

# Would Citizens Contribute their Personal Location Data to an Open Database? Preliminary Results from a Survey

Vilma Jokinen\*, Ville Mäkinen\*, Anna Brauer\*† and Juha Oksanen\*

\* Finnish Geospatial Research Institute FGI, National Land Survey of Finland

† Department of Computer Science, University of Helsinki

**Abstract.** The amount of movement data that people record using their mobile phones via different tracking apps is vast. In a typical case the data can be viewed within the app but using the data by the third-party for other purposes is cumbersome, or practically impossible. One way to improve the situation is to establish an open trajectory data repository, where the users could save their movement tracks as open data. However, this data is considered personal data and the users may not be willing to share full trajectories as they might reveal for example their home locations. Thus, the trajectory data must be processed to minimize the amount of information that can be used to identify person while keeping the utility of the data as high as possible. We launched a survey of peoples' opinions about sharing their movement data and what kind of privacy guarantees they would expect. Based on the preliminary results, a large part of the potential users appears to be interested in sharing their tracking data, when adequate privacy-preserving pre-processing is performed.

**Keywords.** Human trajectory data, mobile tracking data

## 1. Introduction

The diffusion of smartphones and other GNSS-enabled mobile devices among the population has been exceptionally fast and extensive (Bento 2016), and a notable amount of people are recording their movement by fitness and other tracking applications provided by private companies. Typically, the companies keep the data mostly to themselves and allow access to users' data often only through their web sites. Exceptions, such as Strava



Published in "Proceedings of the 16th International Conference on Location Based Services (LBS 2021)", edited by Anahid Basiri, Georg Gartner and Haosheng Huang, LBS 2021, 24-25 November 2021, Glasgow, UK/online.

<https://doi.org/10.34726/1784> | © Authors 2021. CC BY 4.0 License.

Metro (Strava 2021), do exist, but they provide only aggregated data and limit access to the data only for actors directly involved in active transportation infrastructure planning. This leaves out large number of potential data users, like scientists, from the end users (Personal communication 2021). Thus, the data is fragmented and heavily processed, and access to the data is arbitrarily restricted, which inhibits its full utilization potential. On the other hand, there are various reasons hindering sharing of the tracking data openly. Most importantly, human mobility data is personal data whose use in the EU is regulated by the General Data Protection Regulation (GDPR).

In the GeoPrivacy project, we study methods on how human mobility data could be collected, used, and shared in a privacy-protected open database. One objective of the project is to find out citizens' attitudes about sharing their mobile tracking data, and possible motivations for it. We conducted an online survey and in this abstract, we present our preliminary findings related to 1) citizens' willingness to contribute to an open location data repository, 2) the adequate level of privacy protection, 3) motivation and 4) obstacles for sharing personal location data.

## 2. Materials and Methods

The survey was conducted using an online questionnaire form. The form was distributed through a number of email lists and social media channels, mainly for Finnish audiences. The survey consisted of 17 questions about the participants' mobile tracking habits, their opinions and worries about sharing their tracking data publicly, and background information. A Finnish and an English version of the questionnaire were distributed.

The preliminary findings presented in this abstract are based on the 325 responses collected between the 28th of June and the 31st of August 2021. The main method used for analysing the survey data was cross tabulation.

The four questions under investigation in this abstract were presented in the survey as follows:

1. Would you consider sharing your tracking data to an open data repository? (A five-step Likert scale from one ("definitely no") to five ("definitely yes"))
2. Which level of privacy protection should the repository guarantee, so that you would be willing to contribute? (Table 1)
3. What would motivate you to share your tracking data? (Table 2)
4. What would prevent you from sharing your tracking data? (Table 3)

**Table 1.** The answer options and their abbreviations for Question 2 “Which level of privacy protection should the repository guarantee, so that you would be willing to contribute?” (Single-choice)

Answer option	Abbreviation
Level 0: Your tracks are published as they are.	Level 0
Level 1: Your tracks are published as they are, but your identifying information is replaced by a pseudo-identifier (for example, a random number).	Level 1
Level 2: Places that you visit (and spend a considerable amount of time) are removed, including your start location and destination. The remaining parts of the tracks are published as they are.	Level 2
Level 3: Your tracks are processed before publication, so that it is very unlikely that any part of the track can be traced back to you.	Level 3
Level 4: Your data is combined with the data of all other users, and only this summarized data is published. Information about any individual cannot be extracted.	Level 4
You are not willing to contribute, no matter what the privacy guarantee of the repository would be.	Never
Other (free text field)	

