OPENPATHS: EMPOWERING PERSONAL GEOGRAPHIC DATA

Brian House, Department of Music, Brown University, Providence, RI, USA. E-mail: <brian_house@brown.edu>. *

Abstract

OpenPaths, created by the New York Times Company R&D Lab, is a platform that demonstrates the collective value of personal data sovereignty. It was developed in response to public outrage regarding the location record generated by Apple iOS devices. OpenPaths participants store their encrypted geographic data online while maintaining ownership and programmatic control. Projects of many kinds, from mobility research to expressive artwork, petition individuals for access to their data. In the context of locative media practice, OpenPaths expands the notion of the tracing to address the components of an ethical implementation of crowd-sourced geographic systems in the age of "big data".

Keywords

locative media, big data, data visualization, privacy

The "tracing" as a mode of locative media art practice was established through projects such as *Amsterdam Realtime* [1]. Conducted in 2002, participants were given Global Positioning System (GPS) devices to carry as they traversed the city in the course of their everyday lives. The result was a compelling display of the collective routes, a tracing of the urban topology through which, moment to moment, the city was given form. At the time, such work was speculative, anticipating mobile phone networks. Yet *Amsterdam Realtime* already hints at the eeriness of a city inhabited solely by disembodied, moving coordinates, and the centralization required to pull off the project demonstrates the involvement of commercial or military infrastructure in tracing and its potential use for surveillance. Such concerns fuel a broad critique of much subsequent locative media and data visualization practices that do not address the political implications of their technological underpinnings [2]. Recently, the disclosure of the PRISM initiative of the United States' National Security Agency provides a dramatic confirmation of the danger of centralized data gathering and the collusion of state and corporate interests in tracking individuals [3].

Fig. 1. ***Amsterdam Realtime***
© 2002 Esther Polak

Nevertheless, there is great potential for the tracing to serve the public interest. In 2006, Mark Hansen and colleagues at the Center for Embedded Networked Sensing at UCLA introduced the term "participatory sensing" [4]. Recognizing the ubiquity of mobile phone users and the devices' capacity to gather data, they proposed that individuals might opt-in to ad hoc sensor networks to address issues in "urban planning, public health, cultural identity and creative expression, and natural resource management". They note that "we know something about what distributed sensing can be used for in the sciences, industry and the military. We know much less about its function and utility in the public sphere when the components are owned and operated by everyday users". Personal geographic data in this context might be used to observe mobility patterns and identify opportunities for improving public transport or to allocate social services. Such research is appealing as, due to the penetration of device ownership, the potential reach is vastly larger than what would be possible with traditional methods.

Yet this potential remains untapped, even as vast datasets are gathered for commercial purposes. iPhone and Android users, which as of June 2013 make up 56% of the adult US population [5], have at least two corporations tracking and storing where they are at all times. This is the network operator, such as AT&T, who by definition knows your location in the course of delivering cellular service, and the software provider, such as Apple, who actively monitors your location to enhance their applications. The result, for these companies, is so-called "big data", a buzz word signifying both databases of a magnitude that requires specialized computational techniques as well as an epistemological approach that places an absolute value on emergent patterns [6]. However, despite, or because of, its value, these corporations do not have interfaces or policies in place that would allow the release of these datasets to the individuals who generated them, let alone to user-endorsed third-party research programs. As Natasha Singer of the *New York Times* reports,

> ...when I called my wireless providers, Verizon and T-Mobile, last week in search of data on my comings and goings, call-center agents told me that their companies didn't share customers' own location logs with them without a subpoena [7].

Location data are commonly generated in three ways. Network operators find the position of a device by the triangulation of its signal strength to nearby cell towers. Additionally, most contemporary smartphones are equipped with a GPS sensor, by which it may locate itself in latitude and longitude via signals from geosynchronous satellites [8]. Finally, a device may note the identifiers of nearby cell towers and Wi-Fi nodes and infer its position from a database that lists the coordinates of these signals. Apple's iPhone uses this latter method together with GPS in what is known as "hybrid positioning" [9]. Originally, Apple leased their database from Skyhook Wireless, but in 2010 implemented a system to generate their own [10]. Essentially, Apple employs the iPhone-carrying public as a giant "wardriving" [11] sensor network – the location of novel Wi-Fi nodes and cell towers detected by iPhones are logged and sent back to Apple to contribute to an extensive map of the topology of wireless signals across the world [12].

