

# Big Data, Method and the Ethics of Location: A Case Study of a Hookup App for Men Who Have Sex with Men

Social Media + Society  
April-June 2018: 1–10  
© The Author(s) 2018  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/2056305118768299  
journals.sagepub.com/home/sms  


Ben Light<sup>1</sup> , Peta Mitchell<sup>2</sup>  and Patrik Wikström<sup>2</sup>

## Abstract

With the rise of geo-social media, location is emerging as a particularly sensitive data point for big data and digital media research. To explore this area, we reflect on our ethics for a study in which we analyze data generated via an app that facilitates public sex among men who have sex with men. The ethical sensitivities around location are further heightened in the context of research into such digital sexual cultures. Public sexual cultures involving men who have sex with men operate both in spaces “meant” for public sex (e.g., gay saunas and dark rooms) and spaces “not meant” for public sex (e.g., shopping centers and public toilets). The app in question facilitates this activity. We developed a web scraper that carefully collected selected data from the app and that data were then analyzed to help identify ethical issues. We used a mixture of content analysis using Python scripts, geovisualisation software and manual qualitative coding techniques. Our findings, which are methodological rather than theoretical in nature, center on the ethics associated with generating, processing, presenting, archiving and deleting big data in a context where harassment, imprisonment, physical harm and even death occur. We find a tension in normal standards of ethical conduct where humans are involved in research. We found that location came to the fore as a key—though not the only—actor requiring attention when considering ethics in a big data context.

## Keywords

big data, ethics, locative media, hookup apps, dating apps, gender and sexuality

## Big Data, Ethics and Location

The challenges that big data bring are epistemological, methodological and ethical. As danah boyd and Kate Crawford (2012) have argued, “big data reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorization of reality” (p. 665). Big data also brings with it ethical questions about the loss of human autonomy where new actors and tools are engaged to generate knowledge about our activity (Schroeder & Cowls, 2014). Andrej Zwitter (2014) further suggests rethinking the premises of modern ethics in the light of big data. When talking of group privacy, Zwitter states we need to avoid seeing anonymization as a form of harm reduction where large data sets are concerned. He concludes that the

anonymous in the sense of being de-individualized, groups are always becoming more transparent. (Zwitter, 2014, p. 4)

Where this occurs, Zwitter (2014) argues, the possibility increases to create incentives and disincentives for said groups, and this may happen with a lack of transparency of purpose. In effect, groups may be interfered with by unidentified others, who have unknown motivations, and those motivations may be of varying acceptability to the group in question. In this article, we are concerned with the potentials for such interference where location is a sensitive data point. Big locational data, we argue, can require a higher degree of

Anonymization of data is, thus, a matter of degree of how many and which group attributes remain in the data set. To strip data from all elements pertaining to any sort of group belongingness would mean to strip it from its content. In consequence, despite the data being

<sup>1</sup>University of Salford, UK

<sup>2</sup>Queensland University of Technology, Australia

### Corresponding Author:

Ben Light, School of Health and Society, University of Salford, Salford M6 6PU, UK.

Email: b.light@salford.ac.uk



ethical attention than big anonymous data regarding individuals and groups.

The past 15 years has seen the rise of the geoweb, the global positioning satellite (GPS)-enabled smartphone with its integrated location-based services, volunteered geographic information, and geo-social media—all of which are predicated on the sharing of personal or personally valuable geographic information. Noting that studies have estimated “up to 80% of Big Data is ‘spatial’,” Agnieszka Leszczynski and Jeremy Crampton (2016) have called for a more nuanced understanding of what they term “spatial big data” and the “anxieties of control” it engenders (pp. 1, 2). A key “anxiety” that attends to spatial big data is a surveillant one characterized by a fear over loss of privacy, with location being seen as a “uniquely sensitive” data point (Leszczynski, 2015, p. 966). According to a 2016 Pew Research Center report into attitudes toward privacy and information-sharing by American mobile phone users, location data is “especially precious in the age of the smartphone,” as it offers a “special intimacy” for the individual user (Rainie & Duggan, 2016, p. 5).

The question of geoprivacy in the era of spatial big data has therefore become of critical interest to researchers in Geography and in Geographic Information Science.<sup>1</sup> Moreover, the opening-up of mapping technologies via public-facing, “neogeographical” interfaces such as Google Maps since c. 2005 has raised new questions about the ethics of mapping when the question of “who’s doing the GIS” (Scull, Burnett, Dolfi, Goldfarb, & Baum, 2016, p. 26) has been broadened significantly beyond GIS professionals. Despite, or perhaps because of, these dramatic shifts in the nature and forms of geographic data and geovisualisation technologies, Scull et al. (2016, p. 25) argue that theoretical and ethical considerations in Geography, and particularly in Geographic Information Science have not kept pace with the rapid growth and evolution of geographic data. Although the study described here does not deal with particularly large volumes of geographic data (the dataset under consideration comprises roughly 12,000 sets of geographic coordinates), it addresses directly the question of the ethics of location—the ways in which location itself becomes ethically charged—and the ethics of mapping or geovisualising ethically charged location data.

Given the increasingly spatial nature of digital media, location is similarly emerging as a particularly sensitive data point for digital and social media research, particularly with regard to mobile or locative media (see, for example, de Souza e Silva & Frith, 2012; Frith, 2015; Mitchell & Highfield, 2017; Wilken & Goggin, 2015). Where much of the existing research focuses on the privacy implications of the location-sharing practices of individual mobile-phone users, in this study we are interested in ethical sensitivities that accrue around particular, digitally mediated sites of user activity. The ethical sensitivities around location are further heightened in the context of research into public sexual

cultures—that is, research that involves the study of sexual practices in places such as parks, shopping centers and sex clubs. In this article, we investigate ethical considerations involved in using digital methods to generate big data to analyze a web-based geo-social app for users contributing information about public sex and acting upon it.

