

**Ayoung Yoon**, Assistant Professor, Department of Library and Information Science,  
Indiana University Purdue University Indianapolis (IUPUI), [ayyoon@iupui.edu](mailto:ayyoon@iupui.edu)

**Kyle M. L. Jones**, Assistant Professor, Department of Library and Information Science,  
Indiana University Purdue University Indianapolis (IUPUI), [kmlj@iupui.edu](mailto:kmlj@iupui.edu)

**Lydia Spotts**, MLS Adjunct Lecturer, Department of Library and Information Science,  
Indiana University Purdue University Indianapolis (IUPUI), [lspotts@iupui.edu](mailto:lspotts@iupui.edu)

# Ethical Use of Lifelogging Data for Research

## Perceived Value and Privacy Concerns of Wearable Camera Users

### Abstract

*Video data generated from wearable cameras is now available online, as the concept of “lifelogging” has been introduced to many citizens due to the spread of wearable camera equipment. Usually, these wearable cameras automatically capture images or record videos from a first-person point of view; they collect a new form of information that cannot be captured through other means. Citizen data Harvest in Motion Everywhere (CHIME) project pays close attention to the value of this new type of resource, particularly regarding the video data that cyclists record using wearable cameras over a long period of time. These contextually rich data capture community members’ infrastructure experiences and interactions with other transit modes, as well as environmental changes. If curated and made publicly accessible, there is great potential for various stakeholders, including public historians, researchers, city planners, and citizens, to use the data. However, making these videos open to the public and to researchers raises ethics issues, as the data include sensitive, location-based information that may intrude into private lives. Additionally, the videos include the accidental collection of data from secondary participants (bystanders). In this paper, we will describe the potential value of citizen-generated video data using the CHIME project example and discuss the privacy and ethical considerations related to the use of this type of data for scientific and citizen research.*

**Keywords:** lifelogging data, data reuse, ethics, value of data

### Introduction

Wearable technologies gave rise to the concept of “lifelogging.” Today, Google Glass, the Apple Watch, the Narrative Clip, Fitbit, and the GoPro camera represent common commercial technologies individuals use to

lifelog. These devices generally produce a dataset that consists of continuous streams of sensor data (Rawassizadeh, 2012). There are many benefits for individual users of using lifelogging devices, for instance: documenting a neighborhood, gathering evidence to make a case, enhancing memory retention, and monitoring personal behaviors (Hoyle et al., 2014; Shilton et al., 2009).

Users often share these lifelogging data online through social media for different reasons, such as to keep a personal record and to share information with their network of friends or even with the public. The shared data introduce new benefits to society in many ways beyond personal use, including in historical studies, crowd health monitoring, and group behavioral studies (Eagle & Pentland, 2009). Researchers are now interested in not just using lifelogging data that are already publicly available but also encouraging users to share their lifelogging data with citizen science projects. Many citizen science projects engage the public as scientific data contributors, and the collected data from citizens become available online to public (e.g., *Zooniverse family of projects*, Savage, 2012; *eBird project*, Sullivan et al., 2009).

Among the various types of lifelogging devices, wearable cameras usher in a new evolution of lifelogging applications, as they capture and produce image and/or video data from a first-person perspective, allowing for the documentation of citizens' daily lives (Mann, 1997; Templeman, Korayem, Crandall, & Kapadia, 2014). Sharing private sensor data (e.g., geo-location) collected by existing lifelogging devices is not new (e.g., Toch et al., 2010; Wang et al., 2013), but sharing image- or video-based lifelogging data (first-person camera footage) is considerably different because of the rich information contained in the data (Hoyle et al., 2014).

Despite the wide potential and known benefits of wearable cameras, one type of lifelogging devices, to users and society, sharing data from wearable cameras raises ethics issues and privacy questions. Collected data often include personal, private, sensitive, and confidential information—not only of the users themselves but also of their friends, family, and even strangers around them. And making these data publicly available or posting them on social media, even to a network of friends, introduces the risk that personal information might be shared with unintended parties (Motti & Caine, 2015). Kelly et al. (2013) argued that potential harm from data becoming public is greater with image information, such as video data from wearable cameras, and the confidentiality and security of the data take on a greater importance for that reason. These concerns increase when even publicly shared data are reused beyond the original purpose of sharing or the intentions of the data producers, the users of wearable devices (see Nissenbaum, 2010).

Previous studies have explored the issue of privacy in relation to wearable devices, as this is one of the main challenges yet to be solved in wearable technology. However, many studies focused on developing a privacy protection mechanism (e.g., Chaudhari, Cheung, & Venkatesh, 2007; Templeman et al., 2014) and on the ethical use of wearable cameras for research purposes (e.g., Doherty et al., 2013; Kelly et al., 2013). Several studies addressed users' privacy concerns (e.g., Hoyle et al., 2014; Motti & Caine, 2015), yet little is known about users' perspectives on their data and their views on privacy considering the use of the data in various usage contexts. In addition, many studies aim to help lifeloggers better manage their privacy, but they do not address privacy and ethical considerations for sharing and reusing the data from wearable cameras, even though it seems important to understand users' perspectives as data producers.

