The Use and Misuse of Geolocation Mobility Data:

A Literature Analysis and recommendations for ethical research

February 2025

nicholson CONSULTING

Table of Contents

Authorship

The Literature Analysis was largely conducted and authored in early 2024 by Pascarn R. Dickinson (Senior Analyst, Nicholson Consulting; Adjunct Research Fellow, Te Herenga Waka Victoria University of Wellington, School of Geography, Environment and Earth Sciences), with the first draft of the Māori Data Sovereignty table and analysis compiled by Ben Ritchie (then Service Lead, Nicholson Consulting). Feedback on draft versions was provided by Dion O'Neale and Emily Harvey (COVID Modelling Aotearoa).

Suggested Citation

Dickinson, P. R. (2025). *The Use and Misuse of Geolocation Mobility Data: A Literature Analysis and recommendations for ethical research*. COVID Modelling Aotearoa & Nicholson Consulting. https://www.covid19modelling.ac.nz/geolocation-data/

Executive Summary

This document provides a summary of the literature regarding the appropriate use of individual geolocation mobility data sourced from advertisements on mobile phone apps. It seeks to answer whether the ends (i.e. benefits) justify the means (i.e. collection and use) of this data from an impact, accuracy, and ethical perspective.

While this Literature Analysis originates from an Aotearoa New Zealand context – and frequently refers back to that context – the discussion is relevant for any researchers or analysts considering the use of individual geolocation app data in any context.

The following 12 Key Insights from the literature are identified and discussed:

- **Insight 1:** The technical accuracy of mobile geospatial data is uncertain.
- **Insight 2:** The patterns of who is missing and who is present in geospatial mobile data are unevenly distributed and create further inequity.
- **Insight 3:** The collection or use of individual geolocation data is not justified unless it is put to good use.
- **Insight 4:** Individual-level geographic data raises special ethical concerns.
- **Insight 5:** Tension between the privacy and the potency of geolocation app data exists within a broader tension field.
- **Insight 6:** Invasions of individual geoprivacy may be more or less justifiable based on social context.
- **Insight 7:** Truly 'informed' consent from users for the collection of geolocation app data is largely a myth.
- **Insight 8:** The Privacy Act (2020) raises considerations for the use of geolocation app data.
- **Insight 9:** Geolocation app data systems do not align with a te ao Māori perspective, as expressed in terms of Māori Data Sovereignty.
- **Insight 10:** Several guides exist for the safer use of geolocation app data.
- **Insight 11:** Comparably invasive datasets in Aotearoa are subject to strict data protections.
- **Insight 12:** The use of raw data collected in the geolocation app data industry is near incompatible with ethical practice.

A series of ethical self-reflection questions and a decision-tree are presented in the concluding section of this document to aid researchers in deciding whether their use of geolocation app data might be ethically justifiable.

This Literature Analysis concludes that the ends largely *do not* justify the means when it comes to the use of commercially collected individual geolocation app data.

1 Introduction

This Literature Analysis **investigates the accuracy and ethical use of geolocation data sourced from mobile app advertising**, and **produces guidance for ethical research**.

This research was motivated by the need to understand the ethics and accuracy of using geolocation datasets for COVID-19 related modelling and analyses to support future pandemic planning. This means that COVID-19 applications are frequently used as an example in this document, and in the supporting literature. However, the technical and ethical discussions of geolocation app data within this document will also apply to researchers considering using such data for other purposes (e.g. transport mobility analyses, urban planning, tourism research, etc.). This document focuses on assessing how geolocation app data might be used appropriately, seeking to inform researchers of several relevant considerations; other sources discuss the potential beneficial uses of such data for future research in greater depth (e.g. see Campbell, 2024).

Mobile app geolocation datasets are commercially available from several different private organisations, or 'data brokers'. Such datasets contain a large quantity of de-identified individual-level high-resolution GPS data (latitude and longitude) over time. Some also include inferred personal characteristics such as ethnicity, gender, age, and occupation. The data is usually sourced from mobile app advertisements, likely acquired by the data broker (organisations such as SafeGraph, Azira, Babel Street, and many more) on the real-time bidding (RTB) market.

While some discussion in this Literature Analysis is related to the broader ethics associated with either mobile phone app data or geolocation data in isolation, the core focus is on discussing the use of datasets that bring these types of data together. In other words, discussion centres on large scale individual geolocation data derived from mobile phone apps, referred to as *geolocation app data* throughout this document. Nevertheless, many insights from this analysis will be applicable to researchers considering using *any* form of individual-level geolocation data in their research projects.

Section 2 surfaces and discusses 12 key insights arising from the literature across a broad range of topics, covering both the likely accuracy and the ethical use of geolocation app data. Section 3 provides concluding comments and recommendations for future research practice. Concise advice to researchers is also dispersed throughout the Literature Analysis in blue 'Research Advice' callout boxes. Appendix A provides a glossary of commonly used terms and acronyms, while Appendix B outlines the research questions that motivated this Literature Analysis, and also provides a brief description of the approach to undertaking the search of the literature.

2 Key Insights

This section provides a summary of the key overarching insights on ethical geolocation app data use gleaned from reviewing the literature. Ultimately, this Literature Analysis seeks to answer the question: bearing in mind the ethics and accuracy of geolocation app data, do the ends justify the means? In other words: **should such data be used?**

First, however, a note that should be considered alongside the interpretation of the insights below: by nature, much of the critical discussion in the literature tends to focus on potential inaccuracies, flaws, or ethical issues relating to the use of the data. Alternatively, sources focus on discussing the benefits that can be derived from such data without extensively considering the aforementioned negative aspects of its use.

In their discussion of geospatial data use in *Data Power*, Thatcher and Dalton (2022) highlight and critique exactly this dichotomous tendency:

> "Non-fiction narratives about technology tend to be either utopian or dystopian: eschatological visions of mobile applications ending pandemics or of drone strikes silencing political dissent. Accounts of Google's attempted smart-city in Toronto or Cambridge Analytica make for great stories, but they miss the forest for the trees. Both tropes oversimplify complex processes and contexts, hamstringing attempts to understand how individual cases reflect broader systems. Processes of profit-seeking and capital accumulation frame recent discussions around technology, delimiting what is thought possible and desirable for technology to do. *That need not be the case. More alternatives are possible*." (p.2, italics original)

In other words, geolocation data is neither inherently completely good nor completely bad – it can be *both* good *and* bad. Therefore, the question of 'do the ends justify the means?' is to some extent a personal one. To put it in very simple terms: the badness of the bad things associated with the data and the goodness of the good things will likely vary in strength based on an individual's (or collective's) personal philosophy, ethics, and broader values.

However, existing discussion in the literature and the ideas in collective moral frameworks (e.g. legislation or Māori Data Sovereignty principles) provide some guidance to help individuals better understand likely issues and to form their position. This Literature Analysis therefore provides a literature-based overview on the topic of geolocation app data use to support the ethical decision-making of researchers considering the use of it.

The key insights section begins with a summary of the literature's discussion on the accuracy of geolocation app data in Section 2.1, seeking to first understand whether the data is reliable enough to warrant using. Discussion then turns to summarising broader ethical and human rights concerns in Section 2.2, exploring whether the use of the data is likely ethically justifiable. Section 2.3 discusses the Aotearoa New Zealand specific context surrounding any potential use of individual geolocation app data, while Section 2.4 concludes the key insights with an overview of existing guidance around the appropriate use of such data.

2.1 Is the data reliable enough to bring the promised benefits?

If geolocation app data is as accurate and representative as many claim, then there is no doubt that it has great potential to be used for public good, including for COVID-19 modelling. Indeed, many have lauded the potential of geolocation app data for disease disaster response, including for COVID-19 specifically (e.g. Buckee et al., 2020; Cinnamon et al., 2016; Gao et al., 2020; Garattini et al., 2019; O'Connor et al., 2021).

While geolocation data is seen as useful throughout the four stages of the disaster cycle, for disease disasters it is often positioned as most useful for the immediate pandemic *response phase*. As O'Connor et al. summarise: "The data … holds great potential for improving management during all aspects of the disaster cycle, but provides particular opportunities during the response phase where accurate and rapid data is crucial" (2021; S217-S218).

However, the assumption that geolocation app data is accurate and representative is a problematic one; the robustness of such datasets should be interrogated (as all datasets should be) to ensure that any insights derived from them are grounded in reality. This section discusses the accuracy issues associated with mobile geolocation data to support informed decision-making regarding the use of it.[1]

Many groups have a vested interest in asserting that geolocation app data is accurate. For instance, researchers and state organisations want to have a cost-effective, large-scale dataset that can provide robust results, and so may (unintentionally) overestimate the usefulness of such datasets. Ultimately, however, those with the greatest incentive to present big data as infallible are those earlier in the data product value chain with the greatest access to the data itself; those collecting and selling or re-selling the data.

Much of the literature discusses the profit-driven collection of data (so-called 'data capitalism'), and the problems that arise from this system. For example, Dalton et al. (2016) emphasise that the big data that is collected under data capitalism will best reflect what is profitable, not necessarily what is comprehensive and accurate:

> "When we map Big Data we map the contours of capital, one intrinsically limited by the uneven contours of data as it plays out across space. … In parallel, data silences or gaps result from the kinds of data deemed worth creating and storing. Simply put, **corporate data is meant to create a profit, its veracity secondary to its economic value**." (Dalton et al., 2016: 6, bold added)

[1] Note that many accuracy and representativeness issues also have ethical dimensions to them. For example, underrepresenting certain demographics mean those groups of people will not be adequately considered in pandemic response decision-making. Such issues will be discussed in this section rather than in the more general ethics discussion of Section 2.2.

Given this profit motive, information from the private companies that specialise in collecting and manipulating geolocation app data will likely present an optimistic view of the accuracy and overall usefulness of their data. Such businesses have the most to lose should the data be inaccurate, meaning that they tend to overstate the reliability of their products (Thatcher & Dalton, 2022).

While important to bear in mind when assessing the accuracy of big data, the profit motive behind such data should momentarily be set aside to determine the usefulness of the data as objectively as possible; just because businesses in data capitalism have an incentive to overstate the usefulness of their product does not mean that they do. Dalton et al. continue from the above to assert that "researchers … must simultaneously accept the epistemological limits set by the profit imperatives of much 'Big Data,' and ask what data should not be collected or analyzed … 'What's missing' (and who) thus remains of paramount importance" (2016: 7). Here, the authors mainly refer to the ethics of using data that underrepresents missing groups, but their point nonetheless still stands from the perspective of understanding accuracy issues associated with missing data.

Fortunately, several researchers have specifically investigated the accuracy and representativeness of geospatial or mobile data in general, as well as geospatial mobile data specifically. Issues associated with accuracy and representativeness can be broadly separated into two categories: 1) the technical accuracy of geolocation app data collection, and 2) the equity of geolocation app data collection coverage. The following two insights address these categories.

Insight 1: The technical accuracy of mobile geospatial data is uncertain

Bähr et al. (2022) provide a comprehensive description of some of the variables that can impact the reliability of geolocation app data collection. These variables range from characteristics of the mobile device (e.g. the operating system) to individual uses of devices (e.g. leaving devices behind when traveling or exercising). Ultimately, Bähr et al. (2022) assert that this renders a notable portion of mobile geolocation datasets inaccurate or potentially missing altogether:

> "several factors can potentially influence whether measurements are recorded or valid. For example, built-in sensors from different manufacturers might differ in their measurement quality, different operating system (OS) settings affect geolocating accuracy, third-party apps stop data collection, and participants themselves affect measurements by their handling of the smartphone (e.g., handgrip and walking style; Blunck et al., 2011). However, even ignoring these issues, **smartphone sensors might offer a limited perspective of human movement and physical activity**. If we assume that some users do not always carry around their devices or turn them off from time to time, then immobility inferred from smartphone data might just represent a device forgotten at home or resting in a desk drawer at work, a purse, or a gym locker (see e.g., Casilari et al., 2016, p. 2). **In any given smartphone data collection of geolocation, a significant proportion of the measurements that were designed to be collected will not be successful**. This missingness can be caused by intended or unintended user behavior, such as switching off the smartphone or running out of battery, by characteristics of the

> location (no signal), or by technical problems related to the hardware or OS of the smartphone (e.g., energy-saving modes)." (Bähr et al., 2022: 213-214, bold added)

Bähr et al. (2022) are not alone in their concerns, with several other authors also questioning the reliability of GPS-sourced geolocation data for particularly granular analyses. While relevant to any applications of geolocation app data, discussion in recent years has often focused on the use of individual level data to research social distancing and the spread of COVID-19 from person to person. For example, Gao et al. assert that aggregate GPS data is not reliable enough to get an indication of social distancing for COVID-19 analyses owing "to the mobile phone Global Positioning System horizontal error and uncertainty" (2020: 2), while Thatcher and Dalton note that the "spatial resolution of location tracking exists at a scale other than that necessary to control an airborne virus" (2022: 61). Chen (2020) points towards accuracy issues inherent to the use of GPS mobile phone data being likely to result in many 'false positives' for COVID-19, with healthy, low-risk people being unnecessarily quarantined.

Also discussing COVID-19 analyses specifically, the Human Rights Watch notes that "Inaccuracies associated with mobile location tracking programs raise questions about whether the restrictions they impose on privacy are necessary to safeguard public health" (2020). The Human Rights Watch also identifies other research that indicates GPS data is of uncertain reliability for COVID-19 analyses, and calls into question whether such data would be useful even if it were accurate:

> "A key consideration is whether mobile location technologies can accurately determine whether a person is in close contact (within 6 feet of someone for 10 or more minutes) of someone who is infected. Technology researchers have found that cell site location information and **GPS signals are unlikely to provide location estimates with the level of precision required to meaningfully predict the risk of Covid-19 transmission**. … Furthermore, proximity tracing alone says very little about the nature of the interaction, such as whether people were in a closed space or outdoors, whether they were wearing masks or not, or whether someone sneezed during the interaction." (Human Rights Watch, 2020, bold added)

The Human Rights Watch (2020) and others raise similar accuracy concerns to Bähr et al. (2022), noting that device sharing or multiple device ownership has complicated disease disaster location tracking efforts in the past (see also Chen, 2020; Cinnamon et al., 2016). Thatcher and Dalton (2022) extend on this discussion by noting several forms of intentional 'active resistance' to data capitalism that prevent or falsify mobile location tracking efforts, including turning off location services on a phone, GPS spoofing (i.e. fabricating geolocation coordinates), using a VPN, or using ad-blockers. Thatcher and Dalton (2022) even identify a subset of people who attempt to 'escape' from data collection altogether by using a 'dumb phone', though they acknowledge that this ability to escape from contactability in developed nations is generally reserved for an economically privileged (or extremely underprivileged) few.

Some authors have sought to determine the extent to which adding mobility data to COVID-19 case models brings predictive benefits. For example, Abrar et al. (2023) found that the mobility data in US county-based COVID-19 models were of limited usefulness compared to individual regional COVID-19 case prediction models without mobility as a source of

information. They evaluated model performance using nine mobility datasets across four different data supplying companies (SafeGraph, Apple, Google, and Descartes), and identified that "at most, 60% of counties improve their performance after adding mobility data", and that these "performance improvements are modest" (Abrar et al., 2023: 1).

Not all authors are so critical of mobile geolocation data, however. Baron and Musolesi (2020), for example, conducted a study to ground truth *inferred* personal data (e.g. inferred demographics, activities, interests, characteristics) from geolocation to test its accuracy. They found that most information (75%) was rated as relevant by their app users, though there was arguably still a notable portion that was not clearly relevant (14%) or not relevant at all (11%). However, it is worth noting that Baron and Musolesi (2020) used a specific method for inferring personal information; the authors' method likely differs at least somewhat from the methods used by commercial companies, so does not necessarily speak to their accuracy. Moreover, their study is based on only 69 users of the app, so should not be interpreted as broadly representative of the large commercially-available mobile app datasets – further research (including in relevant local cultural contexts) is likely needed on the accuracy of inferred personal information in mobile app datasets specifically.