## Public Sexual Cultures among Men Who Have Sex with Men

Public sex is contentious. It transposes what is generally considered to be a private activity onto a public space, it is often marked as illegal and morally unacceptable, and its study inevitably raises ethical concerns. Public sexual cultures among men who have sex with men (MSM)<sup>2</sup>—with or without the use of digital technology—can be considered as forming subaltern counterpublics (Fraser, 1990), or those parallel discursive arenas where members of subordinated groups circulate counter discourses and generate oppositional identities, interests, and needs. Given a significant proportion of MSM public sex culture is framed around *where* it is possible to have sex, location information is central to MSM discourse, both online and offline.<sup>3</sup> MSM cultures operate in a variety of physical spaces that are “meant” for public sex (such as gay saunas and dark rooms in bars) as well as spaces that are *not* ostensibly meant for public sex (such as parks, public toilets and truck rest stops).

Jamie Frankis & Flowers, (2005) differentiate between these two types of public sex space with the terms Public Sex Venues (or PSVs) and Public Sex Environments (or PSEs). Where PSVs are those spaces “meant” for public sex, PSEs are those spaces “not meant” for public sex. Information about PSVs and PSEs has always circulated within MSM counterpublics, but apps like the one we discuss in this article are able to foreground and frame discussion around location and provide access to MSM public sex location information in a way that pre-digital MSM culture would have been less able to.

MSM-based public sexual cultures have a significant modern history that predates the decriminalization of homosexuality in many countries. Indeed, landmark work, although methodologically controversial, was undertaken only 3 years after the decriminalization of homosexuality in the United Kingdom (Humphreys, 1970). The study of these cultures has remained significant, and this body of work points to their importance as a phenomenon that operates internationally, involves distinct practices, is subject to various legal statuses, and can involve vulnerable people (Frankis & Flowers, 2005, 2009). Public sexual cultures are often marked as illegal and morally unacceptable, not only by the general populace but also by MSM who align themselves with the need to mainstream their sexualities.

Public sexual cultures research usually deploys informal conversations, interview, and observational methods (Frankis & Flowers, 2005, 2009). In this article, we draw from a

project that adds to understandings of public sexual cultures among MSM as they are often enacted today, with digital media. To do this, we adopted digital methods whereby life is understood *with* digital media and not just through it (Rogers, 2013). Our work attends to the collection and analysis of post-demographic data (Rogers, 2013), such as user preferences and practices, in addition to anonymous demographic data such as age, ethnicity, and sexual orientation.

The case study is a web-based geo-social hookup app that facilitates public sex among MSM. We have chosen not to name the app because it is one of a number that serve a similar function, and we are primarily interested in the kinds of broader ethical questions raised by analyzing apps like this one rather than the specificity of the app itself. We developed a web scraper that allowed us to carefully collect selected data from the app and the data was analyzed using a mixture of content analysis using Python scripts, geovisualisation software and manual qualitative coding techniques. We draw on our experiences of conducting this research to elaborate on this ethically charged methodological challenge. Our findings, which are methodological in nature, center on generating, processing, presenting, archiving and deleting data. Overall, we find an interesting tension in normal standards of ethical conduct where human beings are involved in research. In the case we report here, we found that location rather than the individual, or group, becomes a key actor requiring attention when thinking about the potential for harm.

### Public Sex: Media, Methods and Ethics

The app provides a digital directory of public sex locations and contains thousands of entries covering a variety of locations across the world. These locations include parks, shopping centers, gyms, public toilets, saunas, and adult stores. Each entry is user-contributed and includes data such as where one might meet people at the location, the types of people who frequent such places, whether there is disabled access, and the optimum times to go. Directory entries also allow users to provide comments on locations (to post warnings about police activity, thank fellow users for good sex, and so forth). These comments are attached to pseudonymous profiles, which also contain information about individual users, such as their sexual preferences and safer-sex practices. Each entry is formally georeferenced, with a full address and coordinates. For each location, a link is provided to Google Maps allowing for easy navigation to the site on foot or by public or private transport. In the same way that TripAdvisor and Yelp are review and recommendation systems for travel and local business built on user-generated content and location-based services, the app studied here is a location- and user-generated-content-based recommendation system for public sex locations. A notable difference, however, is that, unlike TripAdvisor or Yelp, this app does not have or enable a global map view. That is, although a user of this app can open an individual public-sex location entry in

Google Maps and be presented with directions to it, that user cannot explore the full database (or even a subset of it) via a map-based interface to obtain a “god’s eye view” of the public sex locations in a given area.

In the following sections, we reflect upon our experiences of collecting and engaging with the data produced by the app and its users for the purposes of trying to generate understandings of public sexual cultures. In terms of generating data and processing data, we discuss the ethical considerations we faced when deciding whether to collect data, what data to collect and the institutional and commercial context of this. On presenting data, we refer to our decisions regarding the selective use of geovisualisation software. Finally, we discuss the challenges we faced in relation to the archiving and deletion of data and modes of data collection. We also point to some of the limitations of our work.

