SHARED SIGHT:
SIRENS AND CYCLOPS

by

DANA CUFF*
MARK HANSEN**
JERRY KANG***

Note to readers: This draft was prepared for a slightly different audience, so the arguments focusing on visual privacy do not appear in the text.

Version 1.0 :: 10/14/2006

* Professor of Architecture and Urban Planning, UCLA, dcuff@aud.ucla.edu

** Associate Professor of Statistics, UCLA, cocteau@stat.ucla.edu <http://www.stat.ucla.edu/~cocteau>

*** Professor of Law, UCLA School of Law. <kang@law.ucla.edu> <http://jerrykang. net>.

Research assistance provided by: Nathaniel Ross.

Workshopped at: Preliminary versions and drafts of this paper were workshopped at Google, UCLA School of Law, Center for Embedded Network Sensing, Computer Science Technology Board of the National Academy of Sciences.

Thanks to: Deborah Estrin.

Supported in part by: UCLA School of Law; UCLA Academic Senate.

TABLE OF CONTENTS

INTRODUCTION

[1] Within academic research, pervasive computing[1] in the form of embedded networked sensing has made the leap from the laboratory to the environment. Simultaneously, outside academia in the domain of personal communication and corporate marketing, pervasive computing has entered every backpack, purse, and coat pocket in the incipient form of mobile phones. We characterize this contextual shift as "urban sensing." To consider this shift merely scalar or a change in degree is an error. Rather urban sensing represents a change in kind, a qualitative shift into the realms of politics, aesthetics, interpretation, and motivation. A careful, transdisciplinary examination of these new dimensions of pervasive computing provides insight into the directions future research should take.

I. THE FOREST: FROM THE LABORATORY TO NATURE

[2] In Lake Fulmor, a small body of water in the San Jacinto mountains, seven moored buoys dangle strings of thermistors to acquire time series of temperature at several different depths. The lake is long and thin, roughly five times longer than it is wide, and the buoys are arranged in a line running the length of the lake. Suspended from each buoy, at half a meter below the surface, is a submersible fluorometer recording chlorophyll concentrations. A team of biologists and engineers from the University of Southern California oversees the system, and collects sensor data wirelessly from shore; visualization tools help this group examine both the physical and biological dynamics in the lake. For a more complete picture of the local environment, data from the buoys are combined with windspeed and other microclimate measurements from a nearby weather tower at the

[1] In prior work, we defined "pervasive computing" as the form of computing that arises from the convergence of three trends—increased *ubiquity* of broadband, wireless communications; increased *embeddedness* of computing into physical objects, environs, and infrastructure; and increased animation in both sensing and actuating. See [PerC]. This paper highlights the sensing aspects of pervasive computing as it rolls out in urban contexts.

James Reserve, a biological field station that is part of the University of California Natural Reserve System.

[3] The lake ends roughly at an intersection with Highway 243 and extends lengthwise to the northeast. As you move along the line of buoys toward the highway, the lake deepens. There is also a slight, nearly imperceptible flow in this direction. At four time points during the spring, summer and fall of 2006, a robotic sensing device is installed at the deep end of the lake. The activity is sponsored by the Center for Embedded Networked Sensing, an NSF Science and Technology Center. The robotic systems consists of a cable that spans the lake at its widest point, oriented perpendicular to the line of buoys and the flow through the lake. A small shuttle rides along this cable carrying with it a sensor node that is dipped into the lake at regular intervals. The shuttle submerges the node, and its cluster of sensors, allowing it to dwell briefly and take measurements at several depths. The resulting data form a grid, profiling temperature, chlorophyll concentration and about a dozen other variables in the plane of the cable system. We present the regular grid of measurements taken by the robotic node in one pass over the lake transect; both chlorophyll concentrations (top) and water temperature (bottom) are displayed. When paired with the static buoy data, a model can be formed that captures the important chemical, physical and biological processes in the lake.

[4] At the nearby James Reserve, ten wireless sensor nodes have been deployed in a two-by-five grid spanning covering a 2mx75m area. Each node connects to a series of sensors, buried at depths of 2cm, 8cm and 16cm, and recording soil temperature, moisture and CO2 concentration. In addition, below ground at each grid point, a robotic camera rides along a transparent acrylic tube to collect close-up images of roots, fungi, and the surrounding soil structures. In the next few months, the existing traditional CCD cameras will be replaced with low-power, CMOS devices supporting programmable image processing and wireless communications. The robotic camera can be operated manually; provided with a regular sampling plan (along the tube); or can be triggered by environmental conditions like rain events. Above ground a micrometeorological station records air temperature, pressure, humidity and PAR (photosynthetically active radiation, designating the spectral range of solar light from 400 to 700 nm). Researchers at the University of California in Riverside are examining the output from this system to correlate below ground biological activity and surface CO_2 flux. Eventually, an above-ground robotic tether-and-shuttle system (similar

to that at Lake Fulmor) will scan the 75m transect, taking surface temperature measurements with an infrared thermometer and canopy profiles with a laser range-finding system.

[5] The James Reserve also plays host to a number of imaging devices, from standard video cameras, to smaller webcams to even smaller "mote class" imagers (the so-called Cyclops cameras are the result of a joint project between CENS and Agilent). Some of these cameras are mounted out of doors, monitoring the development of flowering plants in a meadow, or the response of mosses to rain . A number of these imagers are mounted in a network of nestboxes. These images are collected once every fifteen minutes and are subjected to a series of images processing algorithms that determine whether or not the box is occupied as well as higher-level ``events" or milestones like nest building, egg laying and hatching. The use of imagers as biological sensors is a new research thread for CENS.