Overall, geolocation app data has several inherent technical limitations that call its reliability into doubt. That is not to say that the data is useless; simply that the accuracy of the datasets should not be overstated, and that careful attention should be paid to how the data is cleaned and prepared for analysis (for example, by following the 5-stage geolocation data preparation model laid out by Bähr et al., 2022). Depending on the intended uses of the data some of these accuracy issues may not matter greatly – for example, while geolocation app data might be unsuitable for measuring the spread of COVID-19 from person to person, it may still be accurate enough to understand the regions or neighbourhoods that people travel to or operate within for other purposes.

However, some accuracy issues related to geolocation app data are currently insurmountable regardless of intended use: we do not have adequate information to determine how many (and what type) of geolocation records are missing, misleading, or completely false.

Research Advice:

Nuanced communication of the findings from geolocation app data is essential – researchers using such data must ensure that they are clear about the likely limitations inherent to the datasets for their audiences.

Insight 2: The patterns of who is missing and who is present in geospatial mobile data are unevenly distributed and create further inequity

Several authors discuss inequity relating to the collection and use of geolocation app data. Inequity in data collected is concerning because unequal data can be misleading, resulting in poor decision-making that may lead to further entrenching any existing inequities. Geolocation app data inequities exist in the form of bias in datasets relating to numerous different characteristics such as wealth or class, urbanity or rurality, race, gender, age, or

ability. One source sums up the accuracy implications arising from these biases for epidemic analyses specifically as follows:

> "In our study we show that biases embedded in data can have a substantial impact on the pattern of epidemic spread. Biases can cause a severe over- or underestimation of key epidemic properties such as the severity of the peak, or the arrival time of the epidemic; both are characteristics which governments use for planning and responding to pandemics. These effects are especially deceiving because of their heterogeneity, as they can heavily impact certain regions but leave others unaffected, and can even go unnoticed when focusing on national averages" (Schlosser et al., 2021: 8)

Biases in data therefore impact different groups unevenly, with those already underprivileged the most likely to be missed out. Indeed, the Human Rights Watch (2020) emphasise that there is a tendency for minorities and vulnerable groups in general to be excluded from mobile phone tracking data. As they explain, at the core of this issue of bias and representation is that access to smart phones and the internet is unequal along several lines, and that this inequality may bring yet further inequity if geolocation app data is relied on as a source of truth:

> "Disparities in access and use of mobile devices based on location (urban versus rural) and gender are also well documented and generally reflect and entrench broader patterns of inequality. Older people – a group that is at increased risk of severe disease and death in the Covid-19 pandemic – are also less likely to use specialized apps or have smartphones or even access to the internet. In the US, a 2019 Pew survey found that 68 percent of older people between ages 55 to 73 own smartphones, compared to 93 percent of people ages 23 to 38." (Human Rights Watch, 2020).

Like the Human Rights Watch, several sources focus specifically on one or two characteristics associated with bias in mobility datasets. For example, Schlosser et al. (2021) largely focus on the impacts of biases associated with phone user wealth or class in their research. They find that, across the countries in their study, the wealth of phone users has a notable impact on the quantity of mobility data that they create: "high-wealth users are overrepresented, with the wealthiest 20% of users … contributing approximately 50% of all recorded trips, while the poorest 20% … produce less than 5% of all trips. Taken together, the bottom 80% of users produce approximately the same amount of data as the wealthiest 20%" (Schlosser et al., 2021: 3). Schlosser et al. go on to explicitly identify the modelling issues that arise from this overrepresentation of wealthy individuals: "the aggregate travel network is skewed towards the mobility patterns of rich people, … this distorts the outcomes of dynamic simulations, such as the conclusions we can draw from epidemic models." (Schlosser et al., 2021: 3).

Cinnamon et al. share these concerns, noting from their own research that "With phone ownership, there tends to be a bias towards the more wealthy and mobile, meaning the results may not necessarily reflect the wider population's movements (Tatem et al., 2014), and may in fact be an overestimate (see also Xu et al., 2016; Zhao et al., 2016)" (2016: 258). Schlosser et al. (2021) compare the differences between a 'de-biased' mobility network and the original source mobility network, and find substantial differences in the sub-national spreading pattern of the epidemic through those networks. In other words, the biased version of their network based on the source geolocation app data was likely wrong.

Several authors also note concerning rural-urban biases in geolocation app mobility data, with rural groups generally underrepresented (Chen, 2020; Cinnamon et al., 2016; Human Rights Watch, 2020; Schlosser et al., 2021). However, Schlosser et al. do note that the impact of biases in remote regions warrants further research "to make clear whether there is, in fact, a connection to the demographic properties of the region" (2021: 8), while Cinnamon et al. (2016) indicate that rural representativeness will likely continue to increase with time alongside rising mobile phone use and infrastructure development. Further, accuracy issues also exist for highly urbanised areas, and may be difficult to overcome: GPS tracking can produce errors when people are indoors, particularly in high-density urban areas with many multi-storey buildings (Chen, 2020).

Moreover, analysis from Abrar et al. demonstrated an important equity and accuracy dimension to COVID-19 model improvements stemming from mobility data, with the authors noting that "improvements were lower for counties with higher Black, Hispanic, and other non-White populations as well as low-income and rural populations, pointing to potential bias in the mobility data negatively impacting predictive performance" (2023: 1).[2] The authors conclude with a warning to researchers considering the use of commercially available mobility data in COVID-19 modelling:

> "Our analysis shows that **purchasing county-level mobility data will not benefit many counties, and decision makers should proceed with caution accordingly**. It is also concerning that the extent to which mobility data improves predictions is in part a function of the composition of the population in the county. In fact, across most of the mobility datasets, correlation improvements were lower for counties with higher Black, Hispanic, and other non-White populations as well as low-income and rural populations. As older and minority patients have been disproportionately affected by the COVID-19 pandemic, we would hope to provide more and better resources to these groups to ameliorate the disparities. Instead, we see that **mobility data could serve to entrench these disparities**, providing decision makers in counties with more vulnerable populations with worse-performing models, **leading to worse-informed policy decisions**." (Abrar et al., 2023: 18, bold added)

These concerns will apply to the use of mobility datasets for other modelling purposes as well. Based on the findings of Abrar et al. (2023), instead of leading to positive outcomes, including commercially available mobility datasets in modelling may only serve to mislead decision-makers and to further entrench existing inequities.

[2] Abrar et al. (2023) point out that bias might impact the collection of COVID-19 case data as well as geolocation data, but that given they are comparing both mobility-based and non-mobility based baseline models they are confident that the performance differences between the two are specific to the mobility data.

Geolocation Data Inequity in Aotearoa

Several Aotearoa-specific considerations are also worth bearing in mind here, particularly regarding the consequences of digital inclusion and exclusion. The Digital Inclusion Research Group (2017) identified the following groups as most at risk of digital exclusion in Aotearoa:

- families with children in low socio-economic communities,
- people living in rural communities,
- people with disabilities,
- migrants and refugees with English as a second language,
- Māori and Pasifika youth,
- offenders and ex-offenders, and
- seniors.

Their report highlights several stark (albeit now somewhat dated) statistics with implications for who is missing from mobile geolocation datasets. For example, the authors point to research indicating that only 14% of people who are blind have access to data enabled accessible mobile technologies, and that only 68% of households with incomes below $35,000 had internet access, in contrast with 99% of households with income over $100,000 (Digital Inclusion Research Group, 2017). More disadvantaged groups in Aotearoa will likely therefore be underrepresented in any datasets based on mobile phone app usage.

Several sources discuss the equity implications of Māori digital exclusion in Aotearoa in particular, even if not in the context of mobility data specifically. For example, Timutimu (2023) identifies that Māori and Pasifika have lower internet usage than other ethnic groups in Aotearoa, and points to a lack of access to affordable internet and computing equipment as a key driver of this growing digital divide. With implications for wealth, education, and ability as well, Timutimu (2023) also notes that 22% of Māori have below essential digital skills, in part driven by low household income (32% of people in this category lack essential digital skills), lower levels of education (28% of people in this category lack essential digital skills), and higher rates of disability (42% of people in this category lack essential digital skills). Timutimu provides a stark warning of the implications of this digital divide for Māori: “if Māori continue to be excluded from the digital and technology space Māori will be left behind entirely” (2023: 3).

Taken together, the above information indicates that geolocation app datasets will generally underrepresent those who are already disadvantaged, and will overrepresent those who are more advantaged – perhaps excluding a hyper-privileged few who can afford to keep their data protected or to avoid creating it in the first place. If decisions are made based on models created using geolocation app datasets with these biases, then not only will they likely be incorrect, but we can also expect existing inequities to become even more greatly entrenched, including here in Aotearoa.

Research Advice:

Researchers should critically review their location datasets to understand the biases present, and should clearly communicate such biases to avoid decision making that may have unintentional inequitable impacts.

Insight 3: The collection or use of individual geolocation data is not justified unless it is put to public good use

While not strictly accuracy related, a relevant concern raised by some authors discussing the accuracy of geospatial data is whether big data analysis for public good purposes will actually be put to use for the public good – in other words, the efficacy of the data for positive impact. Accuracy is nearly irrelevant to emergency responses if the insights from the data are never applied.

Oliver et al. (2020) in particular provide commentary here, primarily reflecting on a lack of uptake for mobile big data among governments even after research demonstrating its possible uses in the prior Ebola response. They offer five key explanations for this low uptake, one of which focuses specifically on poor communication between researchers and those directly responding to the disaster:

> "**researchers and technologists frequently fail to articulate their findings in clear, actionable terms that respond to practical political and technical questions**. Researchers and domain experts tend to define the scope and direction of analytical problems from their perspective and not necessarily from the perspective of governments' needs. Critical decisions have to be taken, while key results are often published in scientific journals and in jargon that are not easily accessible to outsiders, including government workers and policy makers." (Oliver et al., 2020: 2, bold added)

In light of this observation, any future research drawing on geolocation app data that uses pandemic preparedness as a justification must ensure that the research is communicated well. Oliver et al. (2020: 4) provide four key principles to help improve the effectiveness of such research:

(i) the early inclusion of governments,
(ii) the liaising with data protection authorities early on,
(iii) international exchange, and
(iv) preparation for all stages of the pandemic.

If the research does not progress in line with these four principles, then Oliver et al. have more to say: "Many insights derived from mobile phone data analytics do not have practical implications—such analysis and the related data collection should be discouraged until proven necessary." (2020: 5).

Oliver et al. (2020) and many others argue that there must be a strong public good justification to consider conducting research with personal geolocation app data given the various ethical issues associated with the data. The following section focuses on exploring these ethical concerns in greater depth.

2.2 What is the broader ethics and human rights discussion around the use of such data?

In short: it's complicated. The literature on ethical research using geolocation mobility data (including but not limited to app data) is diverse, complex, and spans multiple disciplines and sub-disciplines. However, one common recognition is that there are unique dimensions to geolocation data in particular. This uniqueness brings strong potential for in-depth analysis, but importantly it also brings unique opportunities for harm and unethical use that are less present for other forms of data; the discussion of which is the focus of the first insight in this section.

Insight 4: Individual-level geographic data raises special ethical concerns

Several sources directly discuss the fact that geolocation data presents unique ethical challenges (e.g. see Baron & Musolesi, 2020; EthicalGEO, 2021; Keßler & McKenzie, 2018; Thatcher & Dalton, 2022; Zhang & McKenzie, 2023). EthicalGEO succinctly summarise the reason that location data *in particular* must be carefully safeguarded from harmful use:

> "Many of the sources of potential harm identified in other fields can occur in relation to location data, including bias in datasets, privacy intrusion, and misuse of power imbalances in markets. But while the examples are very useful, it is important to recognize that **risks relating to use of location data can have impacts that are specific to those uses, to people and to places**. To be followed around where you go in the world has parallels with being tracked around the internet, but it is not the same thing. Also, while a lot of data regulation is framed around the rights of individuals, use of location data can affect groups of people, including some already under pressures relating to where they are, like refugees and other migrants. **Location data lends specific powers, which can imply specific responsibilities**." (EthicalGEO, 2021:3, bold added).

Several real world examples of harmful, targeted use exist. One oft-cited example is of a conservative group in the US who spent millions of dollars on sex app geolocation data (from the RTB market) to track the celibacy of Catholic priests (Ryan & Christl, 2023a; 2023b). Data on one priest's use of a gay app and his visits to private homes was made public by the organisation, resulting in the individual leaving the priesthood. Ryan and Christl note that "beyond the severe intrusion into privacy, the example shows how a similar operation can be conducted to blackmail or manipulate … personnel and leaders" (2023a; 2023b:16).

An even more recent example comes from Wyden (2024), which surfaced the use of the company Near Intelligence's commercial dataset by anti-abortion groups to aggressively target distressing quantities of anti-abortion messaging (14.3 million ads in 2020) at anyone who visited Wisconsin family planning clinics. Cox (2022) similarly surfaced how geolocation app data was being packaged by the company SafeGraph on planned parenthood visitors specifically, allowing for easy and cheap tracking and targeting of that group.

Many such examples are extreme, and appear easily avoided: after all, most researchers do not work for a conservative organisation interested in the sexual preferences of priests. However, such examples also expose the deeply personal nature of geolocation data, and do indicate that it must be carefully secured, and its use by anyone within a team carefully monitored. Even one malicious individual with access to geolocation data (current or historical) can do a great deal of harm. Ethical practice would require that a research team be confident that *nobody* with access to the data will potentially use it maliciously, or even just inappropriately. Could you know for certain that nobody with access to the data is (for example):

- A perpetrator of domestic violence who could track down family members that wish to remain hidden?
- A person who harbours hatred for a particular demographic, and could track and harass or expose them?
- A person that may wish to use personal location information to blackmail those they know, or those in a position of power?
- Simply someone who will inappropriately surveil themselves, their family, their friends, their students, their colleagues, or anyone else that they may know?

It is largely due to the significance of personal geolocation data that Zhang and McKenzie (2023) argue for conceptualising such data as 'platial' (as in 'place') rather than 'spatial' (as in 'space'). While spaces are abstract and relatively easy to quantify, places have meaning ascribed to them, and are generally more fluid and harder to define. Zhang and McKenzie assert that thinking from a platial perspective allows researchers "to rehumanize geoprivacy as it is a concept that involves flesh and blood instead of numbers alone" (2023: 20). They go on to elaborate that:

> "Protecting geoprivacy is therefore more than uniformly masking locations to a certain degree without considering perceived risks from multiple facets. Thinking from the platial perspective, we can discover shared implicit attitudes and move the discipline from analyzing individual concerns towards protecting group privacy." (Zhang & McKenzie, 2023: 20).

This notion of group privacy is frequently raised by the literature (for more, see Mittelstadt, 2017; Zhang et al., 2022), and further complicates considerations around the ethical use of geolocation data – location data is very rarely solely relevant to individuals in abstraction from those near them.

In essence though, perhaps the most important way in which location data is special is its potential to identify individuals. Geolocation data (especially mobile app geolocation data) uniquely allows for easy re-identification of supposedly anonymous individuals, meaning that standard approaches to protecting individuals in datasets will not always work for such data. As de Montjoye et al. explain:

> "**pseudonymization and standard de-identification are not sufficient to prevent users from being re-identified in mobile phone data**. Four data points—approximate places and times where an individual was present—have been shown to be enough to uniquely

re-identify them 95% of the time in a mobile phone dataset of 1.5 million people" (2018: 1, bold added).