In April of 2011, researchers Pete Warden and Alasdair Allan publicized a fact already known in digital forensics circles. Beginning in April of 2010, the data collected by individual iPhones and iPads for Apple's database were stored in a cache file automatically synced to the users' computers via iTunes. By default, this file was not encrypted, and it could be readily examined by anyone with access to the computer [13]. Though Apple stated that "The iPhone is not logging your location. Rather, it's maintaining a database of Wi-Fi hotspots and cell towers around your current location" [14], in practice the distinction was somewhat semantic, as the file clearly reflects location history spanning a year's time. The result was dubbed "Locationgate", a scandal which indicated that users were uncomfortable at how such data were being collected. Senator Al

* Brian House was Creative Technologist at the Research and Development Lab at the New York Times Company, 2010-2012. The views expressed here are the author's and do not necessarily reflect those of the New York Times Company.

Please reference as: Brian House (2013) "OpenPaths: Empowering Personal Geographic Data" in Kathy Cleland, Laura Fisher, Ross Harley (Eds.) *Proceedings of the 19th International Symposium of Electronic Art*, ISEA2013, Sydney. <http://ses.library.usyd.edu.au/handle/2123/9475>
Page numbering begins at 1 at the start of the paper.

Franken demanded that Apple explain themselves [15], 27,000 Koreans sued the company for violation of privacy [16], and even *South Park* weighed in [17]. As Kord Davis puts it, "The decision to use that technological method had clear and direct ethical consequences in the real world" [18].

However, there is a certain irony in the outrage, as consumers were agitating for Apple to restrict access to what was in essence the largest publicly accessible Cartesian document in human history – a year's worth of data for over 50 million iPhone users. Locationgate came to an end on May 4th, 2011, when Apple released iOS version 4.3.3, which no longer logged location data to a cache file. But while users can no longer access these data, Apple certainly continues to collect them. Further, Apple shares individual portions of those data with applications – a large percentage of apps for both iOS and Android request access to a user's location via a confirmation box with the options "Don't Allow" and "OK" that lacks subtlety. An approved app may collect continuous personal geographic data. Yet this infrastructure lacks the means for the user to know exactly what data have been collected or how they will be used, and unless an application developer has built an interface to do so, there is no way for users to access their own data for their own purposes. So while Locationgate helped raise public awareness about the nature of personal data, in the end the discussion fell short of asking what rights individuals should have over their location histories, what might be done with the data as a public resource, and what a more ethical implementation for collecting data might be.

In response to the discussion around iOS cache files, in May of 2011 the New York Times Company Research and Development Lab launched OpenPaths <https://openpaths.cc>. Initially, the platform consisted of two components. First, we wanted to create a tool that would allow non-technical users to locate SQLite location databases within their iTunes backup directories. Our tool, built in Python for both OS X and Windows machines, searched the archives of all devices that had been synced with the computer in question, as well as any connected backup disks. Once presented to the user, the files could then be uploaded to the OpenPaths server. Since Apple's "fix" was already released, this