The purpose of this paper is to explore the perception of data producers regarding their lifelogging data that can be shared as a potential research source, focusing on their perceived value, as well as to explore the potential privacy concerns of producers in sharing the data and making them available for various reuse purposes. Our focus is on video data recorded by cyclists' wearable cameras, particularly because of the richness of the data captured by the cyclists, such as their infrastructure experiences, interactions with other transit modes, and environmental changes, as well as on the potential for these data to be reused by various stakeholders, including public historians, researchers, city planners, and citizens.

Our study was conducted as a part of the Citizen data Harvest in Motion Everywhere (CHIME) project, which

aims to build a cyclist community-driven data archive by integrating cyclists' video data from wearable cameras. The long-term goal of the project is to develop a means for community members and scientists to use community-driven data archives to create, distribute, preserve, and analyze information and, ultimately, to improve society through a shared understanding. Learning about cyclists' views on their data to be shared and any privacy concerns is critical for this project to ethically integrate their lifelogging data into the data archive and reuse the data, because cyclist-generated data, which are current, digital, typically published to the web, and maintained on personal devices, will be an important source of the data archive.

### Literature review

#### *Value of lifelogging data*

The value of personal digital materials (e.g., digital photos, videos, personal website, blogs) to society has been well recognized in the field of archive and historical research. While individuals usually find sentimental, emotional, and legacy value in their personal digital materials (Kaye et al., 2006; Whittaker & Hirschberg, 2001), researchers see the bigger value of these materials: individuals' personal materials are an important part of instituting a collective memory and cultural identity, and so "evidence of me" becomes "evidence of us" (Cunningham, 1994; McKemmish, 1996). Significant parts of personal digital collections are now shared online, mostly through social media platforms, and individuals who published their personal materials online see the potential value when the materials are used by others (Yoon, 2013).

Lifelogging data are considered a type of personal digital materials. The data could record individuals' daily footprints, exposure to air pollution, exercise habits, and frequency of interactions with family and friends. The data can also include much information about individuals' regular locations, habits, and routes (Shilton et al., 2009). Individuals might find personal value in their lifelogging data, but they might also see other value in the data. Wolf, Schmidt, Bexheti, and Langheinrich (2014) found that individuals use wearable cameras to keep evidence in case of an unexpected event, such as a break in, an accident, an assault, or any misconduct. Similarly, other practical values include monitoring safety-relevant situations when involving infants and pets and providing healthcare assistance, such as working as a memory aid (Wolf et al., 2014).

Beyond the personal value found in the data, society benefits from widely sharing lifelogging data as well. Uses of the data in storytelling, historical studies, crowd health monitoring, and group behavioral studies are well-known examples (Eagle & Pentland, 2009). Rawassizadeh (2012) saw potentials and applications in business and scientific research, such as software personalization, learning social patterns and behaviors, recommendation systems, health and media studies, and surveillance studies. Business sectors have been predominantly using lifelogging data for new application and service developments, but previous studies promised opportunities as research sources. Considering the diverse data and the context of the data being created, the potential advantages and uses of these data are beyond what we expect now.

#### *Lifelogging data and privacy concerns*

Despite the benefits of lifelogging data to individuals and society, wearing a camera and capturing data from it raises ethical issues if not legal issues. Users of wearable devices are likely to capture personal or sensitive information about themselves or others around them, particularly when recording in public; consequently, privacy concerns are prevalent. Motti and Caine (2015) argue that privacy issues in this space are still poorly understood, and the nature of privacy concerns remains an open question.

One notable area of research is users' behaviors with wearable cameras to identify what counts as sensitive photos and what privacy concerns are associated with them (Hoyle et al., 2014). Motti and Caine (2015) reported various factors that affect users' privacy concerns, such as the nature of the data collected, their respective levels of confidentiality and sensitiveness, the ability to share and disclose the information, and potential social or criminal implications. These factors were valid across different types of devices, but cameras and microphones with GPS resulted in extreme concerns. Kelly et al. (2013) already noted that the potential harms and risks related to data from cameras are greater than those related to other types of data. Users seem to understand the risks,

particularly when the data are shared with unknown or untrusted parties, as the data may be misused or used in ways users do not intend (Motti & Caine, 2015). Perhaps because of these concerns, past studies reported users wanted to control the data collected through their wearable devices (Kärkkäinen, Vaittinen, & Väänänen-Vainio-Mattila, 2010).

As important as the privacy of users is, the privacy of bystanders (e.g., secondary, unintended participants appearing in videos) is another area of concern. Few studies have addressed the privacy concerns of bystanders from users or bystanders' perspectives. In the study of bystanders' responses to wearable devices, Denning, Dehlawi, and Kohno (2014) found that the majority was either indifferent or negative toward the devices. A study from Hoyle et al. (2014) reported a higher level of acceptability, arguing that acceptability may depend on the level of obtrusiveness and the context of device use. Despite the difference, these studies demonstrated that bystanders may feel uncomfortable with being recorded by devices owned by others. One important finding is that users of wearable devices also care about the privacy of bystanders, and they are willing to discard, modify, or not share their data if it violates the bystanders' privacy (Hoyle et al., 2014).