Clearly, many factors are at play when considering the use of geolocation app data. One of the most effective attempts to synthesise the complexity of these concerns comes from Keßler and McKenzie (2018). They propose 21 'theses of geoprivacy', beginning with seven theses focusing solely on why spatial data is special. Each of these 21 theses (summarised in Table 1) is worth acknowledging individually in considering the use of geospatial data.

Table 1: The 21 theses of geoprivacy. Summarised from Keßler and McKenzie (2018).

Topic area	Thesis
Spatial data is special	"**Thesis 1** Information about an individual's location is substantially different from other kinds of personally identifiable information." (p.5)
	"**Thesis 2** Ubiquitous positioning devices and easy-to-use APIs make information about an individual's location much easier to capture than other kinds of personally identifiable information." (p.6)
	"**Thesis 3** Users of information services have a substantial incentive to share their location with service providers, as location information can significantly improve the quality of a service and make it more useful." (p.6)
	"**Thesis 4** Users often share their current location unknowingly." (p.7).
	"**Thesis 5** Having access to a user's location history allows for a broad range of *location-based inferences*, such as information about their health, consumer behavior, or social status." (p.7, italics original)
	"**Thesis 6** Location-based inferences can reveal information that the user never intended or agreed to share with a service." (p.7)
	"**Thesis 7** Incorrect location-based inferences can have severe adverse effects for affected individuals, with little or no opportunity to rectify these errors." (p.7)
Economic value of location information	"**Thesis 8** Knowing a customer's location is an economic asset for a business." (p.8)
	"**Thesis 9** Users value their own location information based on level of detail and use case." (p.9)
	"**Thesis 10** A new market is currently emerging in which businesses and users directly trade personal-level location information." (p.9)
	"**Thesis 11** Discounts for customers who agree to share their location with a business are effectively penalizing customers who refuse to do so, and may erode the solidarity principle behind collective insurance." (p.9).
Safeguarding geoprivacy	"**Thesis 12** Preserving geoprivacy involves more than obfuscating geographic coordinates. Location can be inferred from non-explicit geospatial information such as interests, activities, and socio-demographics." (p.10)
	"**Thesis 13** Any location-based service offered to a user is limited by the amount of private information the user is willing to share." (p.10)
	"**Thesis 14** Mobile operating systems lack fine-grained control mechanisms for location services, thus severely limiting the degree of control users have over their location information." (p.11)

Topic area	Thesis
	"**Thesis 15** An individual's level of geoprivacy cannot be reliably assessed because it is impossible to know what auxiliary information a third party may have access to." (p.11)
Legal and ethical aspects	"**Thesis 16** The ethical ramifications of advances in location-enabled technology are often viewed as an afterthought and legal concerns over privacy aspects lag behind technological advances." (p.12)
Geoprivacy as a tension field	"**Thesis 17** Geoprivacy as a research topic is situated in a tension field between technological, ethical, economical, legal, and educational aspects that have only been addressed separately so far." (p.12)
	"**Thesis 18** Users often have no way of checking whether the location-aware services and devices they use act within the legal and ethical frameworks and adhere to the provided description and privacy policy." (p.14)
	"**Thesis 19** A higher level of user education in the area of position tracking and location-based services is required to allow them to make more informed decisions about the tools and services they are using." (p.14)
	"**Thesis 20** A better-educated user base can push for more restrictive legislation and force service providers to be more transparent about their data collection and use policies." (p.14).
Conclusion	"**Thesis 21** Constant surveillance of citizens' locations can be used as a tool for oppression and to limit freedom of speech, even in democracies." (p.15)

Some theses focus on the potential benefits of geolocation data to individuals (2, 3, and 5), while others focus on the potential economic value of such data (8, 9, and 10). However, in spite of only one thesis (16) falling under the specific heading of 'legal and ethical aspects', the vast majority of the theses outlined by Keßler and McKenzie (2018) have important accuracy, privacy, and ethics implications for the use of geolocation app data. Perhaps most significantly, theses 7 and 21 note that the use of geolocation data can have severe adverse impacts at both the individual and societal level respectively.

Research Advice:

Regardless of the intended use of geolocation app data by researchers, the potential for the data to be used for harm must be considered, as use of the data in one context sends a signal of support for its legitimacy in other contexts.

Keßler and McKenzie (2018) provide an apt description of the various tensions inherent to the use of geolocation data through their identification of a 'geoprivacy tension field' (thesis 17). This concept is worth discussing in greater detail to frame any consideration of other ethical or privacy issues, and so is the focus of the following insight.

Insight 5: Tension between the privacy and the potency of geolocation app data exists within a broader tension field

A common topic of discussion throughout the literature is the tension between, on the one hand, the privacy, protection, and safe use of data, and on the other hand, the utility of the data for beneficial purposes (e.g. see for example de Montjoye et al., 2018; Chen, 2020; Cinnamon et al., 2016; Garattini et al., 2019; O'Connor et al., 2021; Thatcher & Dalton, 2022;).

Such beneficial utility might be aligned with:

- the state (e.g. increased national security or policing; faster disaster response),
- commercial interests (e.g. more targeted advertising; higher value data to sell), or
- individual uses (e.g. more relevant and targeted services for people).

Regardless of the nature of the utility, the higher the privacy protections, the less granular the data generally is, and therefore the less broadly useful it is. The more granular and personal the data is, the more potentially invasive the data and the higher the potential for benefit through use or for damage through misuse (malicious or otherwise). For the most part, discussion of this tension centres on this simple dichotomy which I label the *data privacy-potency tension* (see Figure 1). In other words, the tension between the privacy protections (or lack thereof) in the data and the potency of the data for analytical use.

Figure 1: The Data Privacy-Potency Tension – each personal geolocation dataset exists somewhere on this spectrum. The more potent the data becomes, the less private it tends to be, and vice versa.

A great deal of discussion in the literature focuses on the 'privacy' aspect, while a similar portion focuses on the 'potency' aspect (i.e. using the data). A smaller – albeit still notable – portion of the literature explicitly acknowledges and discusses the tension between these two factors. However, Keßler and McKenzie (2018) moves beyond this simple dichotomous privacy-potency tension to instead discuss a broader geoprivacy 'tension field' (see Figure 2).

This idea of a tension field more effectively communicates the difficulty of navigating the use of geolocation app data than the privacy-potency tension; the sheer complexity of the data's creation, use, regulation, application, and more all make the appropriate use of data a less-than-straightforward undertaking. Indeed, one could easily argue that the tension field is even more complicated than the one depicted given that Figure 2 does not incorporate groups of users (including their rights, utility, and privacy), any cultural context-specific guiding frameworks beyond the law (e.g Māori Data Sovereignty principles in Aotearoa), or the sheer complexity of ethical considerations.

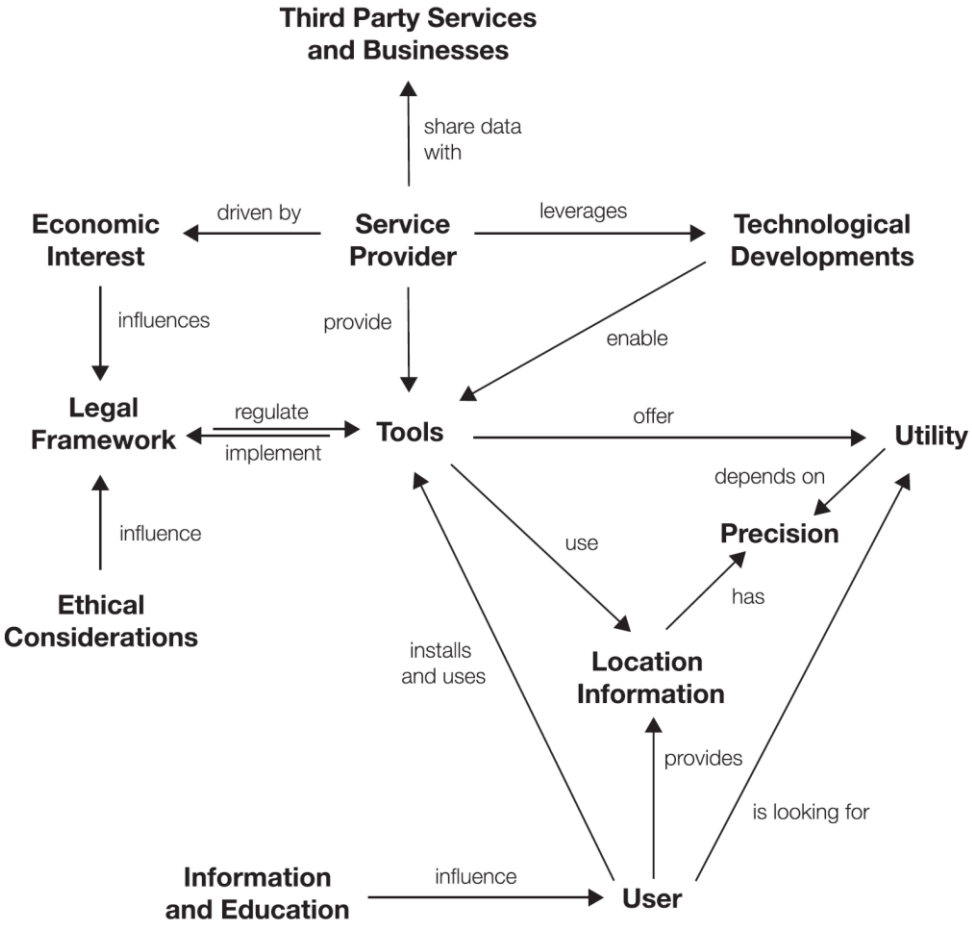


Figure 2: The Geoprivacy Tension Field, as created and explained by Keßler and McKenzie (2018). Original caption from source: "Geoprivacy is influenced by a number of different aspects, creating a tension field that makes it difficult to tackle as a whole".

So, how should we approach using data given the geoprivacy tension field? The 19th thesis that Keßler and McKenzie (2018) describe indicates that improved user education is the key to navigating this tension field, while their 20th thesis asserts that this higher level of education will enable users to push for further regulations and hold companies to account. However, this argument could and should also be applied to researchers considering using geolocation app data.

Research Advice:

As individuals in a position of power relative to the average app user (who are not often empowered to buy, sell, or even view mobile app data) researchers using geolocation app datasets are better positioned to push for regulatory and corporate reform than most people.

Research Advice:

Researchers are therefore also well positioned to continue perpetuating and legitimising any current unethical practices of the geolocation app data industry through their inaction, regardless of their intended uses of the data.

Given the privileged position that researchers operate from, any researcher engagement with geolocation app data must not shy away from genuine, open criticism of the corporate and state infrastructure and institutions that facilitate the collection of the data. Researchers could also consider avoiding data capitalist sources of information altogether, such as by creating bespoke geolocation data collection sources (e.g. see discussion in Bähr et al., 2022) with ethical practice embedded from the outset.

Insight 6: Invasions of geoprivacy may be more or less justifiable based on social context

Many sources in the literature recognise the aforementioned data privacy-potency tension when it comes to the use of personal geolocation data. Likewise, most sources concede that there are certain emergency events that have a strong enough public good justification to *temporarily* outweigh *some* privacy impacts of using sensitive data in order to take advantage of the potency of the data.

One often discussed type of emergency – even prior to 2020 – are disease disasters such as COVID-19. The most important emphasis in many sources here is that such emergencies can only be used to justify invasions of privacy for a *specific*, *time-limited period*, if at all. Several authors assert that users of any form of individual-level geospatial data in responding to disease emergencies risk 'function creep', a process by which the use of sensitive data becomes normalised, and through which the acceptance of corporate and government surveillance is accelerated (Cinnamon et al., 2016; Miller & Smith, 2021; Oliver et al., 2020; Thatcher & Dalton, 2022).

The risks that function creep pose have implications for individual privacy, but many of the risks are substantial enough to have an impact at a *societal* level as well. For instance, Miller and Smith argue that using data only for the purposes it was collected for is essential to avoiding function creep and maintaining trust in government:

> "If liberal democratic governments are to maintain the trust of the community, data collected for one purpose, particularly to address an extraordinary circumstance … must be carefully guarded and not used for broader purposes" (2021: 368).

Moreover, they specifically take issue with the analysis of data collected during COVID-19 being justified through the prevention of future pandemics, and further go on to criticise the later linking of data collected for one purpose with data collected for another purpose. As such, all of Miller and Smith's (2021) criticisms apply to the potential future use of mobile geolocation data for pandemic analysis that might be considered now (including any linkages with other data sources such as the IDI here in Aotearoa), as the world moves well beyond the initial pandemic response phase of the emergency cycle.[3]

[3] The emergency cycle consists of four broad phases: prevention, preparedness, response, and recovery. Any analysis now would likely be contributing to the 'preparedness' phase of future pandemic emergencies.

Like Miller and Smith (2021), Oliver et al. (2020) argue that the ethical justification of disease disasters such as COVID-19 for the use of personal geospatial data is *time-limited;* the justification should not become permanent. Oliver et al. (2020) assert that there is a real risk that ongoing use of the data for COVID-19 will justify its uses for other purposes, potentially impacting civil liberties. Their discussion also emphasises that commercially gathered individual-level geospatial data is inherently more sensitive and potentially problematic than aggregate government or mobile provider data.

Indeed, even some of the advocates for the potential use of mobility data in responding to COVID-19 argue that using *individual*-level data is inappropriate due to the risks that it poses to progress in data protection and personal privacy. For example, Buckee et al. state that even though "data sharing models and data protection laws provide for the legal grounds to use personal data during emergencies … we do not advocate the use of individual data" (2020: 146).

Relatedly, Oliver et al. (2020) highlight that concerns around data protection, privacy, and civil liberties are one of the key reasons behind the lower uptake of individual mobility data to respond to disease emergencies. They point to broader calls for careful consideration of any analysis of geolocation data (even during the emergency response phase) due to the risk of function creep:

> "around the world, public opinion surveys, social media, and a broad range of civil society actors including consumer groups and human rights organizations have raised legitimate concerns around the ethics, potential loss of privacy, and long-term impact on civil liberties resulting from the use of individual mobile data to monitor COVID-19. Control of the pandemic requires control of people—including their mobility and other behaviors. **A key concern is that the pandemic is used to create and legitimize surveillance tools used by government and technology companies that are likely to persist beyond the emergency**. Such tools and enhanced access to data may be used for purposes such as law enforcement by the government or hypertargeting by the private sector. **Such an increase in government and industry power and the absence of checks and balance is harmful in any democratic state**. The consequences may be even more devastating in less democratic states that routinely target and oppress minorities, vulnerable groups, and other populations of concern." (Oliver et al., 2020: 3, bold added).

However, the authors do note that none of the challenges they have raised regarding the use of individual mobility data are insurmountable. To address the risks they have identified, Oliver et al. (2020) emphasise that the use of individual geolocation data when responding to an emergency should be overseen, should follow 'well-articulated' data policies and guidelines, and should be subject to risk assessments. Section 2.4 of this document discusses guidelines for the ethical use of geospatial data in further detail.

Thatcher and Dalton (2022) also directly discuss potential function creep resulting from COVID-19 and the use of mobile geolocation data. They emphasise the possibility of heightened surveillance over the pandemic under emergency conditions instead becoming entrenched as 'the new normal', and criticise much of public commentary for their lack of discussion of the dangers of surveillance in the context of Western nations:

> "And yet, even as outlets like *The Guardian* suggest (likely correctly) that China's increased surveillance to track COVID-19 may likely become "the new normal," left out are the similarities (and failures) of similar systems in other nations. While both individuals and algorithms pored over camera footage in China, around the globe another source of data became intrinsic to supposed disease response: as always, mobile phones. From Norway to the United States to, of course, China, mobile phone location data were touted as a means of tracking and eliminating the disease." (Thatcher & Dalton, 2022: 60-61).

