AGILE: GIScience Series, 4, 21, 2023. https://doi.org/10.5194/agile-giss-4-21-2023
Proceedings of the 26th AGILE Conference on Geographic Information Science, 2023.
Editors: P. van Oosterom, H. Ploeger, A. Mansourian, S. Scheider, R. Lemmens, and B. van Loenen.
This contribution underwent peer review based on a full paper submission.

Human mobility tracks as FAIR data: Designing a privacy-preserving repository for GNSS-based activity tracking data

Anna Brauer [1,2], Ville Mäkinen [1], and Juha Oksanen [1]

[1] Department of Geoinformatics and Cartography, Finnish Geospatial Research Institute FGI, National Land Survey of Finland, Espoo, Finland
[2] Department of Computer Science, University of Helsinki, Helsinki, Finland

Correspondence: Anna Brauer (anna.brauer@nls.fi)

Abstract. Devices with integrated global navigation satellite system (GNSS) receivers have enabled citizens to accurately record activities such as bicycle trips, runs, and walks. Due to its spatiotemporal extent and high level of detail, GNSS-based activity tracking data is a valuable source of information on active modes of transportation. At the same time, movement recordings of individuals are sensitive data and are associated with privacy concerns. In this work, we present a privacy-aware platform where citizens can contribute GNSS tracks to an open repository. The repository is published according to the FAIR data principles: findable, accessible, interoperable, and reusable. This provides the opportunity to use the data as a benchmark for the development of GNSS trajectory processing methods. The platform's privacy module processes each track before publication, concealing stay points, generalizing the tracks in the temporal dimension, and suppressing tracks in sparsely populated areas. This approach mitigates the most likely re-identification attacks and limits the amount of information that could leak if an attacker succeeds with re-identification. As a residual risk remains, the platform sensitizes users to privacy risks and enables them to make informed decisions about publishing their data.

Keywords. GPS trajectories, bicycle data, location privacy-preserving data publishing, VGI, open data

1 Introduction

Safe, carbon-neutral mobility is one of the cornerstones of the European Commission's long-term vision, as transport accounts for about 25% of the European Union's greenhouse gas emissions (European Comission, 2018). In cities, the most sustainable modes of travel are walking and cycling (Pucher and Buehler, 2017). Promoting these non-motorized modes of travel can help to increase urban livability and reduce air pollution and traffic congestion (Hickman and Banister, 2014).

Many cyclists and pedestrians use tracking applications (e.g., Strava) and handheld or wearable devices equipped with Global Navigation Satellite System (GNSS) technology to record their activities. Tracking data has the potential to revolutionize urban planning: it can reveal route preferences, enable detailed analyses of effects and unexpected trends, facilitate the design of policies sensitive to weather conditions, and offer opportunities for increasing citizens' participation in planning (Milne and Watling, 2019).

Despite the popularity of tracking applications, the availability of GNSS tracks from non-motorized movement remains poor. Targeted data collection campaigns are costly, time-consuming, and limited in scope (e.g., Broach et al., 2012; Korpilo et al., 2017), while commercial application providers have little economical motivation to share the data and are obliged to comply with privacy regulations. A few efforts have been made to share tracking data in an aggregated form, notably Strava's Metro (Strava, 2020) and the Bike Data Project (Bike Data Project, 2020). However, aggregated data sets are not always transparent about the methods used for aggregation. Furthermore, aggregation limits the application possibilities of the data significantly.

As GNSS tracks of individuals are sensitive, personal data, they should be shared only if the individuals are able to give their informed consent. In 2021, we conducted a survey among Finnish citizens to investigate their views about sharing tracking data in an openly available data repository (Jokinen et al., 2021). The participants' attitude was positive, with *supporting research* being among the top three motivating factors for considering sharing tracking data.

Encouraged by the survey results, we developed a pilot service[1] that enables citizens to share tracking data recorded during bicycle trips, walks, and runs. Anyone actively moving in Finland can upload data; the service deidentifies the tracks and applies privacy-protecting mechanisms before publishing them in an open repository according to the FAIR (findable, accessible, interoperable, reusable) data principles (Wilkinson et al., 2016). Tracking data that is openly accessible to the public could power innovative applications and enable future studies requiring individual-level data (e.g., (Brauer et al., 2021; Scott et al., 2021)). Furthermore, realistic and heterogeneous GNSS tracks are a valuable asset for algorithm development. Making the data openly available ensures that experimental results are reproducible, lowering the threshold for quantitative comparisons between mechanisms.