[6] The rollout of these embedded networked sensors have coincided with the proliferation of geocoded/geostamped data and the accompanying GIS platforms for their visualization. Services such as Google Maps and Google Earth have driven to nearly zero the cost measured in terms of dollars, time-to-deploy, and technical sophistication required. So-called "mashups" with Google Maps have provided anyone with a web browser the ability to display data (sounds, images, video, "data data") in map layers. In combination with the embedded networked sensors, such systems have made it possible to visualize data in real space, to construct overlaying maps of information, and to analyze locational phenomenon over time.

[7] This move by CENS from the lab to the forest has been a radical leap forward, pushing the capabilities of sensors and robots, as well as offering rich new understanding of the forest itself. This accomplishment directly furthers CENS' mission, which is to grow technology in the context of specific scientific questions. But the choice of the forest as the first out-of-lab experience was neither accidental nor driven solely by the promising nature of environmental applications. Rather the forest was an ideal site for scientists to colonize and manipulate, without interference. In addition, the forest as testbed sidestepped the thorny political and cultural problems of ubiquitous surveillance that have entered public debate. With James Reserve as today's reality, we can ask: What happens tomorrow, when pervasive computing comes out of the woods and goes urban?

II. THE CITY: SENSING GOES PUBLIC

A. Beyond Scientists

[8] Fully centralized instrumentation of the city in the manner of the James Reserve is simply not possible. The money necessary for similar instrumentation in any major city is simply unavailable. Even with requisite funds, scientists lack the property rights to instrument everywhere, and individuals enjoy privacy rights not granted to sparrows. Thus, in the city, we necessarily move away from what might be called the *full centralization* model of the James Reserve. In this model, there are distributed sensors throughout the space, but these sensors, the data they collect, and the ways in which the data are processed are subject to centralized control by the scientists in charge.[2]

[9] From this extreme, one could imagine moving along the spectrum toward the model of *distributed citizen-sensing*. Although a central authority maintains the basic terms and conditions of data collection and also maintains the centralized data repository, that authority employs local data collectors who voluntarily and idiosyncratically observe and input the data. The central authority limits the kind of data that are collected, and provides some basic guarantees about data quality, perhaps removing outlying contributions – but at the very least ensures consistent data formats.[3] A real world example is the Great Backyard Bird Count, in which a snapshot survey of birds across the continent is accumulated from bird watchers of all levels who submit their counts over a four-day period. Centrally managed by the Cornell Lab of Ornithology and Audobon, in 2005 some 52,000 checklists were sent in from across the country, tallying over 6.5 million birds.

[10] Still further along this spectrum, we have what might be called a *fully decentralized* model. Here, we have no central authority to speak of, except some actor providing basic search and storage. This is consistent with an ethos called "Web 2.0," which encompasses those Web sites that derive their principal value from user participation.[4] These Web 2.0 applications allow anyone to post, combine, attract, search/discover and interpret data. As there is little centralized control over the data (besides the initial software infrastructure that permits data sharing), the

[2] Cross-ref NEON and other projects of this scale.

[3] To be sure, there are variations on this approach. For instance, the Great Backyard Bird Count creates more exacting templates for data collection than a wiki, which allows far more flexibility in what might be uploaded as well as deleted.

[4] (InformationWeek article.) [SWAT: check what if any part was quoted.]

distributed individuals who collect and share the data are in charge. There are many examples, from the collaboratively constructed Wikipedia, to the contribution/promotion model of digg.com, or the community organized tagging of del.icio.us. Mashups, which are applications that combine data from several sources, function like redistribution mechanisms, providing opportunities for new interpretations and new inferences.

[11] In sum, urban sensing will involve a decentralization that shifts focus and control away from the scientist at the center. Instead, distributed networks of individuals will play a far larger role in collecting the data and making sense of it. In other words, data will be collected, then interpreted, in ad-hoc ways by everyday citizens going about their daily affairs. But a threshold objection may be raised: citizen-sensing would require individuals to carry around instruments and equipment—how likely is that? Our answer is that we're already carrying them in our pockets.

[12] We mean, of course, the sensor that is known as the "cell phone," with penetration rates reaching almost 70% in the U.S.[5] As these cell phones continue to morph into multi-functioned communicators that leverage high speed wireless data networks and increasingly usable user interfaces, we will regularly carry extremely powerful computers capable of networked data processing. Although we think of cell phones as communication devices, which we episodically and intentionally use and then turn off, we should recognize that they are also passive sensors that can silently collect, exchange, and otherwise process information all day long. Obviously, they are engineered to sense sound–our voices–but they also can sense images and video through their built-in cameras. Still more interesting, they can sense location through GPS receivers and/or basic cell phone triangulation. This bundling of functionality is not solely market driven; it satisfies the Federal Communication Commission's E911 regulations. In addition to sight, sound, and location, in the near future (inside of 15 years), cheap sensors that detect other aspects of the environment may be available as plug ins. Obvious candidates include sensors for various environmental threats, such as pollution, allergenics, and radiation.