In other words, heightened state surveillance being justified through public health emergencies is no less a real concern in Western democratic nations than in the less democratic states that Aotearoa and similar nations are often (favourably) compared to in public commentary. Indeed, even in Aotearoa the New Zealand Human Rights Commission cautions against the use of individual big data due to potential function creep, noting that "operational and procedural safeguards in the social sector are an essential bulwark against the risk of human rights breaches occurring and becoming normalised in relation to personal data" (2018: 44). Similarly, Chen emphasises that "once the pandemic is over it should be a given that any tracking systems must be turned off" (2020).

When it comes to function creep, such sources would therefore likely agree with the Human Rights Watch, who assert that:

> "Even in times of emergency, when states restrict human rights for public health reasons, international human rights law says that measures taken that limit people's rights and freedoms must be ***lawful, necessary, and proportionate***. States of emergency need to be ***limited in duration*** and any curtailment of rights needs to ***take into consideration the disproportionate impact on specific populations or marginalized groups***." (Human Rights Watch, 2020, emphasis original)

Further research and policy work is arguably needed in this area. Even almost a decade later the arguments of Cinnamon et al. (2016) still apply, who emphasise that "more attention to the trade-off between social harm and the public good is needed" in using mobile data to respond to disease disaster, and that this "represents a crucial question for further academic and policy debate, despite some observers claiming that ''data privacy and security . . . are largely irrelevant in the face of an epidemic outbreak" (Koch, 2016, p. 5)." (2016: 262).

Research Advice:

Any researchers considering the use of commercial geolocation app data need to bear in mind the role that they may be playing in legitimising and normalising heightened state surveillance.

Insight 7: Truly 'informed' consent from users for the collection of geolocation app data is largely a myth

At the most basic level of most contemporary models of ethical research involving humans is the idea of informed consent: individuals should have the right to consent to participate in

research after having obtained a full understanding of the research, its aims, and any risks associated with their participation.

The companies that collect, compile, and sell individual mobile phone user data (including geolocation data) will assert that individuals have provided consent – after all, they have accepted the Terms and Conditions or the End User Licensing Agreement (EULA) associated with the service being offered.

However, the literature does not mince words here: **this consent is not informed.**

There are many arguments within the literature that make this point absolutely clear. For example, O'Connor et al. discuss the "serious question over whether agreeing to the terms and conditions of an app can realistically be considered 'informed consent' at all" (2021: S216), providing the example of six London residents who, in 2014 "gave away their first-born children (unknowingly one hopes) in exchange for free Wi-Fi (Fox-Brewster 2014)." (2021: S217). The authors further note the prohibitive length of terms and conditions, citing an example from the Norwegian Consumer Council where it took nearly 32 hours to simply read the terms and conditions of the average number of apps that a Norwegian has on their phone. Thatcher and Dalton (2022) indicate a similar time burden for residents of the US: a word-for-word reading of every privacy policy on every website a US citizen visits would amount to 10 days per person per year. When was the last time you read, in full, the Terms and Conditions for a website or a EULA for an app?

In light of such a prohibitive quantity of information O'Connor et al. rightly argue that "To assert that when individuals agree to such conditions, they are engaging in an informed decision is not credible" (2021: S216) and that the "notion that through granting consent, a user makes an informed choice about the collection, use and storage of their personal information is thus a misconception" (2021: S217). Baron and Musolesi (2020) note that app users are mostly not aware of the privacy implications of the access that they grant apps to their information, especially in regard to location data. In other words, there is no 'informed' consent when it comes to the vast majority of geolocation app data. Such ideas are not new: even over 50 years ago there was clear evidence that the length of consent forms is inversely related to informed decision-making about consenting (Epstein & Lasagna, 1969).

Such a lack of dedication to informed consent practices has potentially important implications for the power dynamics between individuals, states, and corporations. Nearly a decade ago Cinnamon et al. raised concerns about "the ways in which power is being oriented away from citizens and governments towards corporations", noting that "phone users are a little recognized stakeholder group in mobile data use debates, paralleling broader processes of disempowerment of 'data subjects' who neither control nor are well-informed about the uses of their data (see Kennedy and Moss, 2015; Leszczynski, 2015; Taylor, 2015; Zwitter, 2014)." (2016: 262).

Some argue that progress regarding more informed consent will come from dedicated efforts in educating individuals about how their data is likely used (Gluckman, 2017; Keßler & McKenzie, 2018). However, commentary from others demonstrates that these efforts of individual education will be relatively unhelpful without broader systemic change, for a

multitude of reasons. For instance, Thatcher and Dalton compellingly argue that we are offered little choice but to concede our personal location data to corporations if we wish to participate in society:

> "we aren't really offered much of a choice, are we? … Increasingly broad, fundamental swathes of our existence are mediated through location-aware applications, from the romantic, such as Tinder and Grindr, to the economically mandatory (imagine handling your job without email). Tech companies offer an all or nothing choice. There is no negotiation with an EULA, we cannot agree to only certain parts or modify the terms. The benefits are immediate, such as the ability to purchase a commuter train ticket or take a photo and share it with friends, and the costs are hard to recognize." (Thatcher & Dalton, 2022: 74)[4]

The natural extension of the discussion of systemic change is more effective regulation. However, several authors note that regulations and legislation have historically failed to keep up with advancements in technology, including big data collection (Keßler & McKenzie, 2018; O'Connor et al., 2021; Thatcher & Dalton, 2022). These regulations need to take shape and keep pace with technological developments in order to support more genuinely informed consent that provides mobile app users with a choice. It is unsurprising then that O'Connor et al. argue that our "current reality thus begs the question whether even well-informed individuals can appropriately manage their personal privacy in the age of Big Data" (2021: S217).

All of this is without even discussing concerns around the ability of children or other relatively easy to exploit demographics (e.g. people with dementia) to provide consent. *Nor* does this discussion cover the fact that many apps violate privacy law by not even obtaining *un*informed consent for tracking in the first place (e.g. see Kollnig et al., 2021). *Nor* does this section discuss the fact that several geolocation app data companies have been caught illegally selling data (including to military or intelligence agencies) without consent (e.g. see Tau et al., 2023)[5], *nor* that data privacy extends beyond the individual to impact group privacy (Mittelstadt, 2017), with implications for group consent. Greater dedication to ensuring informed consent in the collection and use of geolocation app data is evidently needed.

---

[4] Whether it is a supermarket rewards app that requires you to share your data in exchange for cheaper prices during a cost of living crisis, or a new app or video game you've purchased presenting you with an extensive EULA before you can use the product you've bought; the incentive to provide uninformed consent in the immediate term is so strong that the vast majority of people no doubt don't even read a single word of what they're consenting to. Even if you're particularly privacy conscious, you likely sometimes just scroll down as quickly as possible and tick 'I agree' due to the sheer lack of viable alternatives.

[5] This investigative journalism piece from the Wall Street Journal focuses on the practices of one company (Near Intelligence), who have been troubled with various ethical scandals over the years. The article outlines a culture of not acting on internal ethical/privacy warnings, and includes internal communications from the organisation noting: "We sell geolocation data for which we do not have consent to do so...we sell/share device ID data for which we do not have consent to do so [and] we sell data outside the EU for which we do not have consent to do so."

2.3 What is the Aotearoa New Zealand context surrounding the use of such data?

Concerns around privacy and the appropriate use of personal location data vary depending on the cultural context that the data relates to. Evidence indicates a complex range of perspectives on geoprivacy between Western and Eastern nations, and points to diversity between various Western nations as well (Zhang & McKenzie, 2023).

This subsection seeks to identify any information particularly relevant to the use of individual geolocation data in the specific context of Aotearoa New Zealand.[6] While all of the considerations raised in Section 2.2 apply in Aotearoa, two specific perspectives are worth touching on given the foundation of Aotearoa on Te Tiriti o Waitangi: the legal context for the use of personal data that is set out by The Crown, and the perspectives on the appropriate use of data arising from Te Ao Māori. Neither of these positions will be comprehensively covered, but particularly relevant points will be surfaced.

Insight 8: The Privacy Act (2020) raises considerations for the use of geolocation app data

The Privacy Act 2020 ['The Act'] governs how information is collected, stored, and used by individuals and organisations in Aotearoa. The Act therefore covers the use of data by researchers in Aotearoa as well. While a comprehensive legal review of The Act is outside of the scope of this Literature Analysis, a summary of the relevant principles within still sets important context for research consideration.

The Act contains thirteen principles:

- Principle 1 - Purpose for collection
- Principle 2 - Source of information - collection from the individual
- Principle 3 - What to tell the individual about collection
- Principle 4 - Manner of collection
- Principle 5 - Storage and security of information
- Principle 6 - Providing people access to their information
- Principle 7 - Correction of personal information
- Principle 8 - Ensure accuracy before using information
- Principle 9 - Limits on retention of personal information
- Principle 10 - Use of personal information
- Principle 11 - Disclosing personal information
- Principle 12 - Disclosure outside New Zealand

[6] See also Insight 10 in Section 2.4, which compares the approach to data protection taken by the New Zealand Integrated Data Infrastructure (IDI) with common practice in the commercial geolocation app data industry.

- Principle 13 - Unique identifiers

While all are applicable to the use of purchased mobile app data geolocation datasets, Principles 5, 6, 8, 9, and 10 have particularly relevant passages.

Of these, Principle 10 likely poses the least concern for the potential use of mobile geolocation data. The Principle provides limits on the use of personal information held by agencies. Importantly, relevant requirements in the Principle include that:

> "(1) An agency that holds personal information that was obtained in connection with one purpose may not use the information for any other purpose unless the agency believes, on reasonable grounds, - …
>
> (b) that the information—
>
> (i) is to be used in a form in which the individual concerned is not identified; or
>
> (ii) is to be used for statistical or research purposes and will not be published in a form that could reasonably be expected to identify the individual concerned; or …
>
> (f) that the use of the information for that other purpose is necessary to prevent or lessen a serious threat to—
>
> (i) public health or public safety;" (Privacy Act, 2020; Section 22).

Given the lack of transparency behind exactly *how* most commercially sold data is collected, there is a possibility that the data was collected for a different purpose than research. The subclause (f)(i) above arguably applies to past COVID-19 research, though justification for its application well after active response to the pandemic has ended is less strong. However, given (b)(ii) above, it is likely that commercially purchased individual geospatial data could still meet Principle 10 to be *legally* (though not necessarily *ethically*) used for research if no information is shared that could be used to identify individuals.

However, other principles potentially raise more substantial considerations for the use of geolocation app data. Principle 5 requires that agencies holding personal information securely store and manage access to that information to prevent loss, general misuse, and access, use, modification, or disclosure that is not approved by the agency. Researchers must therefore ensure that there are rigorous processes in place to prevent misuse of data in all forms, as well as unapproved access to the data (even if that unapproved use would be for public good purposes).

Principle 6 requires that agencies grant access to an individual's personal information held by them on request, as well as the right for individuals to request correction to that information. It would be difficult, if not impossible, for many researchers to enforce this Principle when using commercially procured datasets.

Principle 8 requires that "(1) An agency that holds personal information must not use or disclose that information without taking any steps that are, in the circumstances, reasonable to ensure that the information is accurate, up to date, complete, relevant, and not misleading." (Privacy Act, 2020; Section 22). While this Principle does leave some room for interpretation ('in the circumstances, reasonable'), it does clearly place responsibility on the

shoulders of researchers to ensure that the data accurately represents New Zealanders prior to the use of it. Given known inaccuracies inherent to geolocation app datasets (e.g. see Section 2.1), achieving confidence in the accuracy of the dataset may be a substantial undertaking.

Principle 9 requires that "(1) An agency that holds personal information must not keep that information for longer than is required for the purposes for which the information may lawfully be used." (Privacy Act, 2020; Section 22). While analysis for public good would likely meet this requirement (Principle 10 discusses the limits on use of personal information), Principle 9 does raise the question of how long the dataset can justifiably be kept when the end date for when it is 'required' can be extremely ambiguous.

Legislation is rarely ever black and white in its applicability, and problems inevitably arise regarding how it could be used in certain contexts. This is where ethical nuance enters the conversation once again: as Chen noted in a New Zealand context: "most governments can already access this data if they want to – the barrier isn't legal, it is ethical" (2020).

Insight 9: Geolocation app data systems do not align with a te ao Māori perspective, as expressed in terms of Māori Data Sovereignty

Thatcher and Dalton (2022) describe the process of commodification of data[7] and how this serves to obscure the extractive nature of tech companies' interactions with the creators and contributors of data. The commodification of data is a process of *data capitalism*, whereby data that was previously held privately comes to be the asset of a company. Drawing from the work of David Harvey, Thatcher and Dalton describe *data colonialism* as the process of accumulation by dispossession (Harvey, 2004, as cited in Thatcher & Dalton, 2022).

This concept of data colonialism resonates strongly with the experience of Māori[8] with colonisation in Aotearoa, as expressed by Timutimu in his statement of claim on behalf of himself, and of Māori in the Digital Technologies Industry:

> "Māori must be allowed to exercise tino rangatiratanga over their own digital sovereignty, lest we face the dire consequence of also being colonised in the digital realm. A subset of digital sovereignty, is the critical importance and impact of data sovereignty." (2023: 2).

[7] That is, the abstraction of the everyday practices that lead to the creation of data (such as taking a selfie or buying a t-shirt) from their aggregation and sale on RTB platforms.

[8] Those readers outside of an Aotearoa New Zealand context should not disregard this discussion simply because it focuses on a Māori worldview specifically: a growing indigenous data sovereignty movement globally (e.g. see the Global Indigenous Data Alliance) indicates strong relevance in any context where indigenous people(s) reside. Even in non-indigenous contexts, the Māori Data Sovereignty Principles contain many lessons in good practice for the management and use of data.

Indigenous data sovereignty, in turn, is a growing field of discussion in Aotearoa and globally. Kukutai and Cormack articulate the purpose of Māori Data Sovereignty as being "fundamentally about Māori control of Māori data to advance Māori self-determination." (2022: 123). In their Māori Data Sovereignty assessment of the NZ COVID-19 Contact Tracer app, Sterling et al. concisely summarise the importance of strong Māori input into the governance and use of Māori data to meet Te Tiriti obligations:

> "While justifiable in a public health emergency, digital contact tracing is nevertheless one manifestation of state surveillance and marginalised populations with a history of distrust of government cannot be assumed to willingly participate [4, 5]. In Aotearoa, the Indigenous Māori people have a long history of governments using their data and information against them [6]. Māori consistently report the lowest level of trust in government institutions in surveys such as the General Social Survey [7]. Māori disadvantage is structural and systemic and includes, among other things, comparatively higher rates of poverty [8], lower life expectancy [9], higher incarceration [10] and poorer health outcomes [11]. The impacts of the pandemic were heightened for Māori [12] who had more than twice the risk of death compared with European and Other groups [13]. … Article 2 of te Tiriti guarantees Māori tino rangatiratanga [sovereignty] over their lands and taonga [treasures] which, in a modern context, includes data [19–21]." (2024: 2)

The Principles of Māori Data Sovereignty published by Te Mana Raraunga (2018) are a useful and appropriate framework through which to assess the implications of the use of geolocation app data for Māori (for a more comprehensive example of a Māori Data Governance assessment of a digital tool, see Sterling et al., 2024). Table 2 summarises these principles, and sets out an assessment of geolocation app data across all aspects of Māori Data Sovereignty. In short, the data capitalist and data colonialist systems and practices underlying geolocation app data appear to be almost completely incompatible with a Māori world view.