While no one solution can address those privacy concerns, one approach is to develop a system that protects users' privacy by creating algorithms or systems that automatically detect sensitive information and take action. For instance, Al-Muhtadi, Ranganathan, Campbell, & Mickunas, (2003) and Cheng, Golubchik, and Kay (2004) proposed a system that decreased the accuracy of collected data (e.g., allowing users to be "invisible" so that the system did not record) to avoid privacy concerns, while Jana, Narayanan, & Shmatikov (2013) utilized a computer vision algorithm (e.g., image recognition, object tracking, and face detection) to extract features from data in a privacy preserving method. In addition, Chaudhari et al. (2007) developed a protocol that detects and obfuscates human faces recognized in lifelogging video streams, and Templeman et al. (2014) introduced an application that allows the users of wearable cameras to blacklist sensitive spaces (e.g., bathrooms and bedrooms) from image classification based on objects, colors, and textures detection.

These technical solutions would provide some safeguards, but complete protections are not always possible with technical solutions (Bowser & Wiggins, 2015) as human understandings of privacy concerns always drive these technology developments, and those understandings are often framed within socially accepted practices and cultures (Altman, 1977). Wolf et al. (2014) pointed out that capturing data from wearable cameras in public spaces and sharing them for public consumption (e.g., sharing with friends, posting online for non-commercial purposes) are legal in many countries, though laws vary. However, the fact that the act is legal does not necessarily mean that it is socially acceptable. For instance, in the U.S., the laws do not require individuals to consent to data captures and video recordings if they are in public places, but people express discomfort when captured by wearable devices, such as Google Glass (Wolf et al., 2014). In addition, because the usage of the lifelogging device and data is relatively new, Motti and Caine (2015) warned that people might not be aware of potential privacy risks and implications, especially when disparate data are aggregated. This gap between the law and the level of social acceptance and/or awareness raise a question of "ethical" research practices when lifelogging data are used as a research source.

## Methods

Because we were interested in understanding the perspectives of users (data producers) of wearable cameras, we conducted semi-structured interviews with cyclists who utilized wearable cameras and shared the footage on a social media platform or website. Our initial geographic scope was loosely limited to Indianapolis, IN, U.S., because Indianapolis is a city in which a notable cycling culture has emerged, alongside considerable public and private investments in bicycle paths, bikeways, and other infrastructure (Indianapolis Cultural Trail, 2017; Indy Parks, 2004; Simmons, 2014), including a successful bike-share program (Indianapolis Pacers Bikeshare, 2015; Touhy, 2015). We did not have any exclusion criteria for cyclists based on environment (e.g., road, trail, mountain, urban cycling) or purpose (e.g., leisure riders, commuters, competitive riders), thus any type of cyclist could participate in the study.

We identified study participants using purposive and convenient sampling strategies (Bernard, 2000), including searching online for cycling videos published on several social media platforms (e.g., YouTube, Twitter, Instagram) and outreach to cycling organizations and community leaders in Indianapolis. Social media was the primary medium for initial contact with potential participants whom we identified online, and we emailed relevant organizations and community leaders for broader advertisement. We also employed a snowball technique to reach out to other potential participants.

In total, 13 interviews (12 phone interviews, one in-person interview) were conducted from September to December 2016. The semi-structured interviews addressed participants' views on their data (values to them and to society), as well as views on sharing and reusing their data, including potential privacy concerns. Participants were encouraged to speak freely when they had more to share about a particular topic. The average interview lasted 30 minutes. All interviews were recorded either with a smartphone audio recorder application or a handheld digital audio recorder. A transcription vendor professionally transcribed the interviews.

Along with the interview audio and transcription data, we also asked participants to share video clips of their cycling experience to better our understanding of the nature of the video data. In total, 54 clips were collected from the participants through an IRB-approved cloud storage platform, Box, provided by Indiana University. Each participant received an email link to a private folder designated for uploading the videos.

The interview data were analyzed using Nvivo 11 for Mac, a qualitative data analysis software. The data were analyzed using pre-developed codes driven from our interview questions, including several high-level codes of cycling behaviors, perceived value, and concerns with use/reuse and sharing of own data, with a number of sub-codes that emerged from the analysis. The video data were used as secondary sources, as they provided more context regarding to what participants referred while informing of content characteristics (e.g., what and who are recorded, how they are represented, etc.).

## Findings

### *Characteristics of participants*

The participants were male dominant (11 males, 2 females), age-ranged from 20 to 66 (an average age of 40). The majority of our participants were Caucasian. These demographic characteristics aligns with cyclists' general demographic characteristics, although we did not collect a representative sample, as many surveys have reported that the majority of cyclists are men between 25 and 64 years old, and most recreational cyclists are white in U.S. (Pucher, Buehler, & Seinen, 2011; Pucher & Renne, 2003).

All of the participants were experienced cyclists. Many began riding during childhood and continued riding more consistently or intensely when they became an adult. An average year of riding was 10, with a range from 4 to 37 years during adulthood. About half participants (six) reported they rode daily or "nearly daily" year-round for fitness or transit, while remaining half (seven) rode between 1 and 7 days a week, primarily for health or dedicated athletic training. Most cyclists reported fitness as a motivation for riding, four ride purely for pleasure, recreation and fitness, or transit. Some of them (six) were more invested and dedicated than others and training for competitive cycling events, including criteriums, cyclo-cross events, road races, velodrome races, messenger competitions, Ironman competitions, and mountain bike races.