In this paper, we focus on the design of the service's privacy module, which combines three privacy-preserving mechanisms. After describing the state-of-the-art of privacy-preserving trajectory publishing, we analyze the risks and requirements of our case. We then present our solution, introducing the architecture of the service and the privacy module. Finally, we discuss and summarize the work.

2 Privacy-preserving publishing of human mobility tracks

Protecting the privacy of individuals when sharing tracking data, i.e., timestamped sequences of location measurements referred to as trajectories, is a challenging task that has received a considerable amount of attention within the last decade. In general, personal location data is very unique for each individual (De Montjoye et al., 2013; Golle and Partridge, 2009), which makes adapting privacy concepts developed for relational data difficult. Many mechanisms for mitigating privacy risks have been proposed; some of these mechanisms focus on sensitive locations (Huo et al., 2012; Dai et al., 2018), while others partition the tracks (Song et al., 2014) and use location generalization (Gidofalvi et al., 2007; Ma et al., 2010) or perturbation (Seidl et al., 2016) to obfuscate the tracks. Techniques like generalization and perturbation can also be used to render each track indistinguishable from $k-1$ other tracks based on the principle of k-anonymity (Abul et al., 2008; Nergiz et al., 2008), which was introduced by Sweeney (2002) for relational data. Moreover, privacy-protecting mechanisms based on Dwork (2006)'s differential privacy concept have been proposed for trajectory data, primarily generating synthetic tracks from differentially private data representations (Chen et al., 2012; He et al., 2015) or publishing perturbed locations (Hua et al., 2015; Andrés et al., 2013). For a comprehensive overview of mechanisms, we refer to Fiore et al. (2020) and Primault et al. (2018).

3 Analysis of risks and requirements

At the root of designing a privacy protection scheme is a thorough analysis of the risks and requirements, which we outline in this section.

3.1 Risks

In the following, we analyze the risks that arise when an attacker attempts to link the data in the open repository back to individual users with the goal to extend his knowledge about them. The term *attacker* is a theoretical concept referring to any individual person, group of persons, or organization aiming to retrieve new information for a malicious or benign purpose. For clarity, we refer to an individual attacker in the following. We assume that the attacker does not try to gain unauthorized access to the system or intercept communication between the service and the users; these attacks are not in the scope of this paper, although they have been considered in the service design and countermeasures have been implemented.

We differ between two categories of attacks: re-identification attacks aiming to match tracks to users and attribute linkage attacks retrieving sensitive information about individuals from the tracking data. Linking attributes is usually preceded by re-identification, but may also occur when a sensitive attribute is shared by a set of candidate tracks potentially associated with a user (Fung et al., 2010).

Successful re-identification requires the attacker to possess prior information about a user. Potential sources of this information include personal observation, legacy media, and social media, notably also activity sharing platforms. Furthermore, the attacker may have access to confidential data, e.g., data gathered with a location-based mobile service. We identify four principal types of prior information based on Fiore et al. (2020). Sorted by complexity and difficulty to obtain, these are:

1. **Spatiotemporal information.** The attacker has knowledge of a user's whereabouts at a certain time that may be exact (i.e., a spatio-temporal point) or uncertain in the temporal dimension or the spatial dimensions.
2. **Important locations.** The attacker is aware of a user's points of interest, e.g., their home, workplace, or any other location they visit regularly.
3. **Mobility features.** Characteristics describing a user's movement that include, e.g., their preferred mode of travel, average and maximum speed. They may be learned or estimated using, e.g., publicly available tracks of a user.
4. **Mobility model.** The attacker has developed a profile of a user's historical movement, e.g., as a Markov model (Gambs et al., 2014). The model encapsulates typical mobility patterns of the user.

[1] Soon available at https://www.geoprivacy.fi

Additionally, the attacker may utilize auxiliary context data about the built environment, e.g., the road network or information on traffic conditions.

If re-identification succeeds, information about the user can leak to the attacker. We identified five types of information relevant to GNSS tracking data:

1. **Spatiotemporal information.** The location of a user at a certain point in time; this includes revealing the presence but also the absence of a person at an event.
2. **Important locations.** Places that the user visits and their semantic meaning, e.g., home or workplace.
3. **Mobility patterns.** Repeating patterns in the user's movement, e.g., commuting patterns.
4. **Fitness.** The physical condition of a user may be reflected by, e.g., their average speed of travel or performance at a slope.
5. **Social connections.** Meeting disclosure (Shokri et al., 2011) can reveal if and when two individuals meet.

All of this information is sensitive; assessing the degree of sensitivity is a personal matter, and the outcome of the assessment may vary between individuals.