Table 2: Māori Data Sovereignty Principles and their application to the use of geolocation app data. Information on the principles and sub-principles sourced from Te Mana Raraunga (2018).

Māori Data Sovereignty sub-principle; descriptions by principle	Assessment of geolocation app data by sub-principle and principle
Rangatiratanga 1.1: Control. Māori have an inherent right to exercise control over Māori data and Māori data ecosystems. This right includes, but is not limited to, the creation, collection, access, analysis, interpretation, management, security, dissemination, use and reuse of Māori data. 1.2: Jurisdiction. Decisions about the physical and virtual storage of Māori data shall enhance control for current and future generations. Whenever possible, Māori data shall be stored in Aotearoa New Zealand. 1.3: Self-determination. Māori have the right to data that is relevant and empowers	**Rangatiratanga** 1.1: Māori have no practicable ability to exercise control over any of the phases outlined in this principle/expectation. This is highlighted in the discussion of consent – the bare minimum level of control – in Section 2.2 Insight 7 of this Literature Analysis. 1.2: There are significant issues of jurisdiction with the mobility data reviewed here. The RTB market through which much geolocation app data is traded is distributed by definition, and jurisdiction is weakly enforced. Such data is largely not stored in Aotearoa New Zealand. 1.3: There is some *potential* for geolocation app data to support self-determination. If, for example, Post-Settlement Governance Entities had access to such datasets during the time of COVID-19 lockdowns, this

Māori Data Sovereignty sub-principle; descriptions by principle	Assessment of geolocation app data by sub-principle and principle
sustainable self-determination and effective self-governance.	could have supported efforts to enforce iwi "border guards". However the high barrier to entry for using such data (both in terms of cost and key digital skills) is prohibitive for indigenous communities.
Whakapapa 2.1: Context. All data has a whakapapa (genealogy). Accurate metadata should, at minimum, provide information about the provenance of the data, the purpose(s) for its collection, the context of its collection, and the parties involved. 2.2: Data disaggregation. The ability to disaggregate Māori data increases its relevance for Māori communities and iwi. Māori data shall be collected and coded using categories that prioritise Māori needs and aspirations. 2.3: Future use. Current decision-making over data can have long-term consequences, good and bad, for future generations of Māori. A key goal of Māori data governance should be to protect against future harm.	**Whakapapa** 2.1: In the context of this Literature Analysis, mobility data is effectively a type of metadata, because it consists of data collected incidentally to the primary purpose of targeting advertisements. By its nature, this metadata tends to be poorly supported by its own metadata. There is very little transparency and guidance for users around the provenance and purpose of collection, and therefore appropriate use. 2.2: The mechanisms for imputing demographic characteristics, including age, gender, and ethnicity are not transparent, nor do they make use of methods that are they likely to be accurate or consented to (see for example Athey et al., 2021). As noted in Section 2.1 of this document, the Human Rights Watch (2020) describes the tendency for minority groups, including indigenous peoples, to be excluded from such systems. 2.3: Following on from the poor *jurisdiction* sub-principle described above, and as illustrated in the example set out in Insight 12, there are very few governance and control mechanisms in place for geolocation app data that would prevent future harm to Māori.
Whanaungatanga 3.1: Balancing rights. Individuals' rights (including privacy rights), risks and benefits in relation to data need to be balanced with those of the groups of which they are a part. In some contexts, collective Māori rights will prevail over those of individuals. 3.2: Accountabilities. Individuals and organisations responsible for the creation, collection, analysis, management, access, security or dissemination of Māori data are accountable to the communities, groups and individuals from whom the data derive.	**Whanaungatanga** 3.1: There is no evidence of engagement with representatives of collective rights holders in the collection and use of GPS mobile data. As noted by Kukutai et al. (2023), the Privacy Act principles (as noted in Insight 8) do not acknowledge "the gaps in existing data privacy approaches with regards to Indigenous data" (2023: 7). Until such recommendations from Kukutai et al. (2023) are part of the privacy regulation framework in Aotearoa, there is no legal mechanism to protect collective privacy concerns. 3.2: There is no evidence that most geolocation app data providers have acknowledged meaningful accountability to the public in general, nor to specific communities such as indigenous communities.
Kotahitanga 4.1: Benefit. Data ecosystems shall be designed and function in ways that enable Māori to derive individual and collective benefit. 4.2: Build capacity. Māori Data Sovereignty requires the development of a Māori	**Kotahitanga** 4.1: While it is possible that Māori may derive individual and collective benefit from geolocation app data, the surrounding data ecosystems have been developed in ways that privilege technology and capital, and so do not meet the expectation that systems are designed with

Māori Data Sovereignty sub-principle; descriptions by principle	Assessment of geolocation app data by sub-principle and principle
workforce to enable the creation, collection, management, security, governance and application of data. 4.3: Connect. Connections between Māori and other Indigenous peoples shall be supported to enable the sharing of strategies, resources and ideas in relation to data, and the attainment of common goals.	such benefit in mind. See the issues raised by Oliver et al. (2020) covered in Section 2.2, Insight 6. 4.2 and 4.3: Meeting this expectation would require that organisations dealing in geolocation app data invest in indigenous workforce development, and invest in networks between indigenous communities in every jurisdiction in which they operate. Timutimu's (2023) statement of claim under Te Tiriti o Waitangi clearly articulates systemic failures in this regard across the digital system in Aotearoa, while the discussion in Section 2.1 outlines several ways in which the geolocation app data system is not inclusive for Māori.
Manaakitanga 5.1: Respect. The collection, use and interpretation of data shall uphold the dignity of Māori communities, groups and individuals. Data analysis that stigmatises or blames Māori can result in collective and individual harm and should be actively avoided. 5.2: Consent. Free, prior and informed consent [FPIC] (United Nations Department of Economic and Social Affairs, as cited in Te Mana Raraunga, 2018) shall underpin the collection and use of all data from or about Māori. Less defined types of consent shall be balanced by stronger governance arrangements.	**Manaakitanga** 5.1: There is little to no control over the uses to which geolocation app data is put (see Section 2.2 Insight 4 or Section 2.4 Insight 12 for examples), including uses which directly contravene this principle. 5.2: Geolocation app data fails on all reasonable measures of FPIC. As discussed in Section 2.2 Insight 7, consent is not well defined, and is certainly not informed. This weak, ill-defined consent is not accompanied by any strong governance arrangements.
Kaitiakitanga 6.1: Guardianship. Māori data shall be stored and transferred in such a way that it enables and reinforces the capacity of Māori to exercise kaitiakitanga over Māori data. 6.2: Ethics. Tikanga, kawa (protocols) and mātauranga (knowledge) shall underpin the protection, access and use of Māori data. 6.3: Restrictions. Māori shall decide which Māori data shall be controlled (tapu) or open (noa) access	**Kaitiakitanga** 6.1, 6.2, and 6.3: As noted elsewhere in this Literature Analysis, the geolocation app data system is largely governed by the RTB market (see Section 2.4 Insight 12 for examples of how some providers have been able to operate in this environment). As such, Māori have no mechanism for control over the collection, storage, or transfer of geolocation app data. There is no reference to, nor is there any mechanism for, the application of tikanga Māori, kawa Māori, or mātauranga Māori at any stage in the geolocation app data system. Māori have no mechanism for control over the allocation of tapu or noa to geolocation app data.

This poor assessment of geolocation app data in terms of Māori Data Sovereignty highlights that the continued presence of such data in the Aotearoa digital and research landscape is indicative of a lack of robust legislation and policy in Aotearoa, and therefore of the Crown's continuing failure "to meet its obligations and duties to Māori in the digital and technology space" (Timutimu, 2023: 2).

Sterling et al. (2024) conduct a Māori Data Sovereignty assessment of the NZ COVID Tracer App. While that app did not include geolocation app data as discussed in this Literature

Analysis (relying instead on less invasive collection methods of self-report QR Code scanning and Bluetooth data, which are only stored locally), the authors nonetheless noted that "Unfortunately, [the COVID-19 Contact Tracer App] was designed and deployed with minimal Māori input" (Sterling et al., 2024: 2). Their assessment ultimately found:

> "significant room for improvement for future digital public health interventions. … systemic issues in the public health sector create challenges for fully incorporating Māori data governance into intervention design processes, especially when operating with significant constraints (e.g. time). Partnership with Māori is often not considered or considered too late (1.2, 2.1), which can then have significant impacts on how the system is designed and whether or not they are built with Māori in mind (6.2, 6.3). While there is now better recognition of the importance of Te Tiriti and partnership with Māori in the public sector, there are still significant systemic barriers to overcome and correct, where outcomes are just as important as intent." (2024: 7)

These conclusions apply even more strongly to the use of geolocation app data, due to the granularity of the data and central data aggregation. While the COVID-19 Contact Tracer App in Aotearoa may have had 'minimal Māori input' into its design, most geolocation app data and tools derived from it have had none at all. Although the conclusions of Sterling et al. (2024)[9] relate specifically to public health sector workers, they also apply to most if not all researchers operating in Aotearoa or drawing on Māori data.

Research Advice:

Any researchers who want to uphold and respect Te Tiriti o Waitangi should be wary of the use of commercial geolocation app data given the substantial Māori Data Sovereignty concerns associated with it.

2.4 What guidelines exist for the safe use of individual-level geolocation data?

The diversity of geolocation data types, collection points, and data use contexts mean that there are no extremely specific guidelines that will suit every possible combination of them. Indeed, some have noted that agreed upon approaches to navigating the potency-privacy tension for mobile phone data (such as geolocation app data) have largely been missing, which:

> "has left data protection authorities, mobile phone operators, and data users with little guidance on technically sound yet reasonable models for the privacy-conscientious use of mobile phone data" (de Montjoye, 2018: 2)

[9] The article contains extensive relevant discussion for this section, and is strongly recommended reading.

De Montjoye et al note that this has in turn "often resulted in suboptimal tradeoffs if any" (2018: 2). However, there are nonetheless some useful broader frameworks, principles, and guidance for the use of mobile geolocation data that have emerged in recent years.

Insight 10: Several guides exist for the safer use of geolocation app data

In some cases, these are less guidelines and more a set of broad recommendations. For example, Oliver et al.'s discussion of using mobile location data for disease response notes that:

> "any efforts should meet clear tests on the proportionate, legal, accountable, necessary, and ethical use of mobile phone data in the circumstances of the pandemic and seek to minimize the amount of information gathered to what is necessary to accomplish the objective concerned." (Oliver et al., 2020: 5).

While Gluckman likewise asserts that further big data analytics in Aotearoa New Zealand will:

> "require consistent and ongoing attention to transparency, social acceptance, strong data governance, a commitment to data hygiene, curation and quality, upskilling within the policy community, continued vigilance against poor data analysis and interpretation and algorithmic bias, as well as ongoing engagement of policy-makers, service providers, academics and the wider community" (Gluckman, 2017: 3).

Others simply make broad arguments for transparency and for the more complete provision of information about the collection and use of personal big data (including geolocation app data) to ensure adequate social license (e.g. Baron & Musolesi, 2020; Garattini et al., 2019; New Zealand Human Rights Commission, 2018; Zhang & McKenzie, 2023).

However, more explicit guidance does exist. For example, de Montjoye (2018) put forward four models for the use of mobile phone data in a manner that attempts to strike "a reasonable balance" in the privacy-potency tension (de Montjoye et al., 2018: 2). Their four models focus on the management and governance of the datasets (i.e. how they are accessed and shared, and how data is analysed), and include 1) limited release, 2) pre-computed indicators and synthetic data, 3) remote access, and 4) question-and-answer (for more detail on these models, see pp. 3-4 of de Montjoye et al., 2018).

The Human Rights Watch (2020) also provided relevant guidance in the form of a question and answer document focusing specifically on *Mobile Location Data and COVID-19*. The document outlined the minimum limits that should be put on technological responses to the pandemic in order to protect human rights. Among these, several are directly relevant to any potential ongoing analytical work, including that such efforts must:

- "Be time-bound and only continue for as long as necessary to address the pandemic"
- "Be transparent about any data-sharing agreements with other public or private sector entities"
- "Incorporate protections and safeguards against abusive surveillance and give people access to effective remedies"
- "Provide for free, active, and meaningful participation of relevant stakeholders in data collection efforts"

The Human Rights Watch further emphasises several of the above points, and also points to the need to pay specific heed to potential impacts on marginalised groups:

> "Even in times of emergency, when states restrict human rights for public health reasons, international human rights law says that … any curtailment of rights needs to ***take into consideration the disproportionate impact on specific populations or marginalized groups***." (Human Rights Watch, 2020, emphasis original).

While the Human Rights Watch provides some useful high-level guidance about what to do – and crucially, what *not* to do – even more detailed guidance does exist.

Recognising the unique nature of geolocation data, the EthicalGEO group released *The Locus Charter*, "a proposed set of common international principles to support ethical and responsible practice when using location data." (2021:1). An initiative of the American Geographical Society, EthicalGEO created *The Locus Charter* specifically "for individuals and organizations who use location data or have responsibility for activities that create, collect, analyze and store location data." (2021:1). Any researchers using geolocation app data clearly fall within the scope of the charter.

While already included once in this Literature Analysis, it is worth repeating EthicalGEO's emphasis that while there are similarities in ethical issues facing the use of data in general with the use of location data specifically, these issues are not completely the same:

> "Many of the sources of potential harm identified in other fields can occur in relation to location data, including bias in datasets, privacy intrusion, and misuse of power imbalances in markets. But while the examples are very useful, it is important to recognize that **risks relating to use of location data can have impacts that are specific to those uses, to people and to places**. To be followed around where you go in the world has parallels with being tracked around the internet, but it is not the same thing. Also, while a lot of data regulation is framed around the rights of individuals, use of location data can affect groups of people, including some already under pressures relating to where they are, like refugees and other migrants. **Location data lends specific powers, which can imply specific responsibilities**." (EthicalGEO, 2021:3, bold added).

With this in mind, *The Locus Charter* provides 10 foundational principles to apply when using geolocation data. Table 3 on the following page summarises those principles and provides detail (copied from the charter itself) on how they may be implemented in practice.

The Locus Charter (EthicalGEO, 2021) and The Human Rights Watch (2020) provide fairly explicit and clear cut guidance on how to most appropriately use location data, including in the context of disease disaster analysis specifically. Following this guidance will help to protect the rights and safety of both individuals and groups to ensure that the benefits of data use outweigh any potential negative consequences.

Table 3: The 10 principles for ethical location data use. Adapted from EthicalGEO (2021), pages 5-6.

Principle	Description
1) Realise opportunities	"Location data offers many social and economic benefits, and these opportunities should be realized responsibly."
2) Understand impacts	"Users of location data have responsibility to understand the potential effects of their uses of data, including knowing who (individuals and groups) and what could be affected, and how. That understanding should be used to make informed and proportionate decisions, and to minimize negative impacts."
3) Do no harm	"Physical proximity amplifies the potential harms that can befall people, flora and fauna. Data users should ensure that the individual or collective location data pertaining to all species should not be used to discriminate, exploit or harm. Rights established in the physical world must be protected in digital contexts and interactions."
4) Protect the vulnerable	"Vulnerable people and places can be disproportionately harmed by the misuses of location data, and may lack the capacity to protect themselves. In these contexts, data users should take additional care, act proportionately, and positively avoid causing harm."
5) Address bias	"Bias in the collection, use, and combination of location datasets can either remove affected groups from mapping that conveys rights or services, or amplify negative impacts of inclusion in a dataset. Therefore care should be taken to understand bias in the datasets and avoid discriminatory outcomes."
6) Minimise intrusion	"Given the intimate and personal nature of location data, users should avoid unnecessary and intrusive examination of people's lives and the places they live in, that would undermine human dignity."
7) Minimise data	"Most business and mission applications do not require the most invasive scale of location tracking available in order to provide the intended level of service. Users should comply with practices that adhere to the data minimization principle of using only the necessary personal data that is adequate, relevant and limited to the objective, including abstracting location data to the least invasive scale feasible for the application."
8) Protect privacy	"Tracking the movement of individuals through space and time gives insights into the most intimate aspects of their lives. In the rare cases when aggregated and anonymized location data will not meet the specific business or mission need, location data that identifies individuals should be respected, protected, and used with informed consent where possible and proportionate."
9) Prevent identification of individuals	"As an individual's mobile location data is situated within more and more geospatial context data, its anonymity erodes, measures should be put in place to prevent subsequent use of the data resulting in identification of individuals or their location."
10) Provide accountability	"People who are represented in location data collected, combined and, used by organizations should be able to interrogate how it is collected and used in relation to them and their interests, and appeal those uses proportionate to levels of detail and potential for harms."