### Generating and Processing Data

To prepare for the process of generating data and processing, researchers carefully consider what they want the data for, and thus what their research questions are. In this case, our initial interest was rooted in prior work on apps used by MSM and the role these had in shaping the associated cultures (see, for example, Blackwell, Birnholtz, & Abbott, 2015; Brubaker, Ananny, & Crawford, 2016; Fletcher & Light, 2007; Light, 2007; Light, Fletcher, & Adam, 2008; Mowlabocus, 2008, 2010; Race, 2015a, 2015b). In respect of this app, our interests were in how it might be implicated in public sex among MSM. We also had a shared interest in digital methods, and we had a further question regarding how the data generated with and by the app might be helpfully used to understand these public sexual cultures. These two questions combined raised questions about what we *could* collect, and what we felt we *should* collect. At the very beginning of the process, we were aware that some form of harm could be generated by this research simply because we were interrogating a site of significant risk. The most obvious to us was that we could reveal the identity and practices of a person who wished, or needed, to keep these out of public view due to the legal and cultural acceptability of public sex and also the variable acceptability, of any given populace, of sex among men.

As part of this process of deciding what data to collect—and how—we encountered numerous ethical questions. Underlying this was the question of whether to collect the data in the first place. The Association of Internet Researchers Ethics Guidelines offer a significant range of ways to interrogate this situation. We considered the following four questions posed in Appendix 1 to be the most significant.

1. How do terms of service (TOS) articulate privacy of content and/or how it is shared with third parties?
2. Regardless of TOS, what are community or individual norms and/or expectations for privacy?

3. Does the author/subject consider personal network of connections sensitive information?
4. Is the data easily searchable and retrievable? If the content of a subject's communication were to become known beyond the confines of the venue being studied—would harm likely result? (Markham & Buchanan, 2012, p. 18)

Thinking through these questions, we formulated the following ethical position. The app does not provide an application programming interface (API), and the terms and conditions state that data generated by the use of the app cannot be used for commercial purposes. We discussed the TOS with legal scholars and explained to them the sensitive nature of the app. We also had further discussions with other scholars with knowledge of risky research, gender, and sexuality research, and Internet research. We reached the conclusion that collecting selected data for analytical purposes and disposing of the collection method and collected data once this process was complete was appropriate. Because we were being highly selective about the data we were collecting and because of its nature, we felt the benefits of the overall findings balanced against the challenges of using said data without the permission of the app owner or the users even though neither would reasonably expect us to do this. We did not ask the app owners for permission as we feared they may not allow us access, primarily due to their commercial interests in the development. Our decision to go ahead is set within the increasing commercialization of our everyday activities where the digital is concerned. We take the position that how our lives are increasingly being regulated and structured by large platform and app providers is becoming ever more opaque, and that it is our role as academics to seek to make such activity clear (without causing harm). We are aware other researchers may disagree with our position and approach, but it highlights an area in big data ethics that remains contested. In their empirical study of researchers doing online data research, Jessica Vitak, Katie Shilton, and Zahra Ashktorab (2016) found that “the ethics of ignoring Terms of Service” was a critical area “of significant disagreement in the online data research community” and one that “consensus-building efforts” should focus upon (p. 951).

The app in question offers significant amounts of data about MSM-based public sexual cultures and those who engage in them. It does this because it offers a range of ways that facilitate communication within the app such as chat rooms, web camming, messaging, and user-to-user chat as well as user profiles and the digital directory. Although we could have collected a great deal more data generated by users within the app's “public” spaces, we decided that the data we would collect would be highly selective, working from the position that just because we could, it did not mean that we should. Table 1 provides examples details of the data we could have collected, and the data we actually collected for the purposes of this study. This data relate to the profiles

of users who had made comments upon directory entries and the directory entries themselves. To provide a layer of protection for users of the app, we only provide a selection of both the data we collected and the data points we did not collect. This is because if we provided the full list of data points it may be possible to identify the app in question. The data points that we provide appear on multiple apps in this area, and this provides a level of cover.

There are many ways to extract structured data from the web. The use of application programming interfaces (API) is becoming increasingly common in proprietary social media. However, not all providers offer this for a number of reasons, including a desire to prohibit the extraction of data on a large scale, the thought that no-one would want that data or a lack of expertise to offer an API. In the case of the site of our study, no reason is publicly given by the developer for the absence of an API even though these often enable commerce within such spaces, and even though this app has commercial interests.

As there was no API access, we used the Python open-source programming language to collect selected data into a comma-separated values (CSV) file. We structured the data collection into four steps. In the first step, we collected basic location data. At this stage only location ID and location name were collected to set up a basic structure that could be used to scrape data about users and comments. This scraping returned just under 12,000 public sex locations across several countries and continents. In the second step, we collected user comments linked to each location. These data consisted of, for example, the comment, a timestamp for when it was posted and which user id that did the posting. This yielded approximately 736,000 comments overall, or an average of 61 comments per location. In the third step of the data collection process, we collected additional data about each location (see examples in Table 1). Finally, in the fourth step, we collected user data for all users who had posted comments scraped in previous steps (see examples in Table 1). This yielded profile information about approximately 120,000 users. There was one “super user” who had contributed over 3,800 comments on locations in our dataset, but the average number of comments per user was around 6.