---

[5] [Use FCC data.] (Hong Kong has the highest rate at 127%) Gruener, Wolfgang. TG Daily, "US Cellphone Penetration Close to 70%." April 6, 2006 20:48. http://www.tgdaily.com/2006/04/07/cellphone_subscribers/. Hong Kong data from Office of the Telecommunications Authority, The Government of the Hong Kong Special Administrative Retion. June 2006 data. http://www.ofta.gov.hk/en/datastat/key_stat.html

[13] Another technology playing a pivotal role in urban sensing is the radio frequency identification tag (RFID). At their simplest, RFIDs are small antennas with enough memory to radiate a number when a "reader" scans the tag. Given the trends in decreasing size (now as small as a grain of rice) and cost, it is conceivable that a majority of physical items could be uniquely identified by a number, readable by sensors from a distance. Such a fully ennumerated world, complemented by ubiquitous broadband wireless data networks and RFID-reader equipped mobile communicators, makes distinctly possible "an Internet of things."[6]

B. Beyond Science

[14] As we have demonstrated, coming out of the woods means going beyond the centralized control of *scientists*. Relatedly, we must be braced to go beyond *science* itself.

[15] First, consider the concerns around unprofessional data collection and analysis. When data are collected and analyzed by experts, established scientific practices govern how they are handled. By contrast, when individuals are collecting data through cheap, unverified, uncalibrated sensors, the immediate fear is "junk data." This may be merely incidental. Cell phone images frame what the photographer sees and wants to show, with no pretense about neutrality or comprehensiveness. Or it may be more purposeful since there may be no commitment to epistemic objectivity. For example, neighbors documenting traffic focus on congestion and collect data accordingly, rarely recording periods of relatively light traffic. Further, when mathematically unsophisticated individuals without formal training are interpreting data, the immediate fear is "garbage analysis." Arguably, sensing by the masses guarantees bad science.

[16] Second, consider the problem of observation-interactivity. Observation generally and surveillance specifically alters human behavior. For example, video cameras, for traffic or security, are explicitly intended to alter behavior. When data collection is situated "outside" the thing being studied, observation remains arguably neutral. But when data collection is embedded among the actors within a setting, as in participant observation, a cycle of interactivity is launched, in which observation changes behavior changes observation and so on. For example, ubiquitous audio recording might alter what's assumed to be

[6] See CSTB report Radio Frequency Identification Technologies. Washington, DC: The National Academies Press, 2004.

"on the record" in even seemingly private conversations, such as a law school seminar on affirmative action.

[17] Because of unprofessional data processing and inevitable observation-interactivity, urban sensing may be anything but good science. We believe, however, that this concern is overblown and explain why below.[7] More provocatively, we embrace the idea that urban sensing can and should go beyond science into the realm of art and politics.

[18] To make this discussion more concrete, consider, for example the D-Tower, created by Lars Spruybroek/NOX for the city of Doetinchem in the Netherlands. The D-Tower is a public sculpture activated by responses to a website that surveys the mood of the townspeople. If most of the Doetinchemers are feeling fearful, it glows yellow, but when they're in love, the beacon burns red. Here we've added an entirely new dimension to urban sensing – an aesthetic one in which data (responses to the on-line mood survey) are subjected to a heuristic devised for purposes of pleasure, humor, curiosity, and to a lesser extent, information. Contributing to the town's mood-reading through the web is motivated by those same factors, rather than by some scientific ideal like truth. A number of cities, public institutions, and designers are collaborating on similar urban sensing projects, creating dynamic events that engage the citizenry. These "sensing" mechanisms by generating "publicity" intend to spark action, and interaction, rather than merely record it.

[19] As another example, consider "Fade to Black," a project in which Natalie Jeremijenko mounted web-cams on rooftops aimed toward the sky. Over time, they would collect particulate matter on their lenses; the darkening images need little explanation, as we watch the sky over Houston or New York City change from blue to cloudy grey, and eventually, all together black.

[20] When we observe the D-Tower, we can hardly ask whether the citizens are truthfully in love when it glows red, but we can debate whether the tower is responsive to and provokes some community-feeling. (Perhaps in the case of D-Tower, our investment in the outcome has less to do with truth-telling than with participation rates. Again, with sufficient coverage — with hundreds or thousands submitting forms

[7] Even experts quarrel. The debate surrounding adjustment of the 1990 US Census is a classic example; the point being that in all but the most careful designs ("design" referring to an experimental plan), there is room for disagreement even between experts.

– the truthfulness of a few responses is less important.) Likewise, in Fade to Black, we do not know if the particles accumulating on the lens have come from a freak building fire nearby. Still, the demonstration is intrinsically provocative and initiates relevant if subjective debate.

C. The Data Commons

[21] In going beyond scientists and beyond science, urban sensing has the potential to generate a "data commons." By this, we mean a data repository generated through decentralized collection, shared freely, and amenable to distributed sense-making not only for the pursuit of science but also advocacy, art, play, and politics. Defined this way, the data commons resembles what we have previously called a public sphere.

[22] In prior work, Kang and Cuff provided a minimalist definition of the public sphere with four principal attributes: the public sphere must (i) be accessible to diverse members; (ii) provide opportunity for multiple uses; (iii) encourage some sort of (and not always political) exchange among participants (in the case of a data commons, this implies both the sharing and consumption of information); (iv) and be recognizable as such a space. Although these attributes were used to describe physical realms and social practices, they can also be usefully applied to the data commons, which is both a repository for information as well as a set of practices that can operate upon that data.