Most participants reported that they had used wearable cameras for less than 3 years, except for one who had worn one for more than 5 years. This is considerably less than the amount of time they had been cycling, but not surprising given the time when wearable cameras became more popular and affordable to general public. The most common types of wearable camera used by the participants were the GoPro (9), followed by the VIRB Garmin (2), JVC Action Camera (1), and Polaroid Waterproof Sports Action Video Camera (1). Participants reported that there was no real difference in terms of the technical specification for recording video (e.g., image resolution), and the videos produced from the cameras were not geo-coded (no GPS information attached).

### *Perceived value of video data to the public*

Almost all participants discussed why the video data were meaningful to them. The data helped them remember and prompted their memories of social connection, which had nostalgia value. At the same time, the data were a significant source of their security (evidentiary value) to protect them from unforeseen events or accidents. Many also noted that the videos were simply fun to watch later and good sources of entertainment. With different types of personal value, participants also thought that their videos could be useful to the public for various purposes: (1) cyclists' education, (2) promote cycling culture, (3) document drivers' behaviors toward cyclists, (4) capture public infrastructure and safety information, and (5) capture environmental and landscape change.

First, several participants said their videos were useful educational resources for other cyclists. It seemed participants found more educational value for other types of biking, such as mountain biking or bike races, that perhaps require more skills. C13 said the video shows novices "the right methods" of cycling on rough mountain roads. C04 said he knows race cyclists who watch the videos to see how the race should be played out and how to finish and win the race. C01, who is an urban cyclist but also works as a cycling instructor, emphasized the value of videos in educating people about how to ride safely in the city and said the videos were particularly good "to present reality and not some skewed version based on my own filters and biases."

Second, participants saw the potential of the videos as a source for promoting cycling culture. C06 thought the videos could be "invitational to others and help them see how fun the trails can be." C09 echoed that, and argued the usefulness would be even bigger for mountain bikes, which probably have a higher entry barrier for novice cyclists. "Cyclists see videos of extreme sports [like] mountain bikes. They don't understand what it really means to be out in these remote areas and the things we experience." Several believed their videos played that role, as they received positive feedback from other cyclists who watched their videos online. C11 was excited about the "small impact" the video made: "Some comments [from my YouTube channel] are like, 'I thought about [cycling], but I wasn't sure. (...) After watching yours, (...) I'm interested in doing it.'"

A number of participants also mentioned the threat from cars or motorcycles during their riding, especially in urban and city environments. Interestingly, their videos documented "the immense amount of aggressiveness that a lot of cyclists can face on the streets" (C03), "careless and distracting drivers" (C07), who also "honk just to let you they're displeased or for no particular reason" (C03). C07 said it was nice to have these kinds of montages over the ride, and watching the videos often ended with surprises upon recognizing the aggression towards cyclists. As dedicated cyclists, participants hoped the documentation of aggressive interactions, threats, and danger cyclists experienced on the road could "lead to legislative changes, changes in culture" (C01, C07). Not very many, but a few participants captured faulty behaviors or accidents by chance. C07 thought somebody may "need to use [the video] to find the one at fault or prove guilt."

As participants rode in different areas of the city, their videos included images of the infrastructure they used during the rides. Several mentioned that they intentionally or unintentionally filmed unsafe road conditions such as potholes, road hazards (e.g. icy roads), and poorly designed road/trails. One interesting observation was that the videos also captured the use or misuse of infrastructure, as C03 remembered that the videos showed "quite a lot of people over time that start their turns into a bike lane or park in it." He recalled that this was an issue, "particularly in places without good infrastructure for [bikes]," which raised a question of city planning and infrastructure design.

Finally, the videos inevitably captured landscape changes as longitudinal data. For mountain bikers, the changes were more of "the landscape and the colors of the rocks" (C09), which he considered a good source of knowing about changes in many rural or wild areas. The videos from urban cyclists were more about "[what's] knocked down or built new" (C05) or other changes such as new road construction, landmark signs, and city events. C01 believed that this type of information would not only be useful for some historical, longitudinal studies, but also public spaces design.

### *Privacy concerns for sharing*

All participants were already sharing their videos with someone; the majority were with the public through social media platforms (YouTube, Facebook, Twitter, and Instagram), while a few shared only with their network of friends. Many expressed excitement that others viewed their videos. Some with higher numbers of views (e.g. “[my clips] on YouTube [that I uploaded] a couple months ago got over 10,000 views!” (C08)) demonstrated high public interest. However, participants also expressed reasons for not sharing. Concerns on private or personal information in the videos was the most common, but participants also expressed fear of others’ judgments about their behaviors or negative comments/feedback for no reason. C13 noted, “Sometimes people get very wrapped up in their own perspective. (...) I feel like it causes a lot of arguments, They’ll say ‘Oh no, you’re supposed to be on the other side of the road.’ Or ‘You should be in the middle of the road.’ That kind of thing. (...) You can’t make everybody happy or convince everybody.” And even if they decided to share, some of their concerns were still valid, which made participants more selective about their videos. Some manipulated or edited the content before sharing.