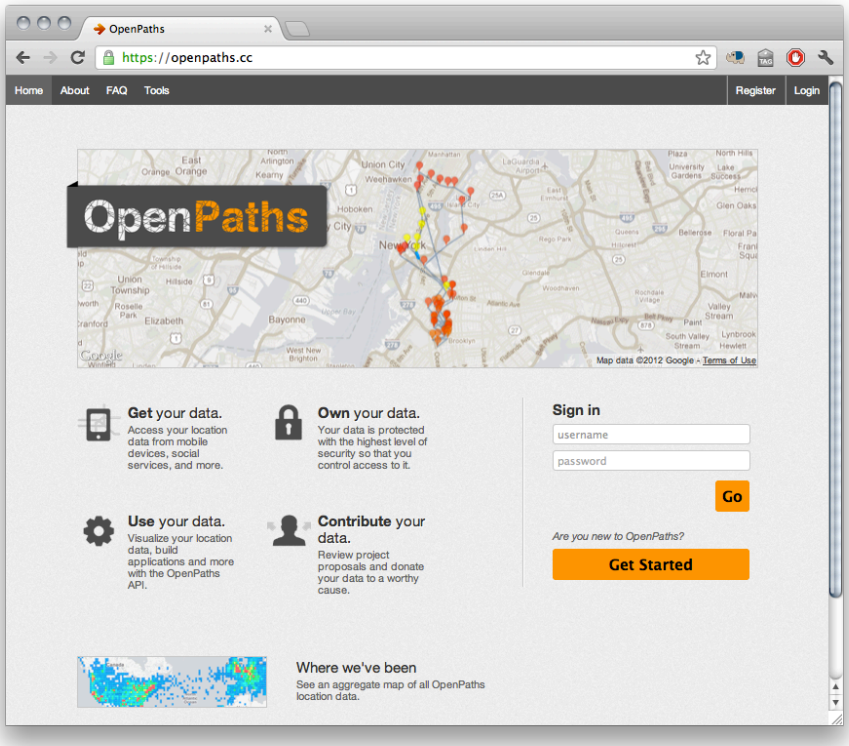


Fig. 2. OpenPaths homepage, © 2012 The New York Times Company

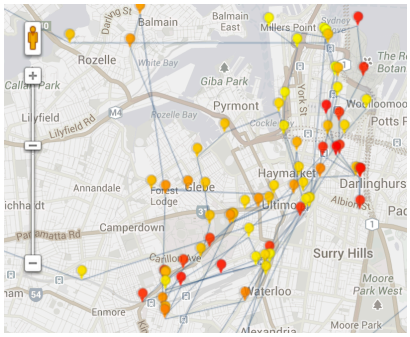
effort was designed to salvage as much historical data as possible before they were deleted or overwritten by updates, and approximately 4000 datasets were collected in this way.

We designed the OpenPaths server as a data "locker" of sorts, one that would embody the idea of "personal data sovereignty". Generally, "data sovereignty" is a business term that acknowledges, marketing language about "clouds" aside, that data lives on physical machines and are hence subject to the local laws in which the data centers are located [19]. This is a liability for corporations if valuable assets are stored by a third-party hosting service, such as Amazon S3, that may be subpoenaed by a state power, such as the US government under PRISM. Personal data sovereignty extends this concept to the level of the individual – it is an alternate model of data collection that empowers that individual with control over the data they generate that is not site-based, but access-based. There are technological and legal aspects of the implementation. From a technological perspective, we propose that data are under your control either when they are stored on a machine to which you physically restrict access or when they are encrypted with a key that only you have. The basic concept of OpenPaths is that by encrypting your data but not storing the key (which is generated from the user password), the service maintains a remote infrastructure without reserving any privileged access to the data themselves (nor is access ceded to the hosting provider, in this case Amazon). This straightforward technological feature is simply a literal interpretation of our user agreements, which state that you own your data, and that your data cannot be accessed without your express permission and participation. The shift that we hope to exemplify is that by leaving out the ability to mine or sell data, the user is no longer an asset in that regard – collective value for OpenPaths users is produced by mutual participation, as we explain below.

Public interest in the project motivated us to provide a means for individuals to continue to collect their data on an ongoing basis without the cache files. Our solution was apps for iOS and Android designed for the single purpose of collecting location data and uploading them to the OpenPaths server with as little friction as possible. The primary technical challenge was to ensure that the apps could run continuously in the background without causing undue battery drain. This largely precludes the possibility of using GPS sensing, which is power-intensive – we use the iOS and Android location services that provide

updates when “significant location change” [20] events occur based on cell-tower and Wi-Fi-node triangulation. The resulting data are similar in resolution to Apple's original location caches, with a topography that suggests a trail of breadcrumbs rather than an uninterrupted GPS path. Likewise, it is of higher quality in dense urban areas with well-documented WiFi – noise is frequently present in suburban locales. Regardless, the apps are effective in tracing individual movements, with a total of ~10000 active users as of this writing.