**Table 2.** The answer options and their abbreviations for Question 3 “What would motivate you to share your tracking data?” (Multiple-choice)

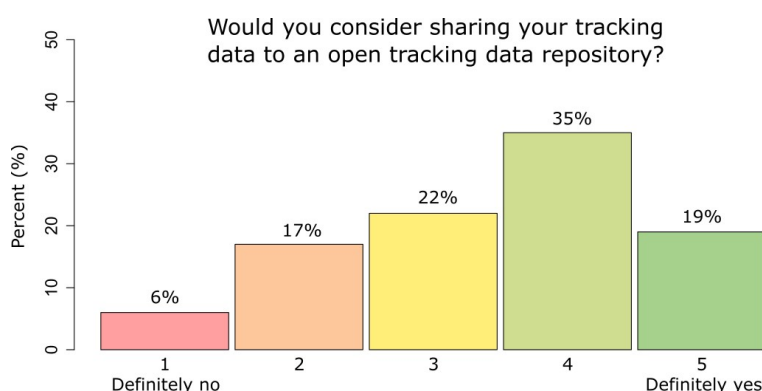
Answer option	Abbreviation
getting summary statistics about yourself (e.g., comparisons to your past activity)	ownSummary
comparison of your summary statistics to other contributors	compSummary
possible improvements in cycling and/or pedestrian lane infrastructure	imprLanes
possible improvements to the safety of lanes used by cyclists and pedestrians/runners	imprSafety
possible improvements to the cost-effectiveness of investments in infrastructure	imprCost
enthusiasm for supporting open data	openData
contribution in scientific research	research
a possibility to influence policy-making in the long term	policymaking
other (free text field)	

**Table 3.** The answer options and their abbreviations for Question 4 “What would prevent you from sharing your tracking data? (Multiple-choice)”

Answer option	Abbreviation
concerns that the data might reveal some personal information	persInfo
concerns about the causes my data will be used for	dataUse
not having enough time	lackTime
not being interested enough	noInterest
technical difficulties in sharing my data	techDiffic
other (free text field)	

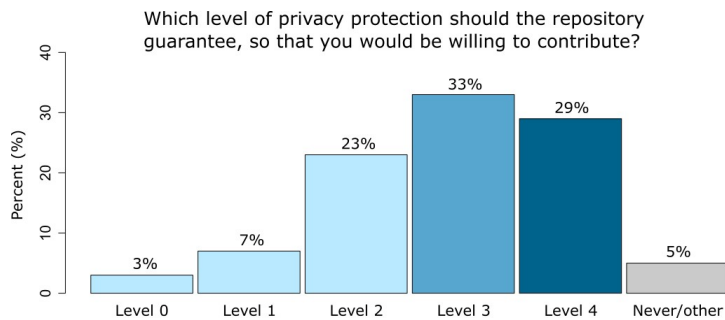
### 3. Preliminary results

*General attitude:* According to the preliminary results, the respondents' attitudes towards sharing their tracking data were mostly positive (Figure 1). The positive general attitude (GA) groups 4 and 5 covered 54 % of the replies, and 77 % had no clear objections to sharing their data.



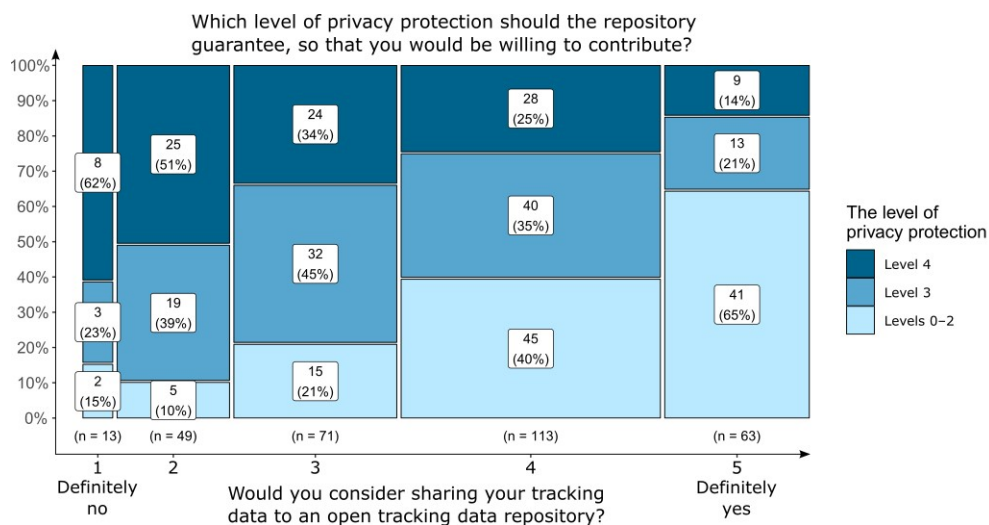
**Figure 1.** Respondents' general attitude about sharing mobile tracking data.