Even though the focus of the video recordings was not to capture images of people and human behaviors, as people became a part of the video content, privacy issue naturally arose. The level of concern varied among the participants, though. Three participants did not have any privacy concerns when sharing videos. However, unlike those three, several participants had strong concerns about the possibility of revealing their personal information. Since some of them published the videos with their names (e.g. YouTube channel or credit information inserted in the videos), they seemed less concerned about their names being released with their videos, while others wanted to protect their names and preferred not to reveal them. Regardless of this difference, all cared significantly about revealing an area of their life, such as house location, neighborhood, and routine path of riding. C02 said, “There’s just so much that I’ve posted (...) If someone is going to track me down, they definitely could do it by now (and) find out where I live.” For this reason, C13 tried not to film within one or two kilometers of his home, and C05 used another application accompanying with the camera, which allows putting a barrier around his house.

Participants also showed different attitudes toward bystanders who were accidentally captured in their videos. For instance, C09 did not care about “anybody, either audio or faces in [the videos].” This was echoed by C02, who said, “they’re just someone randomly in the background who is passing by quickly.” On the other hand, several noted that they “worry about other people (...)” and “[are] very sensitive to the fact that other people have privacy concerns that may preclude them from their information being released” (C01). C01 was particularly concerned about children in the videos, filmed in public spaces (e.g. a park) or more private spaces (e.g. front yards, neighborhoods), as they were more vulnerable. C07 was also aware of the unintentional capture of personal information, such as license plates, during his ride. Both C01 and C07 carefully checked the content of their videos and edited as necessary by cutting out or blurring out to make sure “there’s no real obvious identifiers (C07),” but felt there was still a possibility that they missed some relevant information.

When the video captured any faulty behavior or accident, participants seemed to be more cautious, depending on the type of incident, despite any usefulness they expected. During his night ride, C03 captured a crash in which people drove into parked cars. He published the video online and tagged the police in case they needed the video. However, when C10 recorded an individual fault or violation, even though the behavior or accident was entirely that individual’s responsibility, he “didn’t want to cause a problem.”

## **Discussion**

### *Lifelogging data and the potential for CHIME*

An initial motivation for this study was to move towards a better understanding of cycling lifelogging data and content as a potential source for a community-driven data archive. The project team hypothesized during their planning of CHIME that cyclers who lifelog their bicycling experiences can provide by way of their data important insights into transportation infrastructures, environmental issues, and the everyday experiences of cyclists in an

urban environment, one that privileges cars over pedestrians and cyclists. Yet, the team did not know whether or not these opportunities actually existed, if lifelogs of cycling experiences would inform these issues, or if cyclists even valued this perspective. The interview data indicates that there is, in fact, an opportunity for lifelogging videos to inform a community-centered archive that captures cycling life from the perspective of cyclists.

Lifelogging data is a form of documentation. Our interview data present that cyclists found not just personal, nostalgic, and entertainment value from their videos but also evidential (e.g., documenting drivers' behaviors) and informational value of the videos (e.g., capturing public infrastructure, safety information, and environmental changes). For instance, by recording the lived experiences of cyclists, the data serves to archive how individuals interact with the built environment and its social context in place and time. This documentation, then, serves as an important record worthy of analysis and reflection. As a few participants of this study acknowledged, stakeholders, such as city planners and government officials, can use the videos, audio files, and—potentially—GPS data to better understand how their transportation designs and policies positively and negatively affect the life of a cyclist. Lifelogs data empowers cyclists to tell their stories to community members who can see for themselves the difficulties of using non-traditional transportation (e.g. cars and buses) in a dense, urban environment where cyclists are sometimes met with threatening behavior simply for being on the road.

While documenting the lived experience, and archiving it in CHIME, is worthwhile for the sake of potentially improving the cycling infrastructure, there is also a cultural aspect. Several participants underscored the cultural value of their video by promoting cycling culture and educating novice cyclists. Participants argued the potential of the video data when the data are viewed or used by other citizens. Getting comments and seeing click data (e.g., likes) on lifelogs data published to social media sites excited and encouraged them to continue sharing their experiences. The work the City of Indianapolis has put into cycling paths and the bike-share program has helped seed a developing culture of cycling, and lifelogging data transferred into CHIME would further develop that community of individuals by connecting Indianapolis-area cyclists with a dedicated platform. By connecting individuals together on a centralized site and situating their content within a geographic community, it is arguable that CHIME holds the potential to further unify the local community of cyclists, as well as support their educational and informational needs by hosting, organizing, and creating further value from lifelogs data. Pulling the community together via CHIME also holds the potential to organize Indianapolis cyclists in a way that might enhance their ability to advocate for their cycling agenda in their neighborhoods and in city and state government.