Rigorous scientific attempts to evaluate the risks accompanying the publication of GNSS tracking data have yet to be made (Fiore et al., 2020). In general, the more complex an attacker's knowledge is, the less information he can potentially obtain from the data. For example, an attacker who was able to build a complex mobility model of a user must have access to an elaborate data source. It is unlikely that he will be able to expand his knowledge significantly, although not impossible (e.g., re-identification with a mobility model could enable an attacker to learn about the user's fitness). At the same time, an attacker with extensive prior knowledge is difficult to prevent from performing successful re-identification. In theory, there is no upper limit to the attacker's knowledge; if the open repository is to provide useful data, there will always be hypothetical attackers who are able to re-identify users (Dwork, 2006).

On a different note, the type of data that the open repository is targeting is *activity tracking data*. The creation of this data requires a conscious decision of the user to start the recording and save the track. As the users are usually interested in the stats of their tracks, they stop the recording as soon as they reach their destination or change to a different mode of travel. Thus, the data represents only a fraction of the users' mobility. Uploading data into the repository requires a conscious user decision as well, minimizing the probability of a user publishing data unintentionally.

3.2 Requirements

In the following, we list the technical and conceptual requirements for the service's privacy protection module. In addition to managing the privacy risks defined in the previous section, the privacy-preserving methods deployed in the module should

- **produce output with a high degree of detail.** The sanitized data should facilitate exact street-level analyses, retaining a detailed representation of the movement.
- **preserve the spatial truthfulness of the data at the track level.** The privacy module should not fabricate any data, as this could lead to unpredictable consequences and bias that are detrimental to the data's utility for small-scale analyses (Gramaglia et al., 2017).
- **sanitize each track separately.** This would eventually allow for the privacy module to be executed on the client side; the server would not need to save the original track at any time. Furthermore, methods sanitizing the whole dataset at once are prone to privacy breaches when the dataset is growing dynamically.
- **be customizable.** As privacy requirements vary between individuals, the users should be able to adapt the methods to their personal needs.

Fulfilling all requirements means keeping parts of the data's potentially sensitive information intact. For example, properties like the speed of travel are relevant for traffic analyses but also provide information about the users' fitness.

4 Our solution

The privacy module of our service sanitizes the tracking data that is uploaded by users and publishes the sanitized tracks in the open data repository (Figure 1). The repository grows dynamically, i.e., new sanitized tracks appear in the repository immediately. To enable reproducible research, we follow the FAIR guidelines and 1) publish immutable snapshots of the repository with a persistent unique identifier in the IDA[2] research data storage and 2) provide metadata using the Qvain data description tool to make the data findable in the Etsin finder service.

During the upload process, the users are prompted to indicate their preferred level of privacy protection. With their choice, they determine the parameterization of the mechanisms in the privacy module. The module implements a three-part solution: it consists of a mechanism that protects stay points, a mechanism that distorts the tracks in the temporal dimension, and a mechanism that suppresses tracks

[2] https://www.fairdata.fi/en/

within sparsely populated areas. Additionally, the sanitized tracks are pseudonymized, i.e., user identifiers are removed before publication and replaced with a pseudo-identifier (e.g., a random number).

In the following, we describe the module's three main components in more detail.

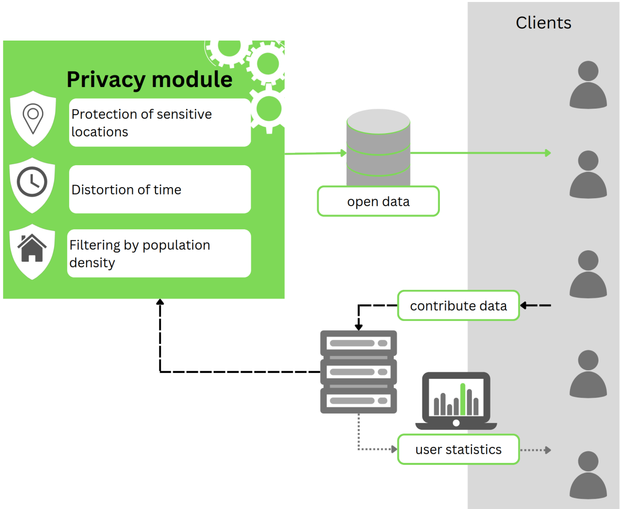


Figure 1. Service architecture. When a user uploads tracking data to the service, the server pipelines the tracks into the privacy module; the data is sanitized and published as open data. Personal user statistics are calculated using the original data that is subsequently deleted from the server.