[23] We are enthusiastic about the possibility of a data commons for many of the same reasons that we are enthusiastic about maintaining the public sphere. In particular, we are skeptical that civic and political commitments will continue to be manifested—if they ever were—through stylized, romantic practices such as voting, political contributions, and face-to-face participation in local townhall meetings. Instead, we believe that individuals will increasingly manifest their civics and politics through engagement of a public sphere unerstood far more broadly. This includes for instance "political shopping," in which individuals use pervasive urban computing to inform their marketplace decisions to further nonmarketplace values.[8] Indeed, in the modern globalized capitalist environment, individuals express their social and political values as much through the consumption choices they make as they do through voting, lobbying, or political donations. Just recently in the United States, U2's Bono and Oprah Winfrey launched the "Product

[8] Kang & Cuff, PerC; Kysar, Preference for Processes.

Red" program, which contributes profits to anti-AIDs initiatives in Africa, with a splashy red iPod.

[24] We also believe that our civic participation will increasingly be manifested through our contributing to the data commons. Current examples that foreshadow this process include reviews of products, tagging of data, comments on blogs, uploading of photographs and other information when newsworthy events talke place and citizen-sensors are there to capture the moment. For example, [murmur], an audio story telling project in Toronto, has placed signs near nondescript local landmarks with a phone number that allows passing cell users to call and hear an individual's personal narrative about the place. In London, a cyclist outfitted his backpack with a carbon monoxide sensor to detect the air quality along his daily commute,[9] and shared that data publicly. In Washington D.C., PigeonBlog equips homing pigeons with GPS enabled air pollution sensors that send real-time, location-specific information to web-based maps. In Tel Aviv, Cellint's TrafficSense system[10] detects location and movement of cellular signals to create a real-time map of what is happening on the road.

[25] In such examples, notice the significance of location and locality. Indeed, many blogs can be thought of as local reporting; recall how we depended on blogs from the site of natural disasters or armed invasions. The very premise of citizen-science applications is to leverage local knowledge in identifying local birds or reporting the strength of an earthquake at the reporter's location. We cite a recent development, the so-called placeblog. These are sites that function somewhere between a local paper and a blog; they aspire to record the details of a place. Similar moves can be found in the upcoming daylife.com and newassignment.net. Such initiatives are consistent with the spirit of participatory GIS (PGIS), which explicitly enlists the community to make a case or to study some aspect of life locally.

[26] Consider finally Neighborhood Knowledge Los Angeles (NKLA),[11] which is an interactive website intended to stem the deterioration of housing and neighborhoods. The UCLA-based site offers data and mapping tools to improve communities. It gives residents the opportunity to document neighborhood problems (see Figure 2); it gives potential builders the location of appropriate sites for infill housing

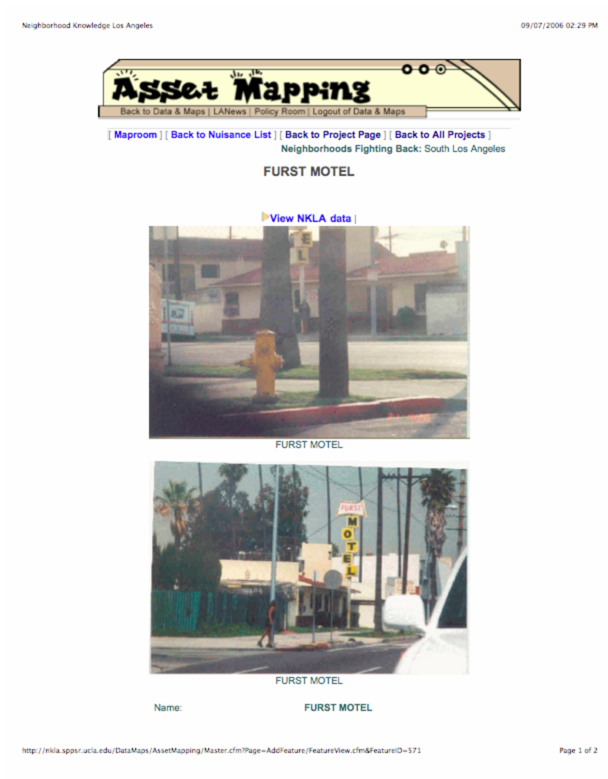
[9] (New Scientist, 9 Sept 06,p 26).

[10] Is this an opt-in system?

[11] The NKLA project was spearheaded by UCLA planner Neal Richman in 1995, with the current website launched in 2001.

development; it provides zip code and census tract demographics. Users can customize maps for specific purposes, such as neighborhood organizations tracking crime or traffic problems.[12] The value of NKLA, like other geographic information systems, is its vast database that is spatially organized and accessible to the public.

[27] In contrast to the James Reserve, users of NKLA move through the layers of information according to their own interests. Rather than receiving the researchers' account of the data, raw data are pieced together by users, who then act on their own understandings. Another important difference is that NKLA can be more readily used with an explicitly non-scientifc agenda: to convince viewers that a neighborhood is plagued by prostitution, for example.



---

[12] Customization of the system has been far less common and far more difficult than its creators had hoped.

III: BUILDING THE DATA COMMONS

[28] Our description of urban sensing and the data commons might be criticized as academic (in the pejorative sense) because the data commons already exists and will come into being without any policy intervention or academic guidance. See Flickr, YouTube, the blogsphere, the basic algorithm of a Google search. But as this Part demonstrates, even these participatory endeavors are no guarantee of a future with substantial data collection, sharing, processing, and practical legibility. Instead, we believe that how and whether a data commons is built depends on legal, policy, and technological (especially user interface) decisions we make now. Therefore, ours is not solely a descriptive project that maps some inevitable progression. It is also a normative engagement that identifies better and worse fates, and how we might collectively guide urban sensing toward the right result.