There is an important science-based motivation behind CHIME, in addition to the plausible community and cultural benefits that might emerge. For cyclists, video and audio content recorded on lifelogging devices serves as a form of digital journaling. But for academics and policy makers, that content creates an analyzable dataset. Each video stream, for instance, can be analyzed via algorithms to identify potential harms to a cyclist. From poor pavement conditions on a bike path to a lack of street light at night, the videos can serve as datasets that motivate inquiry into ways to improve the cycling experience. And when aggregated with other sources of data, say crime statistics or traffic patterns, even more can be understood about cycling in an urban environment. Developing a large, analyzable dataset provides more actionable information for city stakeholders to consider the ways in which they can improve cycling life. And for researchers, such a unique dataset would provide data scientists, anthropologists, and others a new way of understanding a community's way of life.

#### *Privacy and the possibility for harm*

All plausible benefits aside, CHIME's potential for success depends on the degree to which it can account for privacy problems. The interview data suggests that lifeloggers have concerns about what is captured in their lifelogging files, especially with respect to how those files disclose private information about themselves and capture the private lives of others unintentionally. The themes from the data analysis point to a number of significant implications for CHIME.



First, while the participants expressed different levels of privacy concerns, concerns regarding private or personal information in the videos (e.g., revealing personally identified information) were common among most participants. CHIME needs to promote privacy-protecting tools and best practices for lifeloggers to use and follow in order to maximize their participation in the project whilst protecting their privacy—but more importantly by respecting their status as autonomous individuals. Privacy scholars often recognize that privacy is critical to having autonomy and expressing agency (e.g., Kupfer, 1987). If CHIME makes choices on behalf of its participants without their input, then its organizers and scientists limit how lifeloggers make choices in ways that allow them to self-govern according to their own values and wishes. While perhaps not legally obligated, CHIME is morally obligated to create a flexible infrastructure that can account for these important decisions. With that in mind, the CHIME system needs to maximize privacy choices from the very beginning: not as an afterthought. Such an approach will benefit from adopting privacy-by-design principles and frameworks, such as those developed by Schaar (2010) and Cavoukian (2012), when building features like “privacy dashboards” (Zimmerman, Accorsi, & Müller, 2014).

Second, the interview data shows that lifeloggers are aware of and concerned about the ways their videos capture and share sensitive information about their behaviors and environments, including the location of their homes and workplaces. As a result, they limited what they published on social media, and they expended effort to trim and edit videos to limit information disclosures. CHIME can use educational strategies to raise participants’ awareness of how their videos may be disclosing information that puts them at risk. These strategies should be paired with best practices for editing videos before uploading to CHIME for archiving and analysis. Educating users about how their data may be used in an age of Big Data and how they can limit potentially harmful disclosures of private information enhances their “privacy self-management” skills (Solove, 2013).

Third, as a few participants were aware of the privacy issues of bystanders even captured in public spaces, documenting public life via lifelogs can create even more tension around the idea of being private-in-public. Related research to the CHIME project suggested that, with regard to US law, publicly posted videos of public life do not intrude in private spheres of life or lead to appropriation concerns (Wolf et al., 2014); therefore, there seems to be no viable tort recourse available to individuals captured in lifelogging videos. Regardless of the law, the interview data suggests that lifeloggers have a norm of protection in their community and expect that data and information they disclose will primarily be used to the benefit of individuals who participate in this “context” (Nissenbaum, 2010). Cyclists in this study wanted to share their videos to inform their community of like-minded individuals, but at the same time they did not want to cause harm to those captured in their recordings by making the videos widely available and analyzable. More specifically, they were aware that their videos may compromise individuals, at the least, and possibly incriminate them, at the most. The difference between what is legal and what is socially expected among this community means that CHIME needs to be culturally aware of and sensitive to the protections lifeloggers put in place to protect the public, regardless of whether or not such protections may limit research practices and outcomes. Infringing normative expectations would violate the integrity of the context, thus resulting in information flows (e.g., disclosures) not approved by the lifelogging community.

Finally, there was a notable variation in the interview data regarding downstream use of lifelogging videos. While participants discussed how their video data could be used by cycling communities, governments, or advocacy groups, the majority of them did not express awareness of or concern about how these data can be accessed and used by other parties, such as businesses. Moreover, they did not express any understanding with respect to the ways that their videos could be aggregated with other data sources for analysis purposes, like CHIME proposes to do. This gap in the data may suggest that, while participants expressed privacy concerns, they are not fully aware of how their videos may be broken down into analyzable and sharable data to be aggregated with mixed-source datasets, either for research or to be traded and sold in the data market by data aggregators (Tene & Polonetsky, 2012). For CHIME, the ethical practice would be to raise participant awareness around these issues.

CHIME researchers have a moral obligation to consider the ways in which their well-intended work can cause serious harm for their participants. This article is a first step towards addressing the tricky privacy issues that they need to account for, but these are only patchwork solutions if the team does not adopt a data ethic in all that it does. Richards and King (2014) argue that society at large needs to develop “ethical principles and best practices” to govern data flows, and we argue that such principles and practices need to be codified and contextualized. Thus, CHIME—and other research teams like CHIME—need to discuss, debate, and set their own data ethics before system building and data gathering in order to make ethical practice the norm to account for and manage the harms before they emerge.