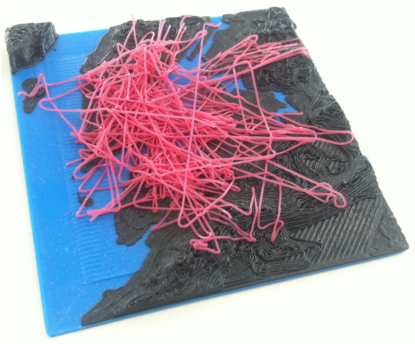
Fig. 3. The author’s path at ISEA2013
Map data © 2013 Google



Participants access their data through the OpenPaths website. Once you are logged in, the platform is able to decrypt your location history and provide access in a variety of ways. CSV, JSON, and KML formats can be directly downloaded, and an OAuth API [21] allows the system to be polled for updates. Our intent is to provide a minimum viable feature set – however, we do include a basic tool to explore your data on a map.

With this interface, you can watch an animation of your travel unfold. Viewing one's geographic tracing is undeniably a compelling framework for personal narrative. When you look at a map of your activity, you see stories, and cannot help but populate the representation with your personal experience. There is, in other words, a meaning in the data beyond the encoding, and OpenPaths has been used by individual artists applying a variety of tools to produce a wide range of interpretive pieces. These include a Processing sketch by Bert Balcaen that recreates a month in New York as a dance of particle systems [22]; a 3D representation of Chris Woebken's path that he printed with a MakerBot (Fig. 4); and a laser-cut necklace showing a network of significant points by Michael Massie commemorating a trip to Zurich [23]; my own work, *Quotidian Record*, which maps 365 days of location data to 365 rotations worth of music on a vinyl record [24]; a tool by the team at CartoDB that estimates total carbon consumption by mode of transport [25]; and Wes Grubbs’s workshop code (for Eyeo Festival, 2013) for finding the distance between two people over the course of their travels, an exercise which proved most compelling when applied to the data of two supposed strangers.

Fig. 4. 3D print of Chris Woebken’s path
Photo © 2012 Brian House



After Sue Huang's phone was stolen in July of 2011, it continued to report its location to OpenPaths. With her assistance, we interpolated positions between each point and pulled the corresponding Google Street View tiles, creating a video showing a point of view as if Google was driving the getaway car, which we called *Joyride* [26]. This project points at the fiction of representation woven by our media platforms with data, and the estrangement possible when personal data are separated from the person. In fact, part of the pedagogical purpose of OpenPaths is to ask what inhabits the tension of that abstraction.

To that end, we have conducted OpenPaths workshops at Rhode Island School of Design, Eyebeam Art and Technology Center, and the School of Visual Arts in New York, following a model initially proposed by design educator Daniel Goddemeyer [27]. Participants use OpenPaths for a week to generate datasets and then anonymously trade with someone else in the group. Each participant develops a presentation on what can be inferred from the data based on cross-examining them with other information together with personal knowledge and intuition. Finally, this report is compared with testimony from the actual subject. We have found, unsurprisingly, that a tremendous amount can be learned about an individual through this process, even without advanced computational tools. The workshops are intended both to increase literacy as to the potential of location data (and the subsequent privacy implications) as well as to further demonstrate that the data are not inert and are subject to narrative and interpretation.

We feel that this is an important exercise in the era of big data. Kord Davis writes,

> Any context we create to turn data into information automatically assigns new characteristics to it, causing data *itself* to become less anonymous and more meaningful. And if we have enough data, we can correlate, extrapolate, query, or extract some very useful new information by understanding the relationships between those characteristics ... while the value of that utility is growing exponentially in our time, so too is the unknown potential for unintended consequences... [28]

Hence a fundamental respect for the individual is necessary when aggregating personal data, as the resulting computational models are tethered to pieces of the real world that carry personal weight. From the standpoint of both pedagogy and practice, we need to cultivate empathy for the people involved in systems [29].

“Participatory” implies individuals who are supplying personal data from a personal device to a study or project because they have an investment or interest in the result. The population of OpenPaths users is constantly collecting data a priori of any particular study, which creates the possibility of assembling ad hoc datasets for larger investigations. OpenPaths includes the infrastructure for “projects” conducted by third parties. Project proposals are not curated by the platform admins, but are sent directly to individual OpenPaths users who then decide whether or not to contribute their personal data. Proposals must include information on how the data will be used, how they will be kept secure, and how the project will benefit the OpenPaths community or the public at large. On average, this opt-in model has produced response rates typically around 600 participants (6%) per project. This is small from a commercial standpoint, but significant for epidemiological or artistic initiatives.