A. Property

[29] The legal institution of (intellectual) property United States copyright law only protects creative expressions; it does not protect the underlying facts or the data itself. Accordingly, once the data are made freely available, the person who labored to collect and upload that data will not be able to monetize their processing in any easy way. If so, one could be concerned that the lack of intellectual property protection for data will undermine both the incentives to collect data and assemble databases. This has been precisely the argument in favor of creating sui generis intellectual property rights over databases.[13]

[30] We believe that this concern—that a lack of intellectual property rights will inhibit data collection and sharing—is entirely overblown in the context of urban sensing. Most important, urban sensing and participation within a data commons will not be driven by financial self-interest. Rather, as we have already suggested, urban sensing will become a part of aesthetic, civil, and political partcipation that is only partly driven by market relations. Countless examples of cooperation, collaboration, and even play, especially mediated through the Internet, demonstrate that many substantial projects are not driven by the

[13] The European Union has created such sui generis rights. Recent empirical research raises skepticism whether these rights have in fact increased the incentives to create databases. See [cite]

prospect of financial remuneration.[14] [add discussion of local motivation]

[31] We can point to data sites like the weatherunderground.com or other "hobbyist" applications in which people build up reputations within the community because of their authority in terms of contributions of high-quality data. Reputation is a strong motivator.

[32] Organizing data can also be motivated as a kind of game or competition. The Google Image Labler is one good example (http://images.google.com/imagelabeler/). Here, people are randomly paired with partners who work together to tag images.

[33] A similar kind of organizational tool is offered by Amazon.com through its Mechanical Turk. Here, participants complete human intelligence tasks (HITs) for small payments. The HITs are advertised as tasks that humans excel, but computers struggle with. In one such example, Aaron Koblin offered 0.02$ for left-facing images of sheep, submitted through a custom Java-based drawing tablet running on the participant's web browser. Tens of thousands of sheep were submitted, attesting to a rather incredible participatory enthusiasm.

[34] We do notice that one of the biggest motivators for citizen-sensing to share data may be attribution. From the dawn of the world wide web, content providers have tabulated hits (of the pageview variety) to determine how many people accessed their site and how they got there. With the rise of blogs, the link's the thing. In both cases, people providing data want to know who else is consuming it. Even the Backyard Bird Count and the USGS's 'Did you feel it?" web sites allow data collectors to see how their data contributed to a larger whole. The same will be true for citizen-initiated sensing. And such attribution can be designed without any expansion of intellectual property rights over data.

[35] If we created robust intellectual property rights over data (for instance, because we mistakenly thought that such rights were necessary to induce participation), we might run into a tragedy of the

[1] [14] We point out that the building a data commons does not preclude building private databases (just as the existence of a public sphere does not preclude, in any way, more private spheres). Firms that want to expend resources to collect, assemble, and cut data can do so. Various laws, such as trespass, contracts, and trade secrets, as well as self-help mechanisms, such as security and encrypted access, can allow these firms to share their data on a restricted basis.

anticommons, a concept introduced by Michael Heller.[15] If too many property rights are created, he argues, the costs of coordinating permissions among multiple, fragmented property rights owners prevent otherwise efficient or socially valuable projects. This insight has been usefully applied, for instance, to the patents over expressed sequence tags (ESTs), which are sequences of DNA base pairs.

[36] To provide an urban sensing example, imagine digital images uploaded to some photo sharing site. Soon these images will by default be embedded with geolocation data, as GPS circuitry is added to digital cameras and multi-functioned indicators. With such information, a third party could plausibly assemble proximate photographs and digitally stitch them together to provide something like a "street-eye view" photographic map. However, if such visualization required IP "clearances" from each and every owner of the data intellectual property right, the transaction costs — even imagining efficient intermediaries — may be too high to justify.

[37] Thinking that such property rights are necessary fails to understand how participation in the data commons is not at its core a marketplace interaction. Further, the possibility of an anticommons tragedy counsels against creating intellectual property rights lest transaction costs prohibit interesting, useful, and dynamic engagement with the data.

B. Privacy

[38] By definition, urban sensing collects information in urban environments. Since those environments are inhabited by and directly connected to human beings, the data collected will often constitute *personal* information. Accordingly, urban sensing raises serious privacy concerns, in a way that surveillance in the woods largely avoided. To be clear, by privacy, we mean *information* privacy, an individual's claim to control how personal data are processed (collected, distributed, and processed).

[39] A patchwork of privacy laws already pertains to various aspects of urban sensing. Existing limitations such as prohibitions on video recording or audio taping under certain conditions could prevent various forms of urban sensing. In addition, the common law tort of invasion of

[15] Michael Heller, the tragedy of the anti-commons, Harvard Law Review.

privacy may also act as a constraint.[16] For instance, imagine sensors that could detect body temperature and pheremones to tell whether a woman is pregnant. Or imagine sensors that can detect a toxin coming from your neighbor. Again, depending on what the underlying technology is, it is possible that the neighbor could state a plausible privacy claim.[17] Rather than providing any standard positive legal analysis of potential privacy claims, we focus on two less observed aspects of the privacy problem: self-surveillance and network solutions.

[40] Typically, we tend to think of privacy claims being stated by the target of observation. In other words, imagine an individual with a cell phone that records background audio.[18] Her brief conversation with a friend is recorded without his knowledge. Our attention is most sharply focused on the friend's privacy claim, and rightly so.

[41] But in our example, the cell phone records also the cell phone owner. Because sensors will often be carried on our bodies, in our automobiles, or sited on our real property, the persons about whom most information will be collected is ourselves. Persuading individuals to engage in such constant self-surveillance and then subsequently to share that data, are themselves nontrivial hurdles entirely independent of and additional to the privacy claims raised by others. This is so even in the world of JennyCam and YouTube exhibitionism.