### Conclusion

The study described herein was motivated in part to better understand why cyclists share lifelogging videos, the ways such videos may be useful for a community-driven data archive, and the privacy concerns cyclists have with relationship to the way their videos capture and disclose information about themselves and the public. CHIME aims to build a platform for sharing and analyzing cycling data to improve cycling infrastructures and develop the Indianapolis-area cycling community. The findings from this study have helped CHIME researchers and system developers better understand how cyclists perceive the value of their videos to their community members and the public, as well as some of their privacy concerns. While the findings are specific to the development of CHIME, they also are relevant to other community archivists who want to include publicly accessible media on social media sites. Specifically, community archivists should take away from this study the important finding that privacy expectations are contextually bound and situated in a community. Thus, community archivists should carefully evaluate what informational norms govern information disclosures and privacy expectations. With that said, one limitation of this study is that it only studied the cyclists’ perspective. It is likely that other stakeholders involved in the CHIME archives project will have different perspectives on the value of including publicly accessible media and related privacy issues.

### References

- Al-Muhtadi, J., Ranganathan, A., Campbell, R., & Mickunas, M. D. (2003). Cerberus: A context-aware security scheme for smart spaces. *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, 489. Washington, DC: IEEE Computer Society. Retrieved from <http://dl.acm.org/citation.cfm?id=826025.826359>
- Altman, I. (1977). Privacy regulation: Culturally universal or culturally specific? *Journal of Social Issues*, 33(3), 66–84. <https://doi.org/10.1111/j.1540-4560.1977.tb01883.x>
- Bernard, R. (2000). *Social Research Methods: Qualitative and Quantitative Approach*. Sage Publication Ltd. Thousand Oaks, CA.
- Bowser, A., & Wiggins, A. (2015). Privacy in participatory research: Advancing policy to support human computation. *Human Computation*, 2(1), 19–44. <https://doi.org/10.15346/hc.v2i1.3>
- Cavoukian, A. (2014). Privacy by design. *IEEE Technology and Society Magazine*, 31(4), 18–19. doi: 10.1109/MTS.2012.2225459
- Chaudhari, J., Cheung, S. S., & Venkatesh, M. V. (2007). Privacy protection for life-log video. *2007 IEEE Workshop on Signal Processing Applications for Public Security and Forensics*, 1–5.
- Cheng, W. C., Golubchik, L., & Kay, D. G. (2004). Total recall: Are privacy changes inevitable? *Proceedings of the 1st ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, 86–92. New York, NY: ACM. <https://doi.org/10.1145/1026653.1026666>
- Cunningham, A. (1994). The archival management of personal records in electronic form: Some suggestions. *Archives and Manuscripts*, 22(1), 94–105.

- Denning, T., Dehlawi, Z., & Kohno, T. (2014). In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, 2377–2386. New York, NY: ACM. <https://doi.org/10.1145/2556288.2557352>
- Eagle, N., & Pentland, A. (2009). Eigenbehaviors: Identifying structure in routine. *Behavioral Ecology and Sociobiology*, 63(7), 1057–1066. <https://doi.org/10.1007/s00265-009-0739-0>
- Hoyle, R., Templeman, R., Armes, S., Anthony, D., Crandall, D., & Kapadia, A. (2014). Privacy behaviors of lifeloggers using wearable cameras. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 571–582. New York, NY,; ACM. <https://doi.org/10.1145/2632048.2632079>
- Indianapolis Cultural Trail. (2017). *About*. Retrieved from <http://indyculturaltrail.org/about/>
- Indianapolis Pacers Bikeshare. (2015). *About*. Retrieved from <https://www.pacersbikeshare.org/what-is>
- Indy Parks & Recreation & Dept. of Public Works. (2014). *Indy Greenways Full Circle Master Plan, 2013-2023*. Retrieved from <https://indygreenwaysmasterplan.wordpress.com/full-circle-master-plan-2/>
- Jana, S., Narayanan, A., & Shmatikov, V. (2013). A scanner darkly: Protecting user privacy from perceptual applications. *2013 IEEE Symposium on Security and Privacy*, 349–363. <https://doi.org/10.1109/SP.2013.31>
- Kärkkäinen, T., Vaitinen, T., & Väänänen-Vainio-Mattila, K. (2010). I don't mind being logged, but want to remain in control: A field study of mobile activity and context logging. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 163–172. New York, NY: ACM. <https://doi.org/10.1145/1753326.1753351>
- Kaye, J. "Jofish," Vertesi, J., Avery, S., Dafoe, A., David, S., Onaga, L., ... Pinch, T. (2006). To have and to hold: Exploring the personal archive. *Proceedings of the SIGCHI Conference on Human Factors in Computing System*, 275–284. New York, NY: ACM. <https://doi.org/http://doi.acm.org.libproxy.lib.unc.edu/10.1145/1124772.1124814>
- Kelly, P., Marshall, S. J., Badland, H., Kerr, J., Oliver, M., Doherty, A. R., & Foster, C. (2013). An ethical framework for automated, wearable cameras in health behavior research. *American Journal of Preventive Medicine*, 44(3), 314–319. <https://doi.org/10.1016/j.amepre.2012.11.006>
- Kupfer, J. (1987). Privacy, autonomy, and self-concept. *American Philosophical Quarterly*, 24(1), 81-89. Retrieved from <http://www.jstor.org/stable/20014176>
- Mann, S. (1997). Wearable computing: A first step toward personal imaging. *Computer*, 30(2), 25–32. <https://doi.org/10.1109/2.566147>
- McKemmish, S. (1996). Evidence of me . . . *Archives and Manuscripts*, 24(1), 28–45.
- Motti, V. G., & Caine, K. (2015). Users' privacy concerns about wearables. In M. Brenner, N. Christin, B. Johnson, & K. Rohloff (Eds.), *Financial Cryptography and Data Security* (pp. 231–244). Berlin: Springer-Verlag Berlin Heidelberg. [https://doi.org/10.1007/978-3-662-48051-9\\_17](https://doi.org/10.1007/978-3-662-48051-9_17)
- Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford Law Books.
- Pucher, J., Buehler, R., & Seinen, M. (2011). Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A*, 45, 451–475.
- Pucher, J., & Renne, J. (2003). Socioeconomics of urban travel: Evidence from the 2001 NHTS. *Transportation*