The process of scraping generated an anonymous data set, assigning a numerical ID in place of the app users' pseudonymous usernames. As noted above, we chose not to collect these usernames, but their already pseudonymous nature means they could arguably be considered to preemptively allow for anonymity, so we should not have been overly concerned with anonymizing these. Furthermore, we searched the open web for pseudonyms, and any comments made by users within the app, and these were not returned in search engine results, providing further evidence that the app already guaranteed a level of anonymity through its pseudonyms and the way that it currently operated as a kind of “walled garden.” However, we were concerned not only with what might happen if someone searched at the same time we



**Table 1.** Examples of Data Collection Choices.

Profile data available	Collected	Rationale
Username	No	Unnecessary information for our research.
City	Yes	Helpful in understanding far someone may travel to a public sex site.
Sexual identity	Yes	Helpful in understanding the demographics of those engaging in public sex.
Cock size	No	Unnecessary information for our research.
Age	Yes	Helpful in understanding the demographics of those engaging in public sex.
Height	No	Unnecessary information for our research.
Ethnicity	Yes	Helpful in understanding the demographics of those engaging in public sex.
Level of engagement with safer sex	Yes	Helpful in understanding stated practices associated with public sex.
Directory data available	Collected	
GPS data about site	Yes	Helpful in understanding location and travel associated with public sex
Type of site (e.g., washroom, park, mall)	Yes	Adds context to location data and comments made on directory pages.
Site address	No	Unnecessary for our research and potentially more harmful data than GPS in its descriptiveness.
Directions to site	No	Unnecessary for our research and potentially more harmful data than GPS in its descriptiveness.
Contact info for listing	No	Unnecessary for our research and potentially harmful data.
Site comment	Yes	Provides insights into public sexual cultures.
User name of commenter	No	Unnecessary for our research and potentially harmful data.

did, we were cognizant of the potential for data to become available in the future. Moreover, we were aware of the aggregation effect whereby, as Kate Crawford and Megan Finn (2014) state, “multiple data feeds [can be] combined which can generate intimate insights without the person’s knowledge” (p. 498). In addition, we noted some pseudonyms contained what appeared to be a user’s given name and, less frequently, family name. Knowing the real or pseudonymous names of the users did not add to our research, as interesting as the nature of pseudonyms might be.<sup>4</sup> For this reason, we decided not to scrape the users’ pseudonyms but rather use an individual, numerical identifier as the means to tell the users apart. Because individual user comments were collected under this numerical identifier, it would be possible for us to re-pseudonymise the dataset, but this would take a significant amount of work to rematch our anonymized data set with the one that was live within the app. Moreover, during our study we noted that older directory listing comments were regularly deleted from public view, to keep them at around 100 comments per entry. This added cover for users as our CSV file would automatically become out of synch with the live app.

In contrast to the collection of usernames, we decided that the very specific GPS coordinates provided in the directory listings were helpful to collect, even though they were risky data. They were risky data in that, assuming they are correct,<sup>5</sup> they provide a very accurate indication of where public sexual activity may take place in the physical world. However, these data, we felt, was central to our analysis given the importance of location in public sexual cultures. However, once we began our analysis, it became apparent that the collection of location data posed its own particular threat to the anonymity of the data set.

As researchers, we raised our concerns with the research ethics committee of the overseeing institution, querying whether we required ethical clearance to conduct this research in accordance with Australia’s National Statement on Ethical Conduct in Human Research (2015).<sup>6</sup> The Statement defines human research as “research conducted with or about people, or their data or tissue” (p. 3). Although the Statement focuses largely on the corporeal—on the ways in which research requires embodied participation or acts upon bodies (e.g., through participants physically taking part in interviews, being physically observed by researchers, or by undergoing medical testing or having their bodily tissues or fluids collected)—it also takes into account access to human data, including personal documents or information (whether identifiable, re-identifiable, or non-identifiable) stored within published or unpublished databases (p. 7). Under the Statement, the requirements for ethical review are framed around the key “themes” of risk (vs. benefit) and consent. Specifically, the Statement distinguishes between low- and negligible-risk human research, where “low-risk” research is defined as “research in which the only foreseeable risk is one of discomfort” and “negligible-risk” research is defined as “research in which there is no foreseeable risk of harm or discomfort; and any foreseeable risk is no more than inconvenience” (p. 13).

In this specific case, because all of the app’s user profiles are pseudonymous and because the process of scraping anonymized individual usernames, it was deemed that our data collection process was of negligible risk to the humans behind the pseudonyms. As all is needed to access the information we collected is to pseudonymously sign up to the app, it was deemed to be public information, and therefore, we did not have to apply for ethical clearance for what was

conceived of as “negligible risk” research. The Statement makes provision for this kind of exemption, explaining that research that “uses collections of non-identifiable data and involves negligible risk [. . .] may therefore be exempted from ethical review” (p. 42). Yet, it does not take much to realize how ethically sensitive location data, in particular, is within this dataset. Although the scraping process anonymised the already pseudonymous app users, we were able to collect specific location data (coordinates) for public sex locations across a variety of countries—sites which, as we have explained from the outset, are contentious ones. In addition to this formally georeferenced data for public sex locations, we also collected informally georeferenced data (i.e., place names) relating to where a user said they were based in their user profile. This informally georeferenced information ranged from the generic (e.g., country or state level) to the more specific (e.g., suburb) depending on how far the user drilled down in providing their location on their profile. Had we chosen to (which we did not), we could quite easily have mapped these users in time and space against public sex locations, and even have mapped out the itineraries over time of individual, anonymised users of the app. It is this process of the analysis of this mutable data set where anonymity, privacy and publicity are concerned which we turn to next.

## Analyzing and Presenting Data

We analyzed the data in two ways. First we analyzed comments made about the public sex locations and categorized these thematically. This gave us, for instance, categories such as users: specifying times they were available, telling others they were not far away from a site, asking about other venues close by, providing warnings about police or security presence in the area, complaining about other users not turning up for a pre-arrange to meet, and commenting on the extent to which a site was busy or quiet. We were able to combine these data with other data points, such as those regarding the types of location, and this allowed us insights into the extent to which comment types were specific to certain physical and geographical environments. For example, the presence of dog walkers was mentioned in relation to parks and security officers were noted in regard to in shopping centers. This part of the research affords insights into certain elements of public sexual cultures among MSM on quite a large scale, and because of this scale we are able to talk about commenting at a very abstract level. Through this process, we provide cover for individuals in that we do not need to quote particular comments to make our point. For example, in one subset of the data (containing 30,000 comments related to a specific geographical area), we were able to establish that the most prominent commenting activity was concerned with users enquiring if anyone was available. Using Python to look for the presence of popular words, we were able to establish the top 5 as: *anyone* (7407), *here*

(7032), *around* (3690), *now* (3681), and *today* (1581), with the average length of a comment being 39.2 characters. This first part of the analysis did not seem to present any very specific ethical challenges. In fact, the big data angle assisted with dealing with a sensitive and under-researched topic (at large scale at least) and fed into our desires to provide insights that would be helpful to those working with minority communities.