However, while the guiding principles of *The Locus Charter* are relatively unambiguous, the broad purpose of the charter means it does not provide context-specific information on how best to enact them. One method of filling this gap is to look to how other, similar datasets are protected or managed, and using reflections on other accepted practices to inform future practice. The following subsection will discuss this possibility in further depth. [10]

Insight 11: Comparably invasive datasets in Aotearoa are subject to strict data protections

One obvious well-documented comparison stands out for the use of highly personal individual data in Aotearoa; the Integrated Data Infrastructure [IDI]. Other private corporate datasets of similar invasiveness no doubt exist in Aotearoa, but are unlikely to follow best ethical practice and are also unlikely to publicly release a great deal of information about their dataset or data management practices.

The IDI – run by Stats NZ – is a large database containing individual level microdata about people and households in Aotearoa New Zealand. While the data is de-identified, the breadth and depth of information within is more than sufficient to identify the vast majority of individuals in the database. In these respects, the data is fairly similar to many of the geolocation app datasets provided by private commercial organisations.

However, unlike these commercial datasets, the IDI has several safeguards in place to prevent misuse of the data wherever possible; the de-identification of the data is only a starting point. In order to access data, researchers must be vetted and approved, including going through confidentiality training. The approach that Stats NZ takes to managing the IDI is directed by *The Five Safes Framework*, which aims to ensure safe people, projects, settings, data, and output (Stats NZ, 2022). See Table 4 on the subsequent page for more detail on how Stats NZ enacts the Five Safes through specific safeguards.

Moreover, researchers using the IDI should adhere to Ngā Tikanga Paihere, "a framework intended to help researchers engage with Māori and other communities to ensure the use of microdata is respectful, ethical, and culturally appropriate" (Stats NZ, 2023). See Figure 3 (immediately following Table 4) for an overview of the Ngā Tikanga Paihere framework.

[10] Other guidance and frameworks do exist, though none are as specific to location data in particular as those covered here. Other potential guidance for the appropriate use of data include Māori Data Sovereignty principles (covered in Insight 9), the Privacy, Human Rights, and Ethics (PHRaE) Framework used by New Zealand's Ministry for Social Development, the CARE or OCAP principles for indigenous data governance, or the FAIR principles for data use. The GSMA (2014) released a set of guidelines for the permittable use of mobile call record data in the Ebola pandemic response. Many other similar examples exist. All of them tend to emphasise appropriate privacy, transparency, and accountability in the collection, management, and use of data.

Table 4: The Five Safes Principles, and the specific safeguards and rules in place around them. Adapted from Stats NZ, 2023.

Principle	Safeguards and rules in place
Safe People: Researchers are vetted and must commit to use data safely before they can access the data	Pass referee checks.
	Attend confidentiality training.
	Sign a confidentiality certificate under the Data and Statistics Act 2022. The certificate is a lifetime commitment to keep the data confidential.
	Sign a contract where they agree to follow our rules and protocols.
	Have capability to use the data.
	Researchers who break our protocols can be banned, blacklisted, or prosecuted.
Safe Projects: To gain access to integrated data, researchers must have a project they can demonstrate is in the public interest	Research projects must focus on finding insights and solutions to issues that are likely to have a wide public benefit.
	The IDI and LBD cannot be used for individual case management, such as making decisions about a specific person or family.
Safe settings: A range of privacy and security arrangements keep data safe.	Data can only be accessed through a secure virtual environment known as the Data Lab, and only in research facilities approved by Stats NZ. A variety of security layers further protect the information (next three boxes).
	The IDI and LBD sit on a separate server that is not connected to the internet.
	Computers are not connected to a network.
	There are no USB ports or printing facilities so users cannot take information in or out of the Data Lab without it being checked first by Stats NZ staff.
Safe Data: Identity is protected. Data has had identifying information removed, and researchers only get access to the data they need.	Data that is available to researchers is de-identified. This means information like names, dates of birth, and addresses has been removed. Numbers that can be used to identify people, like IRD and NHI numbers, are encrypted (replaced with another number).
	Researchers only get access to the data they need for their specific research project. For example, a researcher granted access to the IDI will only have access to the specific datasets they need for their research project: they cannot see all information in the IDI.
Safe Output: All information is checked to ensure it does not contain any identifying results	Researchers must confidentialise their results. We then check all outputs before they can be released from the Data Lab, to ensure information is grouped in a way that makes it impossible to identify individuals. Results that could potentially identify individuals will not be released.
	The Microdata output guide describes the methods and rules that researchers must use to confidentialise output produced from Stats NZ's microdata.

Figure 3: Ngā Tikanga Paihere framework for engaging with Māori and other communities regarding the safe use of individual data. Sourced from: https://data.govt.nz/toolkit/data-ethics/nga-tikanga-paihere/

It is worth noting that even with these frameworks the use of the IDI is neither universally ethically acceptable nor completely effective at protecting privacy. Ngā Tikanga Paihere provides excellent guidance, but its use is not mandatory. Researchers have highlighted that there is much work to do to make the IDI better for Māori and to align with Māori Data Sovereignty, with some also arguing that abolishing the platform and starting over may be appropriate (Greaves et al., 2024). Data breaches happen increasingly frequently as the IDI is used more often, ranging from individuals granting data lab access to non-approved colleagues through to Stats NZ temporarily loading health data without consent from the individuals it represents (Mitchell, 2023).

Moreover, attitudes towards the integration of data vary notably between individuals and communities – Davison et al. (2015) conducted research for Stats NZ into public attitudes on data integration, finding the following:

> "Statistical data integration was more acceptable if the data is completely de-personalised and anonymous. **People tended to feel that data integration is more acceptable if the public is informed about what is happening and why**. If personal,

> sensitive, or complex data is going to be integrated some participants felt that people should have the right to give or withdraw their informed consent. Data integration tended to be seen as relatively unacceptable if there was no demonstrable need or purpose. People also tended to feel that **integration would be unacceptable if it could be misused**, or resulted in harmful or unfair outcomes. Examples of such outcomes include: if integrated data was used for direct commercial gain; to take advantage of vulnerable people; or for profiling, stereotyping, or disadvantaging certain groups or types of people, or people from particular places. **People felt uncomfortable if poor quality data was used in a misrepresentative way, and some were unhappy about data being used for spying and surveillance.**" (Davison et al., 2015: i, bold added)

Ultimately however – regardless of questions around how effective Stats NZ may be in its protection of individual privacy and safety in the IDI – it is undeniable that the safeguards surround the access, use, and publication of IDI data are more transparent, robust, and aligned with ethical practice than the largely non-existent restrictions on the use of commercial mobile datasets. The safeguards that data capitalist organisations selling geolocation app data set around the use of highly personal datasets seem to amount to: buy our data, try not to break local laws, and please don't compromise our ability to continue making money from this in the future.

Given the concerns raised by members of the public in Aotearoa around the integration and use of personal data, it is likely fair to say that many would not support the use of private, for-profit commercial datasets that contain inferred individual personal data and paired geolocation data.

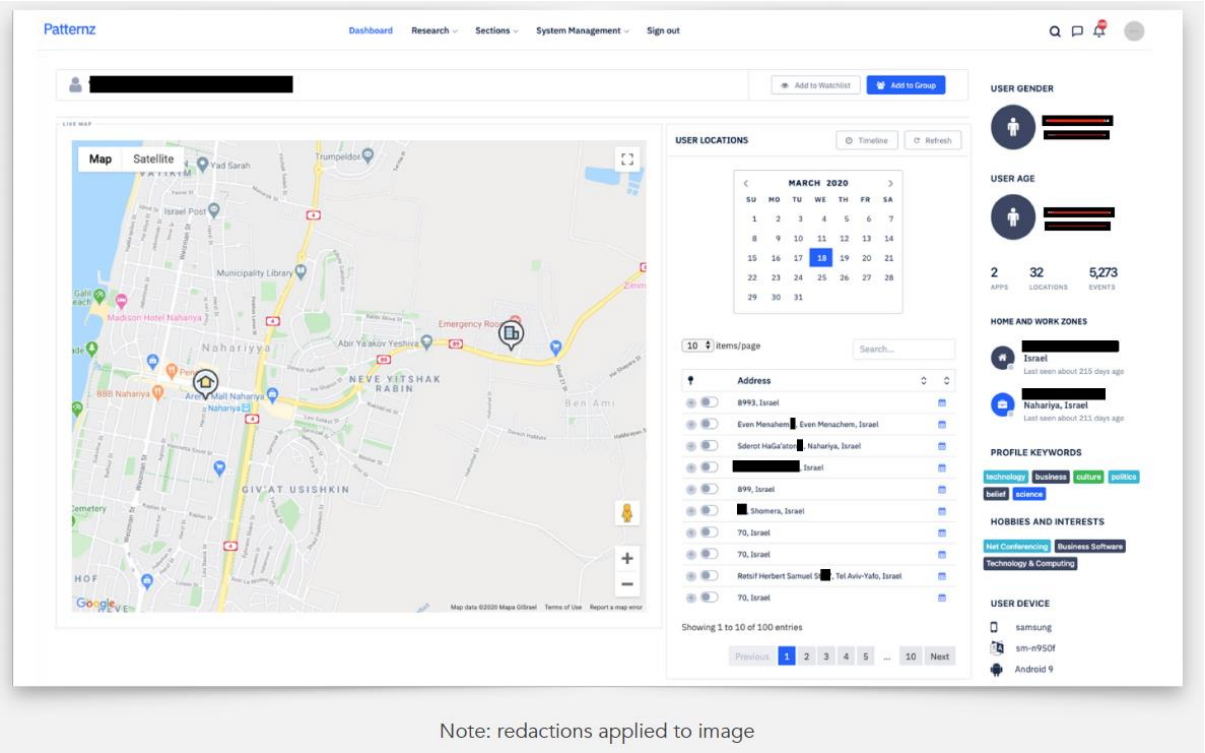
Insight 12: The use of raw data collected in the geolocation app data industry is near incompatible with ethical practice

Rather than looking to frameworks or government organisations for guidance, what can we learn from the approaches of various organisations operating in the data capitalism industry about the appropriate use of GPS mobility data?

Examining both the RTB ad data exchanges that sell geolocation app data as well as the organisations that buy from them, two reports from the Irish Council for Civil Liberties identify the use of RTB data by various commercial organisations to create datasets for the tracking of individuals (Ryan & Christl, 2023a; 2023b). The authors identify issues inherent to RTB data itself, such as "Cambridge Analytica style psychological profiling of target individuals' movements, financial problems, mental health problems and vulnerabilities, including if they are likely survivors of sexual abuse" (Ryan & Christl, 2023a; 2023b: p.4, both reports). Ryan and Christl describe RTB data exchanges as "a torrent of blackmail data" (2023a; 2023b: 15, both reports).

The reports then focus specifically on one tool created from RTB data, called 'Patternz'. This tool claims to have data profiles on 5 billion people, including their movement patterns over the course of several months, who they have met, their hobbies and interests, their demographics, and more (Ryan & Christl, 2023a; 2023b). Patternz can also be used to identify the co-workers and children of a specific target. See Figure 4 for an example screenshot of the Patternz tool, sourced from the reports. Given the 5 billion profiled individuals in Patternz

and the fact that developed nations (where smartphones are more ubiquitous) are likely overrepresented, it is probable that you – the reader – have a page in the Patternz tool as well.



Note: redactions applied to image

Figure 4: A screenshot of the 'Patternz' tool sold by ISA, a private company. Sourced from Ryan & Christl, 2023a; 2023b. The source notes that "Patternz provides a targeted person's current location, historical movements over several months, and who they frequently meet. ISA says Patternz can identify a target's children, co-workers, and their "driving path"" (Ryan & Christl, 2023a; 2023b: 13).

Patternz is not alone: the two reports from the Irish Council for Civil Liberties provide several examples of the use of RTB data by various other data capitalist organisations as well, such as Near Intelligence. Near Intelligence (now known as Azira) are also discussed in Tau et al. (2023), and Wyden (2024) for their illegal selling of data and its use for targeting abortion clinic visitors respectively, while data from SafeGraph – another provider – is similarly discussed in relation to abortion clinic ad targeting in Cox (2022). Earlier, Cox surfaced the US military's purchase of Muslim dating app and prayer app geolocation data from the companies X-Mode and Babel Street, noting that:

> "A former Babel Street employee described … how users of the product can draw a shape on a map, see all devices Babel Street has data on in that location, and then follow a specific device around to see where else it has been… the source said 'we could absolutely deanonymize a person.' Babel Street employees would 'play with it, to be honest,' the former employee added." (Cox, 2020)

Valentino-DeVries et al. (2018) further provides several real-life examples of just how easily individuals can be tracked using geolocation app datasets. The examples here are, however, only the ones that we know about: it is likely that many ethically questionable practices take place behind closed doors that the public is as yet unaware of.

Indeed, several sources have identified practices similar to ethics 'greenwashing' among the companies that sell mobile location data, often described as 'ethics washing' or 'privacy washing' (Bietti, 2020; Thatcher & Dalton, 2022). Thatcher and Dalton concisely summarise such privacy washing – and the data capitalism sources of it – as follows:

> "Read at face value, technology firms' press releases make the case that their top priority is users' privacy and that it has never been easier for users to control how their data are collected and shared. Of course, the seemingly endless succession of data scandals (Bishop 2018), threats of regulations and fines (Information Commissioner's Office 2020), and antitrust and class action lawsuits (Georgiadis and Beioley 2021) suggest that other factors are at play in these promises. Whatever the motivation, some firms are beginning to allow individual users some control and intentionality in how each user's data are collected and shared. Examples include app-specific access control to GPS information and the amount of time before location history data are deleted from a Google account (Morrison 2020).
>
> However, not only does such language deceptively shift responsibility towards the individual, it also does nothing to alter the underlying business model and profit incentives on which these firms rest. **Data capitalism is built upon the creation, extraction, appropriation, analysis, and trade of data. Privacy-washing provides the appearance of conscientious responsibility to ensure the continued profitability of this strategy**." (Thatcher & Dalton, 2022: 67, bold added)

In other words, so long as the profit motive remains, companies participating in data capitalism will do their best to appear ethical while still extracting as much valuable data as possible. Bietti similarly discusses ethics washing among tech companies, arguing that they see philosophy and ethics as "a communications strategy and as a form of instrumentalized cover-up or façade for unethical behavior" (2020: 210). Bietti also identifies current self-regulatory efforts at ethical behaviour within the tech industry as highly problematic, and notes the act of 'ethics bashing', wherein tech companies actively push back against ethics and moral philosophical discussions in regard to their behaviour "as mere "ivory tower" intellectualization of complex problems that need to be dealt with in practice" (2020: 210).