4.1 Protection of sensitive locations

We use the Site-sensitive Trajectory Truncation (S-TT) mechanism proposed in Brauer et al. (2022) to suppress location measurements around sensitive locations. So far, the service does not offer the possibility to indicate sensitive locations manually. Thus, we assume that every location where the user spent more than three minutes is a potentially sensitive location. Additionally, we consider the origin and destination of a track to be sensitive by default. The S-TT mechanism determines a *protection set* for each sensitive location: a set of at least $k-1$ nearby locations whose convex hull includes the sensitive location. We generated the protection sets by clustering all building polygons in Finland. As the clustering is calculated only once, it is not possible to infer the sensitive location based on its protection set.

In an iterative process, the S-TT algorithm truncates the parts of the track leading to the sensitive location and away from it. The truncation process continues until two conditions are fulfilled: the location that is closest to the end of the track must not be any location in the protection set, and the track's heading must point towards none or all of the locations in the protection set (Figure 2). This ensures that the locations in the protection set are *indistinguishable* from the sensitive location with respect to the course of the track.

The privacy level chosen by the users determines the size k of the protection set. To facilitate the choice of an appropriate k, the platform provides the functionality to view the mechanism's output before confirming the upload (Figure 3).

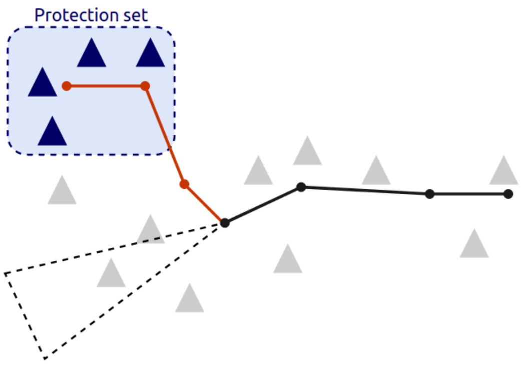
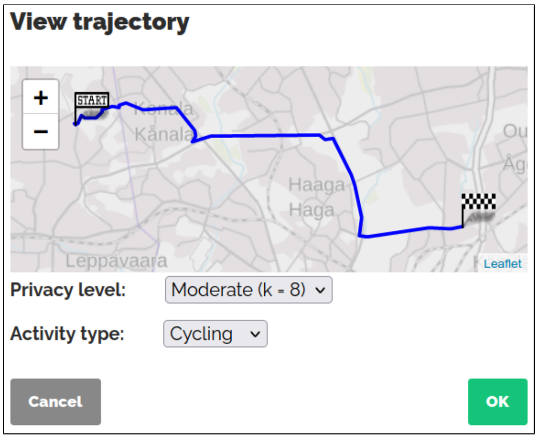


Figure 2. The S-TT mechanism truncates the track until the remaining part (depicted in black) is not in close proximity to any location in the protection set. Furthermore, the sanitized track's heading, modeled as a wedge-shaped field of view, points at either all or none of the locations in the protection set.


View trajectory

Privacy level: Moderate (k = 8)

Activity type: Cycling

Figure 3. Screenshot of a popup window displaying a track processed with S-TT. The window visualizes the effect of the privacy level on the sanitized track.

4.2 Distortion of time

The second mechanism operates solely in the temporal dimension. It subtracts a temporal *offset* from each timestamp in the track; more figuratively, the mechanism *shifts* the track in time. The offset is determined so that the sanitized track starts on the same weekday as the original track, but always at the first occurrence of this weekday in the same month. Furthermore, each day is divided into 6-hour intervals, and the start of the sanitized track is aligned with the start of the interval into which the start of the original track falls (Figure 4). Thus, the sanitized track retains relative differences between timestamps of the original track, as well as year, month, weekday, and time of

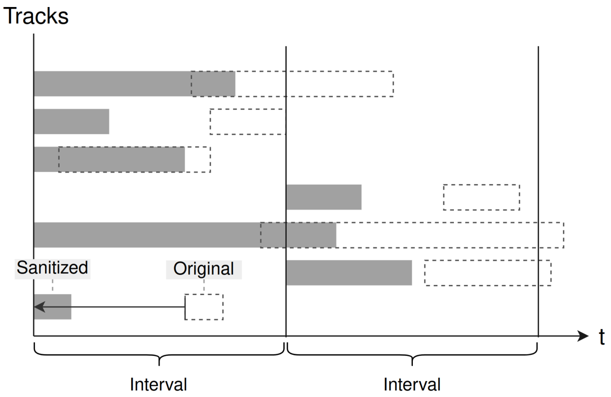


Figure 4. Distortion of time by relocating the start time of the tracks.

day. The precise time it was recorded, however, is not reconstructable.