*Privacy level requirements:* Concerning the respondents' opinions about the privacy protection levels of the tracking data repository, level 3 was the most popular option with 33 % of the replies (Figure 2). The levels 0 and 1 were clearly less popular than the other levels. In the following cross tabulation analysis, they are therefore merged to level 2. Only 3 % of respondents said they would not be willing to contribute, regardless of the privacy guarantee of the repository.



**Figure 2.** Respondents' privacy requirements for sharing their data.

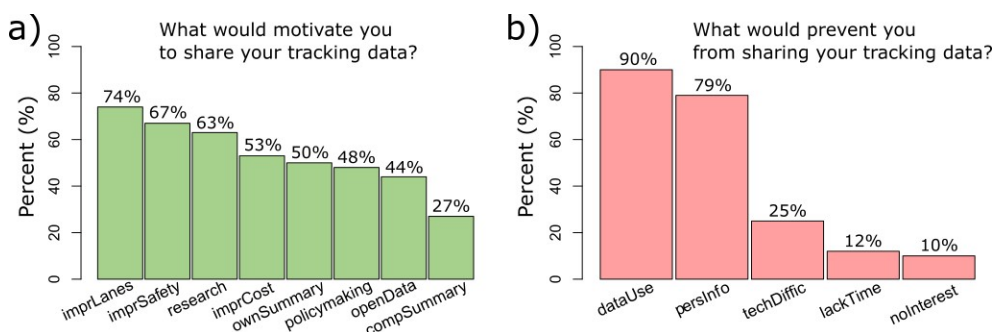
The more positive a respondent's attitude towards data sharing was, the lower the level of privacy protection they would be satisfied with appeared to be (Figure 3,  $n = 309$ ,  $p \leq 0,001$ ). Among those respondents who had either neutral or positive attitudes towards sharing their data (GA groups 3, 4, and 5), 75 % would be willing to contribute if the privacy protection level was 3 or higher. Focusing only on the people who have positive attitudes towards sharing the data (GA groups 4 and 5), 79 % would be pleased with level 3 or higher. In contrast, among the respondents who have negative attitudes towards sharing their data, 53 % require the strictest privacy protection level 4 before considering sharing their data.



**Figure 3.** Respondents' requirement of privacy level vs. their willingness to share their tracking data. The privacy levels 0 and 1 are merged to level 2.

*Motivators and preventors:* The three most popular motivators for sharing data were development of cycling and/or pedestrian lane infrastructure, improvements of lane safety, and the contribution in scientific research (Figure 4a). Of the reasons that could prevent the respondents from sharing

their data, two clearly stand out: concerns about the causes the data will be used for (90 %) and concerns that the data might reveal personal information (79 %) (Figure 4b).



**Figure 4.** a) Respondents' motivators and b) reasons that would prevent them from sharing their data.

## 4. Conclusion

Generally, based on our survey, the attitude towards donating personal tracking data to a privacy-protected open data repository appears to be positive. Most respondents are privacy conscious and require advanced levels of data protection. The most popular motivators for donating the data are related to seeing improvements in biking and pedestrian infrastructure, but such changes are long-term and raise the question of how immediate feedback could be provided for the contributors. The small sample size and bias in the responses must be considered carefully. All findings provide important insights on issues related to designing a privacy-protected open repository for personal location data.

## 5. Acknowledgements

GeoPrivacy project funding from the Finnish Cultural Foundation is gratefully acknowledged.

## References

- Bento, N. (2016). Calling for change? Innovation, diffusion, and the energy impacts of global mobile telephony. *Energy Research & Social Science*, 84–100.
- Personal Communication (2021). Strava Metro Application Update. Personal email June 29th 2021.
- Strava (2021). Better cities for cyclists and pedestrians. On-line: <https://metro.strava.com/>