The second mode of analysis involved us using the Carto platform and generated significant ethical considerations. When we first used Carto in 2016, it was named CartoDB and was marketed as geovisualisation software with the tagline “CartoDB is the Easiest Way to Map and Analyze Your Location Data.”<sup>7</sup> In 2017, the software has lost “DB” from its name and the opening page of its website announces “PREDICT THROUGH LOCATION” and that “CARTO is an open, powerful, and intuitive platform for discovering and predicting the key insights underlying the location data in our world.”<sup>8</sup> Our initial concerns about using a tool to map the locations of public sex sites and then overlaying this with contextual data (such as ethnicity of commenters and times of comments made) is perhaps perfectly illustrated in this shift in the characterization of the purpose of platform between sometime in 2016 and 2017—from analysis to prediction. Carto exists as a freemium, web-based mapping service. We used research funds available to us to pay for an upgraded account because, at the time we undertook the research, the entry-level free account required that any dataset uploaded to the platform be made public. Although the resulting map-based visualization could be set as private, the data underlying could not. We paid in the region of AUD\$1500 for a one-year subscription that allowed us to store data privately. This situation adds weight to the problem that has been raised in academia many times already, particularly in the arts, humanities and social sciences, that only those who can afford to access certain tools will be able to undertake certain kinds of research—a further ethical consideration in big data research.

We organized the scraped data into one CSV file and uploaded it to Carto. This allowed us to geo-visualize the formally georeferenced information (coordinates) in our dataset and to create annotated maps which detailed the locations where MSM purportedly had public sex based on the existence of the directory entry and the comments attached to it. We then had the capacity to overlay these maps with the data collected from user profiles, the data collected about the locations in the directory, and the metadata about the comments (e.g., the times and dates the comments were made). Even taking into account the limitations of our approach and the data (as we discuss in the next section) it is not hard to see how through our attempts at seeking to analyze a situation, we could inadvertently help people to predict and act based on these data.

We have not spoken with the developers of this app about whether the app’s lack of a map interface was or was not a

conscious decision on their part, and so we do not know if there were any specific reasons behind this—for example in relation to user safety. We do, however, know that some other online public sex directories offer this facility and mode of navigation. In any case, it is clear that mapping or geovisualising data in this way can show a very different picture compared with navigating through a database for individual locations. It presents a god's-eye perspective that the user cannot get at the individual database entry level or even in a full tabular dataset and shows the spatial relationships between and among points. Moreover, as Peta Mitchell and Tim Highfield (2017) explain, “when personal location information is visualised on a map—when it is grounded and made apparent in ways that the user might not have anticipated, it puts [. . .] privacy questions into even starker relief,” and they stress that “geovisualisation itself can have unintended or even perverse consequences.”

Using a platform like Carto, it would have been relatively simple for us to exploit its geovisualisation capabilities to create and present a range of annotated static and animated maps in publications and presentations. These maps might have revealed interesting spatial patterns marking the extent and geographic spread of MSM public sexual activity collected within this app. However, disrupting the contextual integrity (Nissenbaum, 2004) of the data in this way also brings greater risk to the community contributing the data, particularly when these maps might draw unwanted attention to the community, its socially contentious practices, and its geographic reach—especially in regions where sex between men remains illegal. We were, therefore, highly attentive to the politics of visualization that arise from research such as this (Kennedy, Hill, Aiello, & Allen, 2016), and the particular sensitivities that attend to geovisualisation (Elwood & Leszczynski, 2011). In this case, a politics of visualization goes beyond the individual and toward the problem of visualizing location. In doing this, it holds the potential to reintroduce a potential risk of harm for individuals. Big data does not provide cover in this sense; rather, it amplifies the potential for harm as the possibility for predicting activity on the ground where MSM are having public sex comes into being. This situation includes those men who may never have used the platform or even know of its existence. To make this very clear, such predictive capability holds the potential for the harassment and criminalization of those who may be vulnerable and isolated, such as those in rural communities, those not out concerning their sexual preferences, those who define as bi-sexual and straight, those forced into marriage, or bound by a particular religion.

Ethical clearance for human research understandably focuses on risks to the human individual, and the anonymous and public nature of the data of the hookup app analyzed here posed few difficulties for us in terms of gaining institutional clearance. And yet in contexts such as this, we argue, the public sex locations (which must be geolocatable for the purposes of the app) call for increased ethical attention. Following his

essay on “Deconstructing the Map,” map historian John Brian Harley (1991) turned his attention to the ethics of cartography, noting that “Cartography seems to be uncritical of its own practices, and both their intentional and unintentional consequences. It certainly lacks,” he said, “a substantial literature in applied ethics comparable to that generated by many of its peer professions in science and technology” (p. 9).