Taking all of the above into account, it is therefore unlikely that any guaranteed ethical dataset exists stemming from data capitalism – at least with individual-level data granularity. Indeed, researchers should be highly sceptical of any supposedly powerful source of data that they are able to acquire from a commercial provider for free or for a much lower price than the multiple millions it could theoretically be worth. However, (to misuse several turns of phrase) it is difficult to put a cat back in a bag or to unspill some beans; these datasets – no matter how ethically problematic – exist, and we must decide how to interact with them. Short of large-scale regulatory reform, the onus therefore falls on the potential end users of the data to either:

1. **Boycott:** Avoiding using the data altogether.
2. **Do the best you can:** Establish and follow robust (potentially expensive and time-consuming) ethical and privacy-protecting processes in your approach to using the data.

3. **Do business as usual:** Act as though the data is not problematic, and proceed to use it unhindered.

Options 1 and 2 are the preferred choices. Under option 1 it is certainly possible to explore alternatives to data capitalist sources of geolocation app data; designing a bespoke, more ethical geolocation data collection method is one such approach that avoids engaging with existing geolocation app data systems altogether. These three options are relatively essentialising: nuance of course exists. Section 3 of this document will discuss decision-making regarding the ethical use of such data in greater detail.

There are a vast number of companies operating in this data capitalism space, and some undoubtedly follow practices that genuinely attempt to be as ethical as possible. Further market research could determine if this is the case (looking for transparent practices is a good first step), but by nature any more ethical company in this space will likely sell data that comes pre-aggregated, and which robustly protects the privacy of individuals and groups. This greater privacy will generally come at the cost of the overall potency of the data[11].

[11] Some exceptions may apply here. For example, companies may exist that possess individualised geolocation data internally (subject to strict data protections), but who provide customised aggregated data suitable for researchers' needs. However, such cases would still have the same technical limitations as other geolocation app data, and would still be vulnerable to data breaches or internal data misuse – so are not necessarily a perfect solution.

3 Concluding Comments

So, do the ends justify the means when it comes to using currently commercially available geolocation app data? The answer – as with many questions in the social sciences – is "it's complicated". Ultimately, however, this report argues that the ends *do not* justify the means. In this final section, it is worth briefly discussing the 'ends' and the 'means' separately.

The ends

Returning to the discussion from Thatcher and Dalton (2022) that opened Section 2, we are cautioned against presenting the use of big data as either solely dystopian or utopian. Geospatial mobility data can be used for good purposes, bringing benefits to individuals, to groups, and to the environment. It can also be a tool for oppression, blackmail, physical abuse, widespread harm, and the erosion of democracy.

These are the 'ends', and they can be competing. One research project may simultaneously bring benefits for public health while also normalising and legitimising state oppression. Another project might seek to provide more tailored services to people in need while also enabling a malicious actor on the team to gain insights into their friends or family that those individuals would have preferred to remain private.

This Literature Analysis does not identify a moral hard and fast rule for what 'ends' are good enough to outweigh the potential for harm. Many researchers will argue that the ends they are trying to achieve are of the utmost importance, but using individual geospatial data will require genuine (and sometimes harsh) critical self-reflection regarding how valuable one's own research truly is.

An extension of this consideration is honesty regarding the pathway to implementation and impact; researchers should reflect on how likely they are to be able to effectively communicate their research in a manner that ensures uptake. In other words, the beneficial 'ends' of a project are only 'ends' if they actually happen. This applies to all research that seeks to have a positive impact, regardless of the specific data (or 'means') used.

The means

This Literature Analysis was focused on the possible use of individual-level GPS data (coupled with inferred demographic information) gained from mobile app advertisements. As 'means', this type of data is highly problematic for several reasons:

- Early in the value chain, we know that RTB-sourced data itself (where many geolocation app datasets come from) has few protections and in many cases may violate relevant local privacy laws where it is collected.
- It is reasonable to claim that very little (if any) geolocation app data was collected under truly 'informed' consent.
- The data is easily used by malicious actors, and there are myriad examples of it being applied to conduct surveillance in a way that many would consider to be highly unethical.

- Those organisations slightly later in the value chain are frequently no less problematic in their practices, with ample evidence of companies knowingly selling illegal data or enabling ethically questionable uses of their products.

The degrees of separation between researchers using the data and the organisations collecting it can imply that the researchers are absolved of responsibility through purchasing the data. Thinking in this manner should be avoided – you still hold responsibility.

Ultimately, data capitalism (and relatedly surveillance capitalism) sits at the core of the problems with commercially available geolocation app data. Many data corporations will generally do what is most profitable, which involves selling data to the highest bidder with as few costly regulations as possible. In addition, the accuracy or correctness of the data is rarely a concern for these organisations; veracity is secondary to profitability. Given this motive, *researchers therefore run a risk of getting data that is neither potent nor private*: a lose-lose situation for all but the companies selling the product.

Fortunately, more ethical and technically robust means do exist, though these means are often more time and resource intensive. For instance, some researchers have created dedicated research apps that collect geolocation (and other) data directly from individuals, with clear informed consent processes and with less profit motive (e.g. see Bähr et al., 2022). However, such apps will have a much lower number of observations than RTB sourced data, and are not free of the potential accuracy issues that are associated with all mobile phone GPS measurements.

Do the ends justify the means?

While this discussion is relevant to many uses of geolocation app data (e.g. transport modelling, military research, etc.), turning to the case of disease disaster research specifically provides ample literature discussing what 'ends' might justify more invasive geospatial mobility data analysis 'means'.

Speaking from an Aotearoa New Zealand context, Chen notes that government analysis is likely *legally* permissible with geolocation app data, but is *ethically* questionable ("the barrier isn't legal, it is ethical"). This report echoes his sentiment that: "key to finding the right balance is proportionality – that the actions taken are proportional to the causes and outcomes, or in other words ensuring that the ends justify the means" (Chen, 2020).

While most agree that geolocation mobile data can be useful throughout the disaster cycle's four stages (prevention, preparedness, response, recovery), many argue that the required sacrifices of civil liberty for the use of *individual-level* data are only excusable during the response phase, when data analysis can have the greatest impact and when rapid action is most urgent (Cinnamon et al., 2016; Human Rights Watch, 2020; Oliver et al., 2020). Even then, some researchers assert that individual data should not be used, with only aggregate data drawn on in order to ensure that data protection and privacy needs are protected (Buckee et al., 2020).

Based on the above evidence, the ends do not justify the means when it comes to the use of currently commercially available individual-level geolocation app data

Based on the content of this Literature Analysis, this report concludes that the ends do not justify the means– at least currently. This conclusion is not specific to pandemic preparedness modelling alone; most public good ends do not justify the commercial geolocation app data ends.

The view of this concluding section reflects elements of much of the literature, and is that individual geolocation data collected via mobile phone apps should only be used:

1. with strict, transparent, externally regulated privacy and human rights protections in place,
2. when there is a clear implementation pathway to positive impact, *and*
3. when no other data is available that would be adequate for the analysis with fewer invasions of privacy.

At present, these safeguards are not suitably in place for the use of such data to be justified, at least in the context of Aotearoa New Zealand. Moreover, the 'ends' of research using geolocation app data are frequently unclear in terms of their benefit, especially given questions around the overall accuracy of the data.

These conclusions echo those of the Human Rights Watch, who advocate not using privacy-invasive technologies for public good without high levels of scrutiny:

> "These technologies are intended for a praise-worthy purpose: protecting public health at a time of public emergency, a situation that can justify some restrictions on rights. But the long history of emergency measures shows that they often go too far, fail to meet their objectives, and once approved, often outlast their justification. No matter how compelling the situation, it is incumbent on public authorities and private actors to ensure that measures do not overstep the permitted legal restrictions on individual rights." (Human Rights Watch, 2020)

While particularly relevant for considering any future COVID-19 related research (especially this long after the immediate pandemic response), these considerations are of equal relevance to researchers considering the use of geolocation app data for other purposes as well. With the above Literature Analysis content in mind, the following subsection provides recommendations and raises considerations for researchers considering research with geolocation app data.

Recommendations and considerations for researchers

Ultimately, the question of whether the means justify the ends for many other research cases may come back to your ethics, moral philosophical standpoint[12], and a need for genuine self-reflection.

So, what should you do if you have read the above Literature Analysis and are, as of yet, unsure on what ethical viewpoint you hold regarding the appropriate use of geolocation app data? An initial recommendation for any researcher is to consider the following hypothetical (yet potentially very real) questions, and what your answers might mean for the use of such data:

- Would you be willing to purchase and use mobile phone app personal re-identifiable location data that may be collected illegally and without informed consent from the individuals to whom it relates?
- Would you be willing to purchase identifiable location data that is collected from tracking devices placed unknowingly on individuals?
- Have you ever considered not purchasing or using a product because of the practices of the organisation that produces it?
- Do you support tight regulation and control of tools that can be used for harm (e.g. guns, car licensing) even if not all people who use them intend to cause harm with them?
- To what extent are you willing to restrict the civil liberties of others to accomplish public good?
- Does informed consent matter to you always, or are there situations where it is okay to have individually identifiable people participate in research without full, prior, and informed consent?
- To what extent do you believe it is possible to engage ethically with a tool and an industry that perpetuate harm?
- How certain are you that your research will lead to public good outcomes if it goes ahead?
- How certain are you that your research will not lead to public bad outcomes if it goes ahead?
- How committed are you to respecting the rights of people – especially those who are marginalised – to have a say in how their data is used?

[12] This is not a moral philosophy Literature Analysis, but uncertain readers may find further answers in that expansive literature base. The arguments of deontologists, utilitarians, social contract advocates, and many others all weigh in on the debate about what ends might justify what means in varied situations.

- Would you be okay with your personal location data being used by anyone who was willing to pay[13] a company for it?
- Would you be okay with the personal location data of your family, friends, and co-workers being used by anyone who was willing to pay a company for it?

In approaching the use of individual geolocation data researchers should also adhere to a decision pathway similar to the one in Figure 5. If at any stage of the process your answer to the question in Figure 5 is not 'yes', then the data should not be used. While actual questions around the use of data in practice will likely require more nuance, the ones in Figure 5 do capture the essence of the minimum decisions that should be actively made as researchers when considering the use of individual geolocation data of *any* form (i.e., not just geolocation app data).

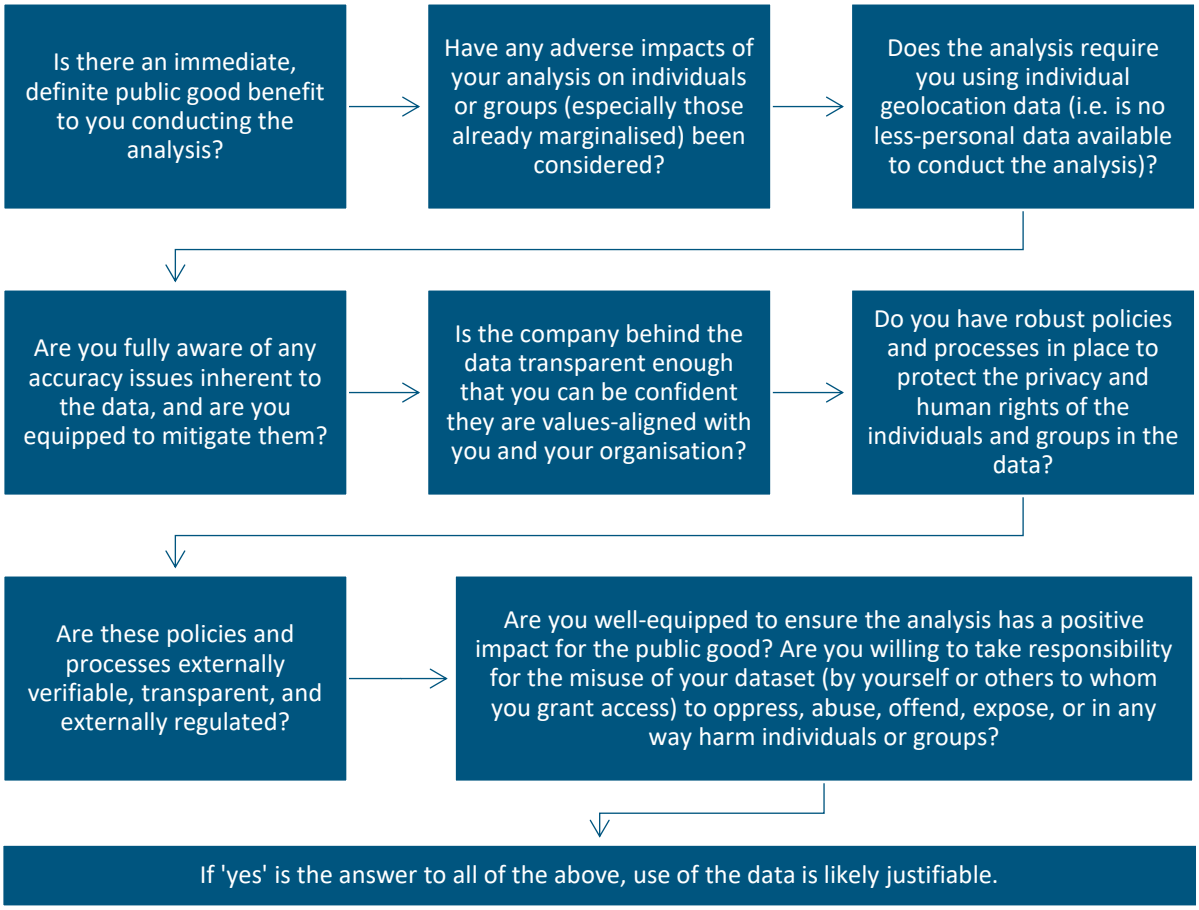


Figure 5: An example of a decision tree for considering the use of individual geolocation data. If the answer to any question is not 'yes', then the data should ideally not be used.

[13] Note that geolocation app data may not be as expensive as you expect: as part of their investigation into SafeGraph, Cox (2022) purchased one week's worth of individualised geolocation app data pinned to over 600 Planned Parenthood clinics in the US for just over $160.

If you are dedicated to using geolocation app data for your research purposes, then the most important recommendation is to at least follow *some* kind of ethical guide. It may involve replicating the approach of the IDI in New Zealand by following the Five Safes framework (Stats NZ, 2022), or it may involve adhering to the Māori Data Sovereignty Principles (Te Mana Raraunga, 2018). It might involve international guidance, such as that provided by *The Locus Charter* (EthicalGEO, 2021) or by the Human Rights Watch (2020).

It might be something different altogether; best practice is likely to remove the profit motive by navigating away from data capitalism, towards the 'more alternatives' to utopian or dystopian data uses that Thatcher and Dalton (2022) suggest are possible. One example of an approach relatively removed from data capitalism could involve you creating your own bespoke data collection approach (including possibly an app) that has transparency and privacy built in from the outset, and which ensures that only the data that is strictly necessary (and that participants are willing to share) is collected (e.g. for an example of a geolocation data collection app designed for a research study see Bähr et al., 2022).

At times the ethical choice can be difficult to accept. As Dalton et al. identify, sometimes "the ethical choice is to *not* collect data, to *not* make a map" (2016: 7, emphasis added). When it comes to the use of geolocation app data, your research may cause more harm than your inaction would.