[42] Whether we decide to engage in self-surveillance, for the purposes of urban sensing, depends in part on what the underlying computing technologies allow us to do. For example, if computer security is weak and information collected even only for personal use are vulnerable to third-party hacks, we will be less likely to collect that information in the first place. Similarly, if personal data cannot be easily scrubbed to become anonymous or the pseudonymous, P will be less likely to share that data publicly.

---

[16] This common law tort is actually a combination of four different torts: (i) intrusion upon one's seclusion; (ii) misappropriation of one's name or likeness; (iii) public disclosure of private facts; and (iv) publicity that puts one in a false light.

[17] Cf. Kyllo (thermal imaging case). Other torts may also be implicated If I collect some data that pertains to you, and it turns out to be inaccurate, might I be subject to defamation? Will such a cause of action turn on a showing of negligence, or might there be a stricter liability?

[18] It might continuously record such content, or be triggered by some realization event such as an interesting sound, rapid acceleration/deceleration, geographic proximity to another communicator, and the like. New projects make use of the mobile phone, running a process on the phone that records data at regular intervals. This kind of image loop is becoming increasingly popular. [reference the software raccoon]

[43] We expect that the network itself will develop services to help individuals negotiate their various privacy relationships. For example, two of the most extensively studied problems in traditional sensor networks are localization and time synchronization. The network knows (or will shortly know) precisely when and where data are published. While these two pieces of data are critical for scientific applications, they are the subject of privacy concerns in urban settings. The network could, if properly designed, implement a kind of resolution control by verifying data up to the resolution that a user is comfortable with. Certify that this measurement was taken in the last week in this census tract; certify that this measurement was taken in the last hour on this city block. The network knows time and location with precision, but can verify these quantities using some notion of user-defined resolution. The tighter the resolution, the more useful the data downstream; but this choice could be left up to the individual provider.

[44] Of course, whenever we think about "choice" in the context of privacy, we must recognize that such choice and the preferences for privacy depend heavily on the background culture. Privacy preferences are adaptive, as should be evident from a cursory analysis of the kinds of information disclosure considered "sensitive" over time. (For example, consider the list of body parts that could be exposed comfortably in public over time.) Thus, one could easily imagine data collection practices that seem exotic and disturbing now appearing entirely banal 10 years hence. Consider, for instance, the default assumption that one is not being videotaped in public spaces. That assumption may not be as accurate as we presume. But, we can easily imagine a near future in which the default assumption is flipped — that one is presumptively videotaped in any urban public space, that nothing much can be done about it, and finally that nothing is especially disturbing or unusual about such surveillance environments. To the extent that privacy preferences are adaptive to the environment in this manner, the policy choices we make today will have long-term path dependent effects.

[45] Given this endogeneity of preferences, any claim that we are achieving some reasonable expectation of privacy through technological controls cannot be deeply satisfying. Further, recent privacy scholarship has rightly challenged whether individual control of personal information should be the defining hallmark of privacy protection. Still, in our attempt to articulate those variables that will influence the successful building of a data commons, we think it important to identify

the self-surveillance aspect of the problem and to highlight potential network solutions.

C. Liability

[46] Even in a fully decentralized model of the data commons, there is likely to be some database intermediaries which provide platforms for individuals to upload their data. Current examples include firms such as Google, MySpace, and YouTube. If these intermediaries can be held liable for the data that third parties have uploaded, they will be less likely to provide such platforms. This question about intermediary liability has been thoroughly vetted in the legal literature, especially as regards to third-party copyright liability. But, our discussion focuses on data, to which copyright does not attach.

[47] On this matter, we highlight some unexpected legal reinforcements that buttress the creation of a data commons by immunizing intermediaries. Consider 47 U.S.C § 230, which was passed as part of the Children's Decency Act, which itself was part of the 1996 Telecommunications Act. Generally speaking, if some unrelated person uploads data onto some data commons platform provided by a data intermediary, that intermediary may not be held liable for the damage that that data might do.[19] In practice, this means that Google can provide database infrastructure for the general public to use, with little concern about liability about what data are in fact uploaded.[20]

[48] In addition, anti-Strategic Litigation against Public Participation (anti-SLAPP) statutes have been passed in a substantial number of states.[21] These statutes were originally enacted to make sure that whistleblowers and political speech would not be chilled by aggressive litigation by large defendant corporations. But such a statute has been used to prevent overreaching by those who would stop data sharing within a data commons.

[49] The celebrated case concerns Barbra Streisand, or more particularly her home. The California Coastal Records Project (CCRP) has taken over 12,000 photographs of the California coast since 2002, in part to

[19] This section reads: "no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." This means that ISPs, websites, chat rooms, or other intermediaries that function as "providers" of an "interactive computer service" will not be held liable for the content uploaded by some third party.

[20] Note important exceptions with intellectual-property.

[21] [SWAT: how many?]

document incremental changes along the Pacific coastline. But some of those pictures included aerial shots of Streisand's house. Asserting a privacy violation, she sued. The CCRP responded with a special motion to dismiss under the California anti-SLAPP statute, and won that motion because Streisand did not have any colorable privacy claim. More important than the quick victory is that the statute requires paying of attorneys' fees.

D. Interface

[50] Even if individuals are motivated to participate and the underlying legal regimes make it possible to do so, we must recognize how significant user interface is to both data collection and interpretation. For example, if collecting and uploading ("slogging") local pollution data is too difficult or too expensive in direct or indirect costs, folks will simply avoid the hassle.