Maintaining the encryption model of OpenPaths while allowing third-party access requires what we think is an innovative security system. We employ an

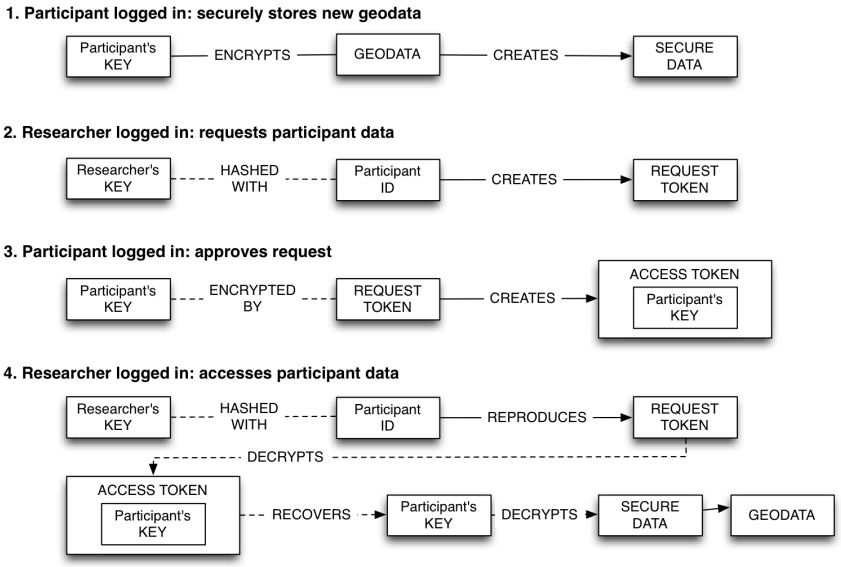
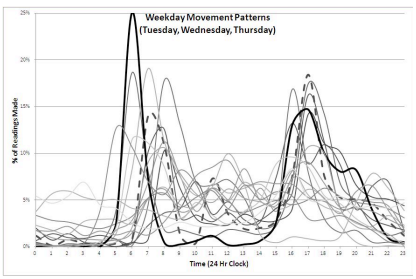


Fig. 5. OpenPaths security model, © 2012 The New York Times Company

exchange of revokable tokens, a simplification of which is as follows. First, project owners (who must be registered OpenPaths users) request participation. This produces a request token for each user that is a hash of the researcher's key and the participant identifier. If the participant approves the request, their key, which is not otherwise stored in the system, is encrypted with the request token to produce an access token. Meanwhile, the request token is eliminated. When the researcher logs in to retrieve the data, the request token is re-created and used to unlock the access token, recover the participant's key, and decrypt the data. This happens in parallel across all participants. The platform facilitates the exchange of data but does not store them unencrypted and so does not maintain for itself any privileged access.

The result has been myriad projects in mobility research, art, urban planning, self-tracking, data visualization, and entrepreneurialism. Highlights include a "re-mapping" of China via longboard [30], a comparison of human mobility

Fig. 6. OpenPaths commuting patterns, <http://researchthecity.com/> © 2013 Niamh Rabbit, Trinity College Dublin

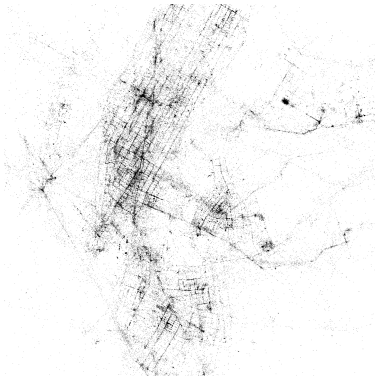


patterns with the spread of the tiger mosquito [31], and the "Science of Getting Lost" [32]. Critically, OpenPaths does not curate or otherwise exclude project proposals, other than to verify their completeness and legibility. Further, the platform supplies participants' unfiltered data to projects – there is no provision to attempt degrees of anonymization, as such a process is likely to fail [33]. This puts the onus on the participants to make informed choices about how their data should be used.