Reference List

Abrar, S. M., Awasthi, N., Smolyak, D., & Frias-Martinez, V. (2023). Analysis of performance improvements and bias associated with the use of human mobility data in COVID-19 case prediction models. *ACM J. Comput. Sustain. Soc., 1*(2). https://doi.org/10.1145/3616380

Athey., S., Ferguson, B., Gentzkow, M., & Schmidt, T. (2021). Estimating experienced racial segregation in US cities using large-scale GPS data. *PNAS*, *118*(46). https://doi.org/10.1073/pnas.2026160118

Bähr, S., Haas, G.-C., Keusch, F., Kreuter, F., & Trappmann, M. (2022). Missing Data and Other Measurement Quality Issues in Mobile Geolocation Sensor Data. *Social Science Computer Review, 40*(1), 212-235. https://doi.org/10.1177/0894439320944118

Baron, B., & Musolesi, M. (2020). Where you go matters: A study on the privacy implications of continuous location tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4*(4), 1-32. https://doi.org/10.1145/3432699

Bietti, E. (2020). *From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy.* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 210-219. https://doi.org/10.1145/3351095.3372860

Buckee, C. O., Balsari, S., Chan, J., Crosas, M., Dominici, F., Gasser, U., Grad, Y. H., Grenfell, B., Halloran, E., Kraemer, M. U. G., Lipsitch, M., Metcalf, C. J. E., Meyers, L. A., Perkins, T. A. Santillana, M., Scarpino, S. V., Viboud, C., Weslowski, A., & Schroeder, A. (2020). Aggregated mobility data could help fight COVID-19. *Science, 368*(6487), 145-146.

Campbell, M. (2024). New Data, New Directions: A Commentary on Emerging Big Geospatial Data for Population Research | Ngā Raraunga Hou, Ngā Aronga Hou: He kōrero mō te Raraunga Rarahi Mokowā ā-Nuku e Maiea ana mō te Rangahau Taupori. *New Zealand Population Review*, *50*, 211-222.

Chen, A. (2020). *The trade-offs for digital data and contact tracing*. University of Auckland, Koi Tū: The Centre for Informed Futures. https://informedfutures.org/the-trade-offs-for-digital-data-and-contact-tracing/

Cinnamon, J., Jones, S. K., & Adger, W. N. (2016). Evidence and future potential of mobile phone data for disease disaster management. *Geoforum, 75*, 253-264. https://doi.org/10.1016/j.geoforum.2016.07.019

Cormack, D., & Kukutai, T. (2022). Indigenous Peoples, Data, and the Coloniality of Surveillance. In, A. Hepp, J. Jarke, L. Kramp (Eds.), *New Perspectives in Critical Data Studies. Transforming Communications – Studies in Cross-Media Research* (121-141). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-96180-0_6

Cox, J. (2020, November 16). How the U.S. Military Buys Location Data from Ordinary Apps. *VICE*. https://www.vice.com/en/article/us-military-location-data-xmode-locate-x/

Cox, J. (2022, May 3). Data Broker Is Selling Location Data of People Who Visit Abortion Clinics. *VICE*. https://www.vice.com/en/article/location-data-abortion-clinics-safegraph-planned-parenthood/

Dalton, C. M., Taylor, L., & Thatcher, J. (2016). Critical Data Studies: A dialog on data and space. *Big Data & Society, 3*(1). https://doi.org/10.1177/2053951716648346

Davison, A., Beetham., Thomas, J., Harding, A., Ivory, V., & Bowie, C. *Public Attitudes to Data Integration*. Opus International Consultants Ltd. https://www.stats.govt.nz/corporate/public-attitudes-to-data-integration

de Montjoye, Y-A., Gambs, S., Blondel, V., Canright, G., de Cordes, N., Deletaille, S., Engø-Monsen, K., Garcia-Herranz, M., Kendall, J., Kerry, C., Krings, G., Letouzé, E., Luengo-Oroz, M., Oliver, N., Rocher, L., Rutherford, A., Smoreda, Z., Steele, J., Wetter, E … Bengtsson, L. (2018). Comment: On the privacy conscientious use of mobile phone data. *Scientific Data, 5*(180286).

Digital Inclusion Research Group. (2017). *Digital New Zealanders: The Pulse of our Nation*. MBIE & DIA. https://www.mbie.govt.nz/dmsdocument/3228-digital-new-zealanders-the-pulse-of-our-nation-pdf

Epstein, L. C., & Lasagna, L. (1969). Obtaining informed consent: Form or substance. *Arch Intern Med., 123*(6), 682-688. doi:10.1001/archinte.1969.00300160072011

EthicalGEO. (2021). *The Locus Charter*. https://ethicalgeo.org/locus-charter/

Gao, S., Rao, J., Kang, Y., Liang, Y., Kruse, J., Dopfer, D., Sethi, A. K., Reyes, J. F. M, Yandell, B. S., & Patz, J. A. (2020). Association of Mobile Phone Location Data Indications of Travel and Stay-at-Home Mandates With COVID-19 Infection Rates in the US. *JAMA Network Open, 3*(9). doi:10.1001/jamanetworkopen.2020.20485

Garattini, C., Raffle, J., Aisyah, D. N., Sartain, F., & Kozlakidis, Z. (2019). Big data analytics, infectious diseases and associated ethical impacts. *Philos. Technol., 32*, 69-85.

Gluckman, P. (2017). *Using Evidence to Inform Social Policy: The Role of Citizen-Based Analytics.* Office of the Prime Minister's Chief Science Advisor (New Zealand).

Greaves, L. M., Latimer, C. L., Muriwai, E., Moore, C., Li, E., Sporle, A., Clark, T. C., Milne, B. J. (2024). Māori and the Integrated Data Infrastructure: an assessment of the data system and suggestions to realise Māori data aspirations [Te Māori me te Integrated Data Infrastructure: he aromatawai i te pūnaha raraunga me ngā marohitanga e poipoia ai ngā wawata raraunga Māori]. *Journal of the Royal Society of New Zealand*, *54*(2), 190-206. https://doi.org/10.1080/03036758.2022.2154368

GSMA. (2014). *GSMA guidelines on the protection of privacy in the use of mobile phone data for responding to the Ebola outbreak*. GSMA. https://www.gsma.com/mobilefordevelopment/resources/gsma-guidelines-on-the-protection-of-privacy-in-the-use-of-mobile-phone-data-for-responding-to-the-ebola-outbreak/

Human Rights Watch. (2020, May 13). *Mobile Location Data and Covid-19: Q&A*. https://www.hrw.org/news/2020/05/13/mobile-location-data-and-covid-19-qa

Keßler, C., & McKenzie, G. (2018). A geoprivacy manifesto. *Transactions in GIS, 22*(1), 3-19. https://doi.org/10.1111/tgis.12305

Kollnig, K., Binns, R., Dewitte, P., Van Kleek, M., Wang, G., Omeiza, D., Webb, H., & Shadbolt, N. (2021). *A fait accompli? An empirical study into the absence of consent to third-party tracking in Android apps*. Proceedings of the Seventeenth Symposium on Usable Privacy and Security, August 9-10, 2021.

Kukutai, T., Cassim, S., Clark, V., Jones, N., Mika, J., Morar, R., Muru-Lanning, M., Pouwhare, R., Teague, V., Tuffery Huria, L., Watts, D., & Sterling, R. (2023). *Māori data sovereignty and privacy*. Tikanga in Technology Discussion Paper, Te Ngira Institute for Population Research. https://www.waikato.ac.nz/assets/Uploads/Research/Research-institutes-centres-and-groups/Institutes/Te-Ngira-Institute-for-Population-Research/MDSov-and-Privacy_20March2023_v2.pdf

Miller, S., & Smith, M. (2021). Ethics, public health and technology responses to COVID-19. *Bioethics, 35*, 364-371. https://doi.org/10.1111/bioe.12856

Mitchell, C. (2023, February 6). Massive government database had rules breached more than 100 times. *Stuff*. https://www.stuff.co.nz/national/300783781/massive-government-database-had-rules-breached-more-than-100-times

Mittelstadt, B. (2017). From individual to group privacy in big data analytics. *Philos. Technol., 30,* 475-494. DOI 10.1007/s13347-017-0253-7

New Zealand Human Rights Commission. (2018). *Privacy, Data and Technology: Human Rights Challenges in the Digital Age*. New Zealand Human Rights Commission

O'Connor, H., Hopkins, W. J., & Johnston, D. (2021). For the greater good? Data and disasters in a post-COVID world. *Journal of the Royal Society of New Zealand, 51*(S1), S214-S231.

Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., De Nadai, M., Letouzé, E., Salah, A. A., Benjamins, R., Cattuto, C., Colizza, V., de Cordes, N., Fraiberger, S. P., Koebe, T., Lehmann, S., Murillo, J., Pentland, A., Pham, P. N., Pivetta, F., … Vinck, P. (2020). Mobile phone data for informing public health actions across the COVID-19

pandemic life cycle. *Science Advances, 6*(23). https://doi.org/10.1126/sciadv.abc0764

Ryan, J., & Christl, W. (2023a). *America's Hidden Security Crisis*. Irish Council for Civil Liberties. https://www.iccl.ie/digital-data/americas-hidden-security-crisis/

Ryan, J., & Christl, W. (2023b). *Europe's Hidden Security Crisis*. Irish Council for Civil Liberties. https://www.iccl.ie/digital-data/europes-hidden-security-crisis/

Schlosser, F., Sekara, V., Brockmann., & Garcia-Herranz, M. (2021). *Biases in human mobility data impact epidemic modeling*. Preprint article December 24, 2021. https://arxiv.org/pdf/2112.12521.pdf

Stats NZ. (2022). *How We Keep Integrated Data Safe.* Stats NZ. https://www.stats.govt.nz/integrated-data/how-we-keep-integrated-data-safe/

Sterling, R., Kukutai, T., Chambers, T., & Chen, A. T-Y. (2024). A Māori data governance assessment of the NZ COVID Tracer app. *Discover Social Science and Health, 4*(32). https://doi.org/10.1007/s44155-024-00092-2

Tau, B., Mollica, A., Haggin, P., & Volz, D. (2023, October 13). How Ads on Your Phone Can Aid Government Surveillance. *Wall Street Journal*. https://www.wsj.com/tech/cybersecurity/how-ads-on-your-phone-can-aid-government-surveillance-943bde04

Te Mana Raraunga. (2018). *Principles of Māori Data Sovereignty*. Te Mana Raraunga: Māori Data Sovereignty Network. https://static1.squarespace.com/static/58e9b10f9de4bb8d1fb5ebbc/t/5bda208b4ae237cd89ee16e9/1541021836126/TMR+Ma%CC%84ori+Data+Sovereignty+Principles+Oct+2018.pdf

Thatcher, J. E., & Dalton, C. M. (2022). *Data Power: Radical Geographies of Control and Resistance*. Pluto Press. https://doi.org/10.2307/j.ctv249sg9w

Timutimu, L. (December 20, 2023). *WAI3311: Statement of Claim*. https://linktr.ee/takiaho

Valentino-DeVries, J., Singer, N., Keller, M. H., Krolik, A. (2018, December 10). Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret. *The New York Times*. https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html

Wyden, R. (2024, February 13). *Signed Near Letter to FTC and SEC*. United States Senate, Office of Ron Wyden. https://www.wyden.senate.gov/imo/media/doc/signed_near_letter_to_ftc_and_sec.pdf

Zhang, H., McKenzie, G., Tomko, M., Egorova, E., & Kim, J. (2022). *Report from the First Workshop on Cyber Ethics in Platial Research*. In: F.B. Mocnik and R. Westerholt (eds.), Proceedings of the 3rd International Symposium on Platial Information Science (PLATIAL'21), 87–92. https://doi.org/10.5281/zenodo.6413003

Zhang, H., & McKenzie, G. (2023). Rehumanize geoprivacy: From disclosure control to human perception. *GeoJournal*, *88*, 189-208.

Appendix A: Glossary of terms and acronyms

EULA: End User Licensing Agreement.

FPIC: Free, prior, and informed consent.

Geolocation app data: Large scale individual geolocation data derived from mobile phone apps. Datasets generally contain a large quantity of de-identified individual-level high-resolution GPS data (latitude and longitude) over time. Some also include inferred personal characteristics such as ethnicity, gender, age, and occupation.

GPS: Global Positioning System.

IDI: Integrated Data Infrastructure, an Aotearoa New Zealand government database containing individual de-identified administrative and survey information on most people in the country.

RTB data: Real-Time Bidding data, sourced from ad data exchanges.

Appendix B: Research Questions and Method

This appendix outlines the two research questions that motivated and informed the approach to searching and reading the literature. It also provides information on the broad steps in the Literature Analysis process, as well as what type of sources were in scope and which search engines and search terms were used.

AB.1 Research Questions

The following primary and secondary research questions inform the investigation underlying this Literature Analysis.

Primary Research Question: What does ethical use of geolocation app data look like for research?

Secondary Research Question: What Aotearoa-specific considerations apply to the use of geolocation app data for research?

AB.2 Approach to Literature Analysis

This Literature Analysis involved a scan of the available literature on topics relating to the appropriate use of individual mobile phone data and/or geolocation data, including several synonyms for both terms. This is not a comprehensive review of all available literature on the subject, though several of the sources included in this summary have reviews of relevant literature themselves if needed. The focus of this summary was on identifying, summarising, and briefly synthesising a selection of key sources relating to the above Research Questions.

The Literature Analysis followed the same broad approach as most investigations of literature, consisting of the following steps:

Step 1. An initial set of sources were scoped based on searches relating to known authors, organisations, and topics in the field.

Step 2. Sources were further investigated and classified based on relevance (judged by the abstract or executive summary), with some being excluded and other new sources continually added to the pool.

Step 3. High priority sources were read, reflected on, and notes were taken.

- Sources that discussed geolocation mobile app data specifically (rather than geolocation data or app data separately) were given highest priority, though this was dependent on the specific topic focus of the source.
- As in many fields, several authors have published multiple articles relating to similar subject matter; in these instances, more recent sources by the same author(s) were prioritised over reading their full body of work in the interest of capturing more diverse perspectives.

- The literature was summarised and synthesised using a semi-structured thematic literature table approach, whereby the relevant information from sources were abstracted and commented on based on their alignment with two key topics: the ethics of geolocation app data use, and the accuracy of geolocation app data. The key messages of sources were also captured, as well as any miscellaneous relevant information that fell beyond the scope of the two key topics. See this table with the raw unaltered notes as filled out in Appendix A.
- Review of the literature was conducted entirely by one person. While this does bring a consistent lens to the analysis, it does mean that the review will not have benefited from the diverse perspectives that multiple positionalities would bring.

Step 4. Another search of the literature was conducted part way through the summary once the scope of the literature was better understood. Reading and note taking continued.

Step 5. Reading largely continued, moving on to the synthesis of key insights from across the literature for inclusion in this document.

- Hundreds of potentially relevant sources were located, with approximately 100 saved for assessment of direct relevance and approximately 50 read with some degree of depth.
- Sources spanned various disciplines and sub-disciplines, with most sources being concentrated in the fields of Big Data Analytics, Human Geography, GIS, or Ethics more broadly.

Key factors in the literature search for this project include:

- **Published Sources:** This Literature Analysis draws primarily on published resources (journal articles, reports, book chapters, grey literature, and similar outputs with a fixed publication date). These sources are more readily available, and are unlikely to be edited or removed in the future, making them robust sources of knowledge to refer back to. Some relevant website pages were included, however.
- **Search Engines:** Sources for this Literature Analysis were largely obtained from Google Scholar and Google. Searches within particularly relevant journals were also conducted.
- **Search Terms:** The search terms for this Literature Analysis reflected the broad data focuses of this research. Searches usually involved some combination of the following (or similar) terms: mobile data, mobile app data, GPS, geolocation, data ethics, appropriate use of data, geoprivacy, spatial data, disaster data ethics, data privacy, COVID-19, disease response data, data accuracy, location data, RTB [Real-Time Bidding] data ethics, Aotearoa, New Zealand, Māori data. Sources were also found through the reference lists of other included sources, and so were located directly through Google, Google Scholar, or similar.
- **Paywalled Sources:** Research for this Literature Analysis focused on using publicly available sources where possible. Many paywalled sources were found but few were included in the review given ample publicly available literature.