The privacy level chosen by the user does not yet impact this mechanism; designing a customizable version of it is a subject of future work.

4.3 Filtering by population density

The third mechanism of the privacy module enforces a simple constraint: tracks are published only if they start and end in a populated area, otherwise, they are suppressed completely. To test this constraint, we use a uniform grid of 1 km x 1 km containing the number of inhabitants per cell. The origin and destination of a track are required to be in a cell having more than five inhabitants.

5 Discussion and conclusions

The design of the privacy module concentrates on attackers with limited means, as these attackers have the most to gain from linking their side knowledge to tracks in the open repository. At the same time, their attacks are the easiest to prevent, with the least compromise on utility. Although this approach leaves the data vulnerable to attacks facilitated by extensive side knowledge, we argue that these attackers have little to gain from their attack in practice. In the following, we revisit the types of prior knowledge and attributes of GNSS data established in Section 3.1 and analyze how the privacy module's combination of mechanisms prevents re-identification attacks (Table 1) and attribute linkage upon re-identification (Table 2). Note that Tables 1 and 2 are meant as rough guidance only.

Pseudonymization does not offer any protection beyond preventing simple re-identification by identifier. Stay point protection with S-TT removes the semantically most important information irreversibly from the tracks, increasing the difficulty of re-identification by important locations significantly. However, as the unicity of location data is very high (De Montjoye et al., 2013) and especially home-work location pairs are very unique among the population (Golle and Partridge, 2009), the protection set should be chosen sufficiently large, especially for regularly repeated movement. S-TT may also protect tracks from re-identification by spatiotemporal information or a mobility model, but only if the attack relies on the semantically important track parts around stay points, which attackers are arguably more likely to possess.

Table 1. Does the combination of mechanisms prevent re-identification attacks with different kinds of background knowledge?

Re-identification by	Pseud.	+ S-TT	+ Time
Identifier	yes	yes	yes
Important locations	-	mostly	mostly
Spatio-temp. inform.	-	potentially	mostly
Mobility model	-	potentially	potentially
Mobility features	-	-	-

Furthermore, the S-TT mechanism prevents the retrieval of any precise information on important locations. However, as S-TT does not alter the timestamps of the remaining track, the disclosure of the presence or absence at an event may still be possible. This serves as one rationale for distorting the timestamps. After applying this second mechanism, the exact time of visiting a place cannot be determined anymore, unless the attacker is able to calculate the temporal offset; this can be possible if the attacker possesses an accurate timestamp of any location along the track.

Similarly, re-identification by spatiotemporal information becomes more difficult if the temporal dimension is generalized, as the tracks within the same interval become more similar. There is, however, no guarantee that there are other tracks in the database that are spatially and temporally similar. This would require enforcing k-anonymity at the track level, which would severely impair data utility. Furthermore, we could argue that it is overly pessimistic to assume that the attacker knows for certain that an individual's data is in the database. Thus, our solution strives for indistinguishability on the level of the population instead of the dataset, which is approximated with the population density filter.

The privacy module does not protect against re-identification by mobility features. The odds of success of this type of attack on tracks of non-motorized movement have not been investigated so far, but we expect the attack to yield only inconclusive results. Moreover, an attacker succeeding at re-identification is able to infer the target individual's fitness based on their speed of travel, which is, however, only an imprecise indicator.

To summarize, the privacy module decreases the prospects of success of the most probable types of attacks considerably. It adopts a utility-friendly approach, enabling the open repository to support detailed mobility analytics and

Table 2. Does the combination of mechanisms prevent the attacker from expanding his knowledge about different attributes, given that he succeeded in re-identification?

Attribute	Pseud.	+ S-TT	+ Time
Important locations	-	mostly	mostly
Locality	-	potentially	mostly
Social connections	-	potentially	mostly
Mobility patterns	-	potentially	potentially
Fitness	-	-	-

facilitate track-level algorithm testing. As successful re-identification may still be possible for an attacker with extensive prior knowledge, empowering the users to make informed decisions about sharing their tracking data is central to ethical data publishing. Finding more efficient ways of communicating privacy risks to users is a subject of future work, as well as handling bias in the open repository.

Acknowledgements. This work was supported by the Finnish Cultural Foundation under Grant [00200814]. We made use of the geocomputing platform provided by the Open Geospatial Information Infrastructure for Research (Geoportti, urn:nbn:fi:research-infras-2016072513) funded by the Academy of Finland, CSC – IT Center for Science, and other Geoportti consortium members.

References

Abul, O., Bonchi, F., and Nanni, M