A collective tracing of New York City, produced daily via participants in the New York Times Company Research and Development Lab's own "Mapping Habitual Geographies", has much in common with *Amsterdam Realtime*. It is a portrait of a city defined by its transitory dynamics. Yet where the earlier work operates aesthetically and carries with it a certain foreboding, OpenPaths projects are characterized by a situated politics. First, the data mirrors that which have already been collected by the network operator and software provider, and subsequently by unknown corporate or government entities. As such, they already hold presumed utility from a commercial or surveillance perspective, but that utility is restated, in an act of détournement, in terms of scientific or artistic value as the participants see fit. Secondly, the dataset held by AT&T, for example, comprises an unwitting collective formed solely by consumer habits and/or the practical necessity of using a cellphone. The voluntary and informed formation of a group of participants in an OpenPaths project has a markedly different nature, and, in contrast, is an intentionally political body.

Mark Tuters and others have identified a post-locative practice that shifts emphasis away from the tracing of individuals to the networks of interactions between objects [34]. Projects like MIT's Trash Track initiative [35] or Christien Meindertsma's *Pig 05049* (2008) [36] exemplify the proposition of theorists such as Latour to consider perspectives beyond the human subject [37]. Yet the post-locative should not ignore the human trace, given its ineluctability, and should seek to interrogate the nature of its data. In other words, the communication protocols, encoding schemes, and user interfaces by which location information is formed are not given – Google Maps, for example, may be the de facto standard on Android, but it is a system with designed biases and can be contested as such.

Fig. 7. "Mapping Habitual Geographies" © 2012 The New York Times Company



OpenPaths seeks to inhabit this inflection point where the collection of location data creates a context in which to assess personal data in general, even while acknowledging the particularly vital connection of the geographic tracing with the personal narrative about how it came to be. We suggest that the erasure of context that comes with the encoding of data can be restored through an actual, functional relationship with the individual via a respectful infrastructure. Our hope is that what we have proposed with OpenPaths will serve as one model for how the ethical exchange of data is both possible and necessary.

References and Notes

1. Esther Polak, *Amsterdam Realtime* (2002), media artwork, <http://realtime.waag.org/>, accessed June 28, 2013.

2. Mark Tuters and Kazys Varnelis, "Beyond Locative Media: Giving Shape to the Internet of Things", *Leonardo* **39**, No. 4 (July 2006) pp. 357-363.

3. Barton Gellman and Laura Poitras, "U.S., British intelligence mining data from nine U.S. Internet companies in broad secret program", *Washington Post* (June 6, 2013).

4. Jeff Burke, Deborah Estrin, Mark Hansen, Andrew Parker, Nithya Ramanathan, Sasank Reddy, Mani Srivastava, "Participatory Sensing", *Proceedings of the International Workshop on World-Sensor-Web* (Boulder, U.S.A.: ACM, October 31, 2006).

5. Aaron Smith, "Smartphone Ownership 2013", *Pew Internet* (June 5, 2013), <http://pewinternet.org/Reports/2013/Smartphone-Ownership-2013.aspx>, accessed June 28, 2013

6. Steve Lohr, "Sizing Up Big Data, Broadening Beyond the Internet", *New York Times* (June 19, 2013).

7. Natasha Singer, "If My Data Is an Open Book, Why Can't I Read It?", *New York Times* (May 25, 2013).

8. "Global Positioning System", *Wikipedia: The Free Encyclopedia* (Wikimedia Foundation Inc., updated July 3, 2013, 19:09 UTC), <http://en.wikipedia.org/wiki/GPS>, accessed June 28, 2013.

9. "Hybrid positioning system", *Wikipedia: The Free Encyclopedia* (Wikimedia Foundation Inc., updated June 27, 2013, 13:36 UTC), <http://en.wikipedia.org/wiki/Hybrid_positioning_system>, accessed June 28, 2013.

10. Brian X. Chen, "Why and How Apple Is Collecting Your iPhone Location Data", *WIRED* (April 21, 2011), <http://www.wired.com/gadgetlab/2011/04/apple-iphone-tracking/>, accessed June 28, 2013.