[51] Part of the success of blogs can be attributed to extremely simple tools for creating and publishing content. The same is true for mapping or GIS applications. Be it NKLA or more "mainstream" sites like flickr.com, mappr.com, or communitywalk.com, it is now relatively easy to create maps of geocoded information. Perhaps more important than content creation, the simple publication or sharing of this information is critical as are distribution mechanisms like RSS that allow people to register interest in content, and (thanks to reblogging tools) republish selected portions. This pipeline or stream model is not dissimilar from what we might expect from citizen-sensing. Easy data discovery, subscription and republication will be crucial.

[52] Far from a database query, it would be a reasonable outgrowth of existing technologies if the data commons were built from disparate sources of shared data, (following simple publication mechanisms), informally organized (as with meta tags), open for discovery, visualization, and comparison, and subject to republication (modeling feed-forward). In part, discovery in the data commons might borrow from these existing services, relying on republication/aggregation/modeling as a kind of "link" between data sources (I fit a regression using your data, and a link is established).

[53] There are large implications for discovery/search. Currently, we navigate a large collection of essentially text and image data via a search engine; lists or tables of results are returned to us, and although millions of items are identified, we rarely examine more than a few handfuls. Are

lists the right structure for discovery within the data commons? Given the significance of space and location, is a mapping metaphor more appropriate? As information timeliness and trends becomes increasingly important (consider, for example, traffic data or ocean safety alerts), might there be other ways to present the data? The dynamic character of the blogosphere, for instance, has resulted in other kinds of summaries (trend displays and so on) to reflect the fact that sources are constantly updating, perhaps in response to current events. Finally, what structure will allow consumers of data to make the kinds of trustworthiness evaluations that are needed to make sense of data? All this, while making sure that the data are amenable to simple, easily readable visualizations.[22]

IV. THE FATES

[54] So far, we have traced how pervasive computing might move out of the woods into the urban context. We have stressed the significance of a data commons, and the legal and technological factors that make its existence more or less likely. But the mere existence of a data commons is not a panacea; the data commons is essentially public infrastructure that can be used for multiple purposes. It is a powerful resource that can be misused or under-realized. In this final Part, we forecast those potential fates portrayed in Greek mythology as Sirens and Cyclops.[23]

A. Sirens

[55] A mantra in the computer science field of embedded network sensing is that it will "make the invisible visible." This has already taken place in scientifically controlled natural environments. It will likely soon take place within our cities, through decentralized processes, generally often without scientific goals or consequences. However, like a Siren the new vision may seduce us toward bad ends or to make poor choices.[24]

[56] First, we may be seduced by *junk data.*

[57] Second, even if the data are accurate, we may be seeing a *statistical mirage.* In statistical parlance, the term "data mining" originally had a

[22] [Note: In a conversation with the Graphics Editor of the National Desk at the New York Times, I was told that you'll never see a scatterplot in the Times; They get lots of complaints when they've tried it; while the bread and butter of statistical practice, these plots are not intuitive to many.]

[23] Medusa for later.

[24] We realize that we're mixing metaphors here since Sirens influenced through their audio more than video; our retort is that in a digital world, bits are bits.

negative connotation. It referred to the process of digging through the data, posing multiple hypothesis tests, each having a probability of incorrectly detecting some kind of feature or structure in the data (a Type I error). In short, making multiple comparisons is like tossing a coin; even if the chance of seeing heads is small, if you toss it often enough, you're likely to see at least one head. In a world where many people are looking at the same data set, the number of interpretations multiply and with them, the chance of data rumor or hearsay. John Tukey, the statistician who literally wrote the book on exploratory data analysis, coined the collective noun "a quarrel" for a group of statisticians reflecting, in part, the fact that while it is possible to get general agreement about an analysis when the data were collected as part of a careful experiment, ad hoc collections of data can be tricky. Even with simple visualization tools like maps, making inferences about associations is complicated. Now, add the effects of "data mining," the active search for associations on the part of many consumers of the data commons, and we have an even more difficult scenario.

[58] Third, even if there is no data mirage, highly salient data or associations might seduce us to *instrumentally irrational decisions.* More information does not necessarily produce more rational (in the sense of instrumental efficacy) decision-making.[25] Well-known cognitive biases might lead us to pay more attention to particular types of data than they rightly deserve. Consider, for example, the rough and ready risk calculation that individuals undertake in deciding where to live. If the data commons offer readable, highly salient depictions of reported violent crime rates in the city, such information might persuade people to move out to the distant suburbs in spite of the far greater mortality risk created through the increased amount of highway driving.

[59] Finally, even if we are making rational decisions based on the data and their interpretations, we could be funneling down a vortext of *self-fulfilling prophecy.* Suppose that we in fact have perfect information about the distribution of violent crime by ZIP code. After objectively considering associated risks and benefits, we conclude that it remains in our self-interest to avoid various portions of the city even though doing so will increase racial segregation as well as environmental pollution (due to the longer commutes). Relying on this highly salient, unidimensional "crime statistic" sort could produce a positive feedback loop that make those areas with high crime rates grow more dangerous,

[25] Hanna experiment?

while those areas with low crime rates get ever lower crime rates. New sensing capabilities can make far more insidious and less visible factors enter into residential location choice, such as buried toxic waste sites or ground water toxins.