Harley's call for an ethics of cartography seems even more important in an era where even those researchers with no training in the principles of cartography or GIS, but who have access to neogeographical tools like Google Maps or Carto, can make maps on the fly with geodata they have scraped from mainstream social media or the more specialist apps like the one discussed in this study. In pointing toward an ethics of cartography, Harley (1991) asks map-makers to consider a number of questions, including “What are the motives and personal engagements of cartographers with the maps they make? What are the relationships between production and consumption in cartography and GIS? [. . .] What are the moral benefits or deficits of particular ways of mapping the world?” and how do what is included or excluded, privileged or elided by the map “actually influence the way people think about and act upon social issues in a democracy?” (p. 14).

Taking into account these critical questions, in regard to our study, we argue it is more ethical *not to map* in the sense of presenting or publishing geovisualisations of the data we have collected. Nonetheless, location is a key part of public sexual cultures among men who have sex with men, and so there are arguments that could be made for the research benefits of mapping and contextualizing user-contributed public sex locations with and through digital methods. Most obviously, one might think of the public health benefits of knowing when and where people are having sex and who “say” they are going to do it unsafely. Outreach workers could operate efficient services, springing from behind trees with a bag of lubricant and condoms, just at the right moment, and all will be well will it not? Of course not. This gave us pause for thought in terms of how such locative data might be put to better use. Outreach workers, and the organizations they are part of, usually know where cruising and cottaging activity takes place within the areas they serve. If they do not, they can easily source that information as they engage with MSM and they can look up directories such as the one we consider here. So in this way location has less value.<sup>9</sup> The value in this data set is that it affords learning, at scale, about the topics of discussion, activities, demographics and tastes associated with different kinds of location where public sex occurs. Included in these kinds of location we mean, for instance, the extent to which a location is rural or urban, a park or a toilet, a beach or a shopping center, a sauna/bathroom or an adult cinema/shop. By focusing upon themes such as these, rather than coordinates and actual places, it is possible, we believe to use big data to help vulnerable groups in a way that distances them from harm.

## Archiving and Deleting Data

The final set of considerations we wish to approach here concerns the question of holding and retaining a highly sensitive archive of data. As we have already discussed, the CSV file we created as a result of the scraping is not easily matched with live data as far as individuals are concerned. However, it stands as a sensitive record of discussion about a diverse range of locations. Although locations for public sex may be mutable in their existence and popularity, many have existed for decades and may continue to do so. Therefore, even though users, discussions, and even directories on the web/apps may change, these data could cause harm in the future, and even more so given we cannot anticipate any potential aggregation effects. We will therefore delete our data and delete our python scripts.

We will only hold the CSV data file for as long as we need to complete our work. At this point, the file will be deleted. This is in opposition to governing institutional data management policies, which require the long-term archiving of research data. It is also misaligned with the move toward open-access data sharing in academia, particularly where research is conducted using public funds. Deleting our research data also means, of course, that our findings cannot be checked by others, which may matter to varying degrees in different disciplines. This research is not part of a PhD student's program of study where an examiner or supervisor may ask to see data, but if it was, this may of course then lead to further issues.

We will delete the Python scripts that have been used to collect and interrogate the data set, which makes it harder for another researcher to acquire (or for us to reacquire) this data. Of course, another researcher or member of the public could write their own scripts or use proprietary software to collect the data in another way (they may even have done this already). On the sharing of scripts, there is another issue to raise here. Due to the nature of our research and our institutional ethics board requirements, we were granted an exemption from the ethical approval process. Had we gone through this process one might expect that we would provide details of our research instruments. For example, it is not uncommon to provide interview protocols or questionnaire designs. Here, an interesting question arises around whether or not programming scripts should be submitted for scrutiny in the ethics process as they become a common instrument of data collection, and/or analysis. That said, in the same way as a data set might be better kept by only a few, and then deleted, the same may be said of programming scripts. We do not have the answer here but merely report on the actions we have taken and provided reasons why.

## Limitations

Like other platforms, the app we are studying is not representative of a group—in this case, all those men who have sex

with men—or even representative of a subset of a group—those MSM who have sex in public places. There are many reasons for this. Certain people from certain countries use this app. Some users may also have multiple accounts and multiple profiles, and may therefore be represented multiple times within the app. There are also, due to the nature of the app, questions regarding the accuracy of the way that profiles are populated. People may adjust their sexual identity or location, for example, to provide cover for themselves, or they may signal they engage only in safe sex when they may, under certain circumstances, change their minds at a meet up. Moreover, the data we have speaks only to those who have commented upon a directory entry. We did not collect the profile data of all users—only those who had made a comment in the directory. The methods we have deployed also tell us little about those users of the app who have browsed the directory (and used information such as how to get there and to work out a good time to go). This “lurking” and “listening” (Crawford, 2009) has been shown in various studies to be an extensive practice. However, overall, we see this data set as affording a plausible, and helpful, starting point for conversations about the nature of public sexual cultures among MSM where large-scale studies have not previously been possible.

## Conclusion

Methods that afford the collection and analysis of big data bring with them political and ethical questions. However, while big data and the methods used to generate and analyze them may be aligned with commercial interests (Frade, 2016), as shown by our work, other politics and ethics may simultaneously be at play. We agree with Carlos Frade (2016) that any engagement with big data and associated methods must include strong elements of critique and reflection. We also argue that while big data may reflect the status quo of society, at the same time it offers the potential to reveal resistance and alternative lives—critically, at scale. The need, from an ethical perspective, is to challenge dominant discourses politics in society without causing harm. That is not to say that we should shy away from the difficult questions and topics. In fact, what we hope we have shown is that big data can sometimes provide safe cover if treated with care as it is collected, analyzed, presented, and stored (or not).