11. "Wardriving", *Wikipedia: The Free Encyclopedia* (Wikimedia Foundation Inc., updated June 27, 2013, 18:31 UTC), <http://en.wikipedia.org/wiki/War_driving>, accessed June 28, 2013.

12. Chen [10].

13. Alasdair Allan and Pete Warden, "Got an iPhone or 3G iPad? Apple is recording your moves", *O'Reilly Radar* (April 20, 2011), <http://radar.oreilly.com/2011/04/apple-location-tracking.html>, accessed June 28, 2013.

14. "Apple Q&A on Location Data", Apple Inc., press release (April 27, 2011).

15. Al Franken, letter to Steve Jobs (April 20, 2011), <http://www.franken.senate.gov/files/letter/110420_Apple_Letter.pdf>, accessed June 28, 2013.

16. Jun Yang, "IPhone Users in South Korea Sue Apple for Collecting Data Without Consent", *Bloomberg.com,* (August 17, 2011), <http://www.bloomberg.com/news/2011-08-17/apple-s-iphone-users-in-south-korea-claim-data-collection-breached-privacy.html>, accessed June 28, 2013.

17. "HUMANCENTiPAD", *South Park* (April 27, 2011), television.

18. Kord Davis, *Ethics of Big Data*, (Sebastopol, U.S.A.: O'Reilly Media, September 28, 2012), p. 1.

19. "Data sovereignty", *WhatIs.com*, (updated March 2013) <http://whatis.techtarget.com/definition/data-sovereignty>, accessed June 28, 2013.

20. See APIs for iOS <http://developer.apple.com/library/ios/#documentation/userexperience/conceptual/LocationAwarenessPG/CoreLocation/CoreLocation.html> and Android <https://developer.android.com/google/play-services/location.html>.

21. See <https://openpaths.cc/api>.

22. Bert Balcean, "35 days in NYC" (September 2012), blog post <http://www.theworldneedsmoredreamers.net/35-days-in-nyc/>, accessed June 28, 2013.

23. Michael Massie, "Geo2Jewelry" (November 14, 2012), blog post <http://www.michaelmassie.com/blog/our-mapping-data-from-openpaths-during-our/>, accessed June 28, 2013.

24. Brian House, *Quotidian Record* (2012), media artwork, <http://brianhouse.net/works/quotidian_record>, accessed June 28, 2013.

25. See <http://geostats.herokuapp.com/>.

26. Brian House, *Joyride* (2011), media artwork, <http://brianhouse.net/works/joyride>, accessed June 28, 2013.

27. Daniel Goddemeyer, Amit Pitaru, Noa Younse, "Data Narratives", workshop <http://www.danielgoddemeyer.com/teaching.php>, accessed June 28, 2013.

28. Davis [18], p. 35.

29. Jer Thorp, "Make Data More Human", *TED* (November 2011), online lecture, <http://www.ted.com/talks/jer_thorp_make_data_more_human.html>, accessed June 28, 2013.

30. See <https://openpaths.cc/projects/GEYDAMBQGEZQ>.

31. See <https://openpaths.cc/projects/GEYDAMBQGA4Q>.

32. See <https://openpaths.cc/projects/GEYDAMBQGIYA>.

33. Nate Anderson, "'Anonymized' data really isn't—and here's why not", *Ars Technica* (September 8, 2009) <http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/>, accessed June 28, 2013.

34. Mark Tuters, "Forget Psychogeography: The Object-Turn in Locative Media", paper presented at the "Unstable platforms: the promise and peril of transition" conference, MIT (May 2011), <http://web.mit.edu/comm-forum/mit7/papers/Tuters_DMI_MIT7.pdf>, accessed June 28, 2013.

35. See <http://senseable.mit.edu/trashtrack/>, accessed June 28, 2013.

36. Christien Meindertsma, *Pig 05049* (2008), print artwork, <http://www.christienmeindertsma.com/index.php?/books/pig-05049/>, accessed June 28, 2013.

37. Bruno Latour, *Reassembling The Social: An Introduction to Actor Network Theory* (Oxford: Oxford University Press, 2005).