[60] Given these dangers, what if anything might we do to avoid the Siren call? Before answering this question head-on, we want to be clear about the we are not comparing a data commons to some information utopia verging upon collective omniscience. Rather, we are comparing a data commons to the status quo, which is partial sight also infected with junk data, statistical mirages, instrumentally irrational decisions, and a self fulfilling prophecies.

[61] First, we want not to exaggerate the Siren problem. For example, the junkiness of data depends on the purposes for which they are used. Issues of data quality may not be particularly relevant for more playful or artistic applications. By contrast, data that are tendered for traditional scientific applications, especially with policy implications, must satisfy far greater rigor.

[62] Second, we point out the possibility of various forms of distributed accountability that make data collection more reliable. For instance, a user may tell the network which sensors should "agree" with her measurements. The more checks users provide when they publish the data, the more trusted the data become.

[63] In addition, instead of going all the way to full decentralization, certain portions of the data commons can be maintained under a distributed citizen-sensing model. [The more organization is present binding participants, the greater the role for web services to help citizens with the planning or staging of measurement activities, so that the resulting data is as interpretable as possible, perhaps corresponding to one or more standards of evidence employed by the relevant governing body. We see precedents for this with newassignment.net and daylife.com in the context of citizen journalism, where citizens are participating to help a specific end-goal.]

[64] Finally, the sensing network itself given its ubiquity can provide the redundancy necessary to identify and interrogate faulty data. Even in so-called bottom-up systems, guarantors of data quality exist. In some sense, this is how we have come to identify reliable sources on the Web; search engines like Google return millions or possibly relevant sources of information, but we understand how to navigate these lists and identify reliable sources. Redundancy was (is?) the promise of an early vision of

sensor networks. In reality, we will never approach an asymptotic sensing limit; this means we might have to rely on network or web services to help us assess trustworthy data.

[65] Third, to minmize mirages, irrationalities, or self-fulfilling prophecies, we encourage data presentation practices that encourage what might be called "rights of reply." Just as blogs often provide spaces for comments, we envision data visualizations linking, perhaps, in some automated manner to counter interpretations that might counter the Sirens with its own melody.[26] If crime rates are being associated with race, a link might provide another visualization that controls also for wealth and income. If race is being associated with siting of SuperFund sites, a link might provide another visualization that shows how a time series of poorer people coming *after* the toxic sites' siting, with racial consequences.

B. Cyclops

[66] In one version of the tale, the Cyclops is granted the power to see the future: again, the invisible suddenly becomes visible. Unfortunately, the Cyclops is deeply saddened because the only future that he is allowed to see is the circumstance of his inevitable death. As they say, when the gods want to punish you, they grant you your wish.

[67] The tragedy of the Cyclops–that is, the impossibility of effecting change, might be visited upon us as part of the data commons. For example, it is possible that distributed environmental sensors could detail with alarming precision the nature and extent of our environmental poisoning. That poisoning could be a function of microclimates sharply delineated (in space) by highway overpasses and (in time) by length of rides on diesel-burning school buses. Those with resources may be able to respond to such information; however, what about those who lack the financial ability to change their circumstance? Quite possibly, what they are left with is the debilitating information about the nature of their demise without any practical ability to change their circumstance.

[68] There may also be political versions of such tragedies. For instance, a distributed network of airplane enthusiasts helped track CIA airplanes hopping from one country to another, which suggested the transport of prisoners into secret European detention centers. But for those who thought such practices were anathema, they were largely powerless to do

[26] This would be the Orpheus strategy. Jason and the Arognauts successfully navigated past the Sirens by having Orpheus sing with his lyre a melody even more beautifuly than the Sirens.

anything about it. For example that crosses the political aisle, consider the widespread deployment of sensors that better track the amount of unauthorized migration across the United States and Mexican border.[27] Notwithstanding self-help, these data may have no practical impact on the net migration, which would require concerted political action from the federal government.

[69] Greater transparency (through urban sensing) could be leveraged to greater accountability on the part of our social and political institutions. We are surely seen such examples with the smoking gun documents and off-the-cuff statements captured on YouTube.[28] Data collection can be a form of whistleblowing, and we reject the sort of cynicism that claims that such information can never alter policy.

[70] Without attempting any grand theory about how and when new information might catalyze change in political, social, and economic systems, we offer one novel idea: arbitrage our ignorance. This draws on the idea of a "veil of ignorance" offered by political philosopher John Rawls. He famously argued in favor of adopting principles of justice that would be agreed upon by persons in an ideal choice position (called the "original position"), which included deliberation behind a "veil of ignorance." This veil prevented persons from knowing what station of life they would find themselves in.

[71] As applied to urban sensing, the idea is this: if urban sensing lifts the veil, by making the invisible visible, is there a way to create some consensus before we learn the new information? After the information arrives, the predictable reaction is for the rich and powerful to remain rich and powerful, and to justify their response through motivated reasoning. But by pre-committing to a particular principled response, before the veil is lifted, we may be able to mobilize the sort of collective resources necessary to avoid a tragedy of the Cyclops.

CONCLUSION

[72] Embedded network sensing has made the leap from laboratory to the natural environment through the careful design of professional scientists. It is now crossing into the urban context, but leaving behind the primacy of both scientists and science. In so doing, urban sensing is enabling a data commons that is essential infrastructure for citizen participation in politics, civics, and aesthetics—as well as science. What

[27] Web cams at the border?
[28] Allen and macaca.

we do today will influence what the data commons comes to be tomorrow. And only through deliberative effort and political engagement can we try to inculcate the spirit of citizen navigators savvy to Siren calls and the tragic Cyclops.