult to craft a data retention regulation that would fit well with future GSI innovations.

It may be possible to address several of the most significant concerns raised by data retention regulation. First, by establishing stiff penalties for violation of data retention limits, regulation could create incentives for compliance even where enforcement would be difficult. Second, a broad rule that sought to establish merely an outer limit for retention would provide a baseline of protection for user privacy without overly constraining future innovation. A broad rule coupled with limited enforcement would also alleviate some of the burden on GSI services – egregious violators could expect scrutiny and the risk of heavy penalties, while small, new services operating in good faith could expect more leniency.

Ensuring Service Transparency and User Control

Rather than regulating technological anonymity or data retention, a third approach would be to impose transparency and access requirements that would allow users to monitor their data themselves. By giving users a “right to delete” or “right to access” their information, they could be given the power to watch the very services that would be watching them. (Ohm 1995)

Regulation could require users to have (a) access to data stored about them; (b) access to audit records of transactions involving their data (such as sale or trade to
services themselves – this addition, user monitoring is than any law-enforcement e following: First, regulation te with fully-anonymized users y the risks and opportunities of sufficient strength to Donald 2008) Finally, it is did not protect user data, Aug. 25, 2009, other words, services that recent study concluded that urgent problems could, in to balance the interests of, 24 IEEE s must focus on supporting monitize their anonymized data. (MacD reading privacy policies could cost users $781 billion heavy cost on users. One reading privacy policies could cost users $781 billion dollars in time and effort to protect themselves – not an considerable sum. (MacDonald 2008) Finally, it is important to note that enabling user access to data creates new privacy risks, as it will create a new loophole for identity thieves to attack. User monitoring offers significantly weaker protection than direct government regulation, but does provide much greater flexibility for the development of GSI. In fact, this flexibility could be good for long-term GSI regulation. Allowing users to publicize specific privacy-threatening activities of GSI services may well shed light on more systemic privacy problems that emerge from the GSI industry as a whole. These emergent problems could, in turn, be addressed by more specific and constraining regulations that might be difficult to implement within an already-rigid regulatory structure.

Conclusion

None of the three methods of user privacy protection considered here provides a complete solution on its own. A combination of all three, however, might be sufficient. One possible solution would be the following: First, regulation would establish minimum data retention constraints and pair them with high penalties for noncompliance. Second, robust user access requirements would allow user monitoring and self-enforcement of privacy standards and data retention regulations. Finally, services that chose to adopt cryptographic techniques of sufficient strength to ensure user anonymity could be offered a safe harbor from other regulatory requirements. In other words, services that truly anonymized their data would be exempt from data retention and user access limitations. This regulation would need to be carefully drawn, both to balance the interests of law enforcement against that of users, and to ensure that anonymization schemes were strong enough to truly protect users from both external risk and GSI services themselves.

This framework would lay some compliance burden both on users and services and create a floor of protection through the data retention limits. In addition, it would provide space for GSI to develop, and empower users to identify privacy risks that require additional regulation. If properly balanced, the safe harbor would create incentives for services that could operate with fully-anonymized users to do so because of reduced compliance costs and the opportunity to retain and monetize their anonymized data. This two-stepped framework would implement user-anonymity where possible and ensure regulatory protection where it would be impossible.

More detailed analysis of a regulatory scheme for GSI is beyond the scope of this paper. Whatever the ultimate solution, however, regulators must focus on supporting GSI’s potential while protecting users who may eventually find themselves compelled by social and professional demands to adopt GSI-ready devices. Although the implications of GSI have had relatively little scrutiny, the technology already exists, and entrepreneurs around the world are rapidly developing new GSI applications. We must act quickly to identify the risks and opportunities within the new technology, and ensure that both users and innovation are protected as it continues to develop.

References


Urs Hengartner, Hiding Location Information from Location-Based Services, Proc. of International Workshop on Privacy-Aware Location-based Mobile Services (PALMS), Mannheim, Germany, May 2007, pp. 268-272.

Robert G. Hollands, Will The Real Smart City Please Stand Up?, City, 303-320 (Dec. 2008).


John Snow, M.D., ON THE MODE OF COMMUNICATION OF CHOLERA, John Churchill (1855).

US v. Garcia, 474 F.3d 994 (7th Cir. 2007), Reh’g and suggestion for reh’g en banc denied, no. 06-2741, 2007 US App. LEXIS 8397 (7th Cir. Mar. 29, 2007) (holding that GPS tracking does not constitute a search and does not require a warrant).