*Quarterly*, 57, 49–77.

Rawassizadeh, R. (2012). Towards sharing life-log information with society. *Behaviour & Information Technology*, 31(11), 1057–1067. <https://doi.org/10.1080/0144929X.2010.510208>

Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest Law Review*, 49, 393–432.

Savage, N. (2012). Gaining wisdom from crowds. *Communications of the ACM*, 55, 13. doi:10.1145/2093548.2093553

Schaar, P. (2010). Privacy by design. *Identity in the Information Society*, 3(2), 267-274. doi:10.1007/s12394-010-0055-x

Shilton, K., Burke, J. A., Estrin, D., Govindan, R., Hansen, M., Kang, J., & Mun, M. (2009). Designing the personal data stream: Enabling participatory privacy in mobile personal sensing. *Center for Embedded Network Sensing*. Retrieved from <http://escholarship.org/uc/item/4sn741ns>

Simmons, A. (2014, March 4). In Indianapolis, a bike path to progress. *The New York Times*. Retrieved from [https://www.nytimes.com/2014/03/09/travel/in-indianapolis-a-bike-path-to-progress.html?\\_r=0](https://www.nytimes.com/2014/03/09/travel/in-indianapolis-a-bike-path-to-progress.html?_r=0)

Solove, D. J. (2013). Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review*, 126, 1880–1903. Retrieved from <https://harvardlawreview.org/2013/05/introduction-privacy-self-management-and-the-consent-dilemma/>

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142 (10), 2282–2292. doi:10.1016/j.biocon.2009.05.006

Templeman, R., Korayem, M., Crandall, D., & Kapadia, A. (2014). PlaceAvoider: Steering first-person cameras away from sensitive spaces. *Proceedings from NDSS Symposium 2014*. San Diego, CA: NDSS. Retrieved from <http://www.internetsociety.org/doc/placeavoider-steering-first-person-cameras-away-sensitive-spaces>

Tene, O., & Polonetsky, J. (2012). Privacy in the age of big data: A time for big decisions. *Stanford Law Review Online*, 64. Retrieved from <http://stanfordlawreview.org/online/privacy-paradox-privacy-and-big-data/>

Toch, E., Cranshaw, J., Hankes-Drielsma, P., Springfield, J., Kelley, P. G., Cranor, L., ... Sadeh, N. (2010). Locaccino: A privacy-centric location sharing application. *Proceedings of the 12th ACM International Conference Adjunct Papers on Ubiquitous Computing - Adjunct*, 381–382. New York, NY: ACM. <https://doi.org/10.1145/1864431.1864446>

Tuohy, J. (2015, April 16). 1 year and 100,000 rides later, Indy bike-share program thrives. Retrieved from <http://www.indystar.com/story/news/2015/04/16/year-rides-later-indy-bike-share-program-thrives/25880751/>

Wang, Y., Leon, P. G., Scott, K., Chen, X., Acquisti, A., & Cranor, L. F. (2013). Privacy nudges for social media: An exploratory Facebook study. *Proceedings of the 22nd International Conference on World Wide Web*, 763–770. New York, NY: ACM. <https://doi.org/10.1145/2487788.2488038>

Whittaker, S., & Hirschberg, J. (2001). The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8, 150–170. <https://doi.org/10.1145/376929.376932>

Wolf, K., Schmidt, A., Bexheti, A., & Langheinrich, M. (2014). Lifelogging: You're wearing a camera? *IEEE Pervasive Computing*, 13(3), 8–12. <https://doi.org/10.1109/MPRV.2014.53>

Yoon, A. (2013). Defining what matters when preserving web-based personal digital collections: Listening to bloggers. *International Journal of Digital Curation*, 8(1), 173–192. <https://doi.org/10.2218/ijdc.v8i1.240>

Zimmermann, C., Accorsi, R., & Müller, G. (2014). *Privacy dashboards: Reconciling data-driven business models and privacy*. Paper presented at the Ninth International Conference on Availability, Reliability and Security (ARES), Fribourg, Switzerland. doi: 10.1109/ARES.2014.27