As we have argued, the ethics of location and geovisualisation in the context of the rise of big data and the digital geo-social requires further investigation. With the growing “pervasiveness of location within and across all mobile networking apps and platforms” (Mitchell & Highfield, 2017), the ethics of engaging with, researching, and re-presenting user-contributed spatial data in mapped form will become increasingly salient. Before disposing of the dataset, as we have indicated above, we are interested in delving further to see what the data might tell us about, for instance,



- The spatiotemporal growth and spread of public sex locations as reported in and through the app
- The differences or similarities in PSE practices and discourses between urban, periurban, and rural public sex locations
- The kinds PSE user mobility we might see by linking public sex locations with user-profile locations (something we have identified as raising heightened ethical concerns)
- The relationship between safer sex practices and location type and what this might reveal about sites of and for safer sex.

We believe that doing ethically attuned work with this data to explore the role of location within MSM cultures, but in a way that is sensitive to the politics of location, may contribute to a growing and ever-more critical understanding of the ethics of location and geovisualisation within digital methods and data cultures.

### Acknowledgements

Our thinking on our ethics around this work has been very much improved through presentation and discussion at the 2015 AoIR Conference—Phoenix, USA; the 2015 Data Power Conference—University of Sheffield—UK; the 2016 CCI Digital Methods Summer School—QUT—Australia; and the 2016 Crossroads in Cultural Studies Conference, University of Sydney, Australia. In particular, we also thank Dr Kylie Jarrett (Maynooth), Dr Sharif Mowlabocus (University of Sussex), Prof Susanna Paasonen (University of Turku) and Assoc. Prof. Nic Suzor (QUT) for their helpful advice and discussions at various points as we have developed this work. Finally, we are grateful for the valuable feedback obtained during the review process. Naturally, any errors and the final decisions on ethics are ours.

### Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iDs

Ben Light  <https://orcid.org/0000-0002-0038-8979>

Peta Mitchell  <https://orcid.org/0000-0003-4523-6685>

### Notes

1. See, for instance, Elwood and Leszczynski (2011) and Miller and Goodchild (2015).
2. The term “men who have sex with men” is used to allow for the inclusion of a diversity of sexualities of men who engage in public sex with men, such as those who identify as bisexual and heterosexual.
3. Another key aspect of MSM public sexual cultures is, of course, time, especially when combined with location. The

availability of a location for public sex is often related to the time of day. This raises a further ethics of time associated with public sexual cultures and with big data, but it is outside the scope of this article.

4. For example, Anna Livia (2002) notes, numbers (between 10 and 25—referring to centimeters) and/or the abbreviations BFT and TTBM are used in complex codes of expressive pseudonyms to denote penis size in MSM hook up groups that operated on the now defunct French Minitel system.
5. Given that the coordinates are volunteered geographical information and manually entered by users of the app, the data show occasional errors, due, for example, to inversion of latitude and longitude.
6. The primary research was undertaken while Light was employed at QUT and written up when he moved to the University of Salford. The University of Salford’s ethics process was not engaged for the findings of this article, but will be as we move forward with our analysis.
7. <https://web.archive.org/web/20160610170347/https://cartodb.com/>
8. See [www.cartodb.com](http://www.cartodb.com)
9. We acknowledge the different legal statues of homosexuality and public sex across different geographical locations.

### References

- Blackwell, C., Birnholtz, J., & Abbott, C. (2015). Seeing and being seen: Co-situation and impression formation using Grindr, a location-aware gay dating app. *New Media & Society, 17*, 1117–1136.
- boyd d & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*, 662–679.
- Brubaker, J. R., Ananny, M., & Crawford, K. (2016). Departing glances: A sociotechnical account of “leaving” Grindr. *New Media & Society, 18*, 373–390.
- Crawford, K. (2009). Following you: Disciplines of listening in social media. *Continuum, 23*, 525–535.
- Crawford, K., & Finn, M. (2014). The limits of crisis data: Analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal, 80*, 491–502.
- de Souza e Silva, A., & Frith, J. (2012). *Mobile interfaces in public spaces: Locational privacy, control, and urban sociability*. Abingdon, UK: Routledge.
- Elwood, S., & Leszczynski, A. (2011). Privacy, reconsidered: New representations, data practices, and the geoweb. *Geoforum, 42*, 6–15.
- Fletcher, G., & Light, B. (2007). Going offline: An exploratory cultural artifact analysis of an Internet dating site’s development trajectories. *International Journal of Information Management, 27*, 422–431.
- Frade, C. (2016). Social theory and the politics of big data and method. *Sociology, 50*, 863–877.
- Frankis, J. S., & Flowers, P. (2005). Men who have sex with men (MSM) in public sex environments (PSEs): A systematic review of quantitative literature. *AIDS Care, 17*, 273–288.
- Frankis, J. S., & Flowers, P. (2009). Public sexual cultures: A systematic review of qualitative research investigating men’s sexual behaviors with men in public spaces. *Journal of Homosexuality, 56*, 861–893.

- Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text* (25-26), 56–80. Retrieved from [https://www.jstor.org/stable/466240?seq=1#page\\_scan\\_tab\\_contents](https://www.jstor.org/stable/466240?seq=1#page_scan_tab_contents)
- Frith, J. (2015). *Smartphones as locative media*. Cambridge, UK: Polity Press.
- Harley, J. B. (1991). Can there be a cartographic ethics? *Cartographic Perspectives*, 10, 9–16.
- Humphreys, L. (1970). *Tearoom trade: Impersonal sex in public places*. London, England: Duckworth.
- Kennedy, H., Hill, R. L., Aiello, G., & Allen, W. (2016). The work that visualisation conventions do. *Information, Communication & Society*, 19, 715–735.
- Leszczynski, A. (2015). Spatial big data and anxieties of control. *Environment and Planning D: Society and Space*, 33, 965–984.
- Leszczynski, A., & Crampton, J. (2016). Introduction: Spatial big data and everyday life. *Big Data & Society*, 3, 1–6.
- Light, B. (2007). Introducing masculinity studies to information systems research: The case of Gaydar. *European Journal of Information Systems*, 16, 658–665.
- Light, B., Fletcher, G., & Adam, A. (2008). Gay men, Gaydar and the commodification of difference. *Information Technology & People*, 21, 300–314.
- Livia, A. (2002). Public and clandestine: Gay men's pseudonyms on the French Minitel. *Sexualities*, 5, 201–217.
- Markham, A., & Buchanan, E. (2012). *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee (Version 2.0)*. Retrieved from <https://aoir.org/reports/ethics2.pdf>
- Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, 80, 449–461.
- Mitchell, P., & Highfield, T. (2017). Mediated geographies of everyday life—Navigating the ambient, augmented and algorithmic geographies of geomedial. *Ctrl-Z: New Media Philosophy*, 7. Retrieved from <http://www.ctrl-z.net.au/journal/?slug=mitchell-highfield-mediated-geographies-of-everyday-life>
- Mowlabocus, S. (2008). Revisiting old haunts through new technologies: Public (homo)sexual cultures in cyberspace. *International Journal of Cultural Studies*, 11, 419–439.
- Mowlabocus, S. (2010). *Gaydar culture: Gay men, technology and embodiment in the digital age*. Farnham, UK: Ashgate.
- National Statement on Ethical Conduct in Human Research. (2015). *National health and medical research council, Australian research council, and Australian Vice-Chancellors' committee*. Retrieved from [https://www.nhmrc.gov.au/\\_files\\_nhmrc/publications/attachments/e72\\_national\\_statement\\_may\\_2015\\_150514\\_a.pdf](https://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/e72_national_statement_may_2015_150514_a.pdf)
- Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, 79, 101–139.
- Race, K. (2015a). 'Party and play': Online hook-up devices and the emergence of PNP practices among gay men. *Sexualities*, 18, 253–275.
- Race, K. (2015b). Speculative pragmatism and intimate arrangements: Online hook-up devices in gay life. *Culture, Health & Sexuality*, 17, 496–511.
- Rainie, L., & Duggan, M. (2016). Privacy and information sharing. *Pew Research Center*. Retrieved from [http://www.pewinternet.org/files/2016/01/PI\\_2016.01.14\\_Privacy-and-Info-Sharing\\_FINAL.pdf](http://www.pewinternet.org/files/2016/01/PI_2016.01.14_Privacy-and-Info-Sharing_FINAL.pdf)
- Rogers, R. (2013). *Digital methods*. Cambridge: The MIT Press.
- Schroeder, R., & Cowls, J. (2014, August 24). Big data, ethics, and the social implications of knowledge production. Paper presented at data ethics workshop, New York, NY. Retrieved from <https://dataethics.github.io/proceedings/BigDataEthicsandtheSocialImplicationsofKnowledgeProduction.pdf>
- Scull, P., Burnett, A., Dolfi, E., Goldfarb, A., & Baum, P. (2016). Privacy and ethics in undergraduate GIS curricula. *Journal of Geography*, 115, 24–34.
- Vitak, J., Shilton, K., & Ashktorab, Z. (2016). Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. In *Proceedings of the 19th ACM conference on computer-supported cooperative work and social computing* (pp. 941–953). New York, NY: ACM.
- Wilken, R., & Goggin, G. (Eds.) (2015). *Locative media*. New York, NY: Routledge.
- Zwitter, A. (2014). Big data ethics. *Big Data & Society*, 1, 1–6.

### Author biographies

**Ben Light** is professor of Digital Society in the School of Health and Society at the University of Salford. His research is concerned with understanding people's everyday experiences of digital media. Light engages science and technology studies bringing it into dialogue with questions of (non)consumption practices, digital methods, gender, and sexuality. He is currently working in the areas of digitally mediated public sexual cultures and dating and hookup apps. He is author of *Disconnecting with Social Networking Sites* (Palgrave, 2014), and his latest book with Kylie Jarrett (Maynooth) and Susanna Paasonen (Turku), exploring the phenomenon of Not Safe For Work (NSFW), is due to be published in 2018 (MIT Press). His research has been published in journals such as *Convergence*, *First Monday*, *Information, Communication and Society* and *New Media and Society*.

**Peta Mitchell** is an associate professor and Chief Investigator in QUT's Digital Media Research Centre. Her research focuses on digital geographies, location awareness and mobile media, algorithmic culture, and network contagion. Mitchell is author of *Cartographic Strategies of Postmodernity* (Routledge, 2008) and *Contagious Metaphor* (Continuum, 2012) and has published numerous articles, chapters, and refereed conference papers that span media and cultural studies, cultural geography, and human-computer interaction. She is also co-founder of the Cultural Atlas of Australia, an ARC-funded digital mapping project that explores Australian locations as they are represented in and through films, novels, and plays, and co-author of the related book *Imagined Landscapes: Geovisualizing Australian Spatial Narratives* (Indiana UP, 2016).

**Patrik Wikström** is professor and Head of QUT's School of Communication in the Creative Industries Faculty. He uses and develops computational methods to examine the dynamics and regulatory challenges of networked markets, with a particular focus on markets for cultural products. Wikström is the author of *The Music Industry: Music in the Cloud* (Polity, 2009) and has published his research in journals such as *Technovation*, *International Journal of Media Management*, *Journal of Media Business Studies*, *Journal of Music Business Studies* and *Popular Music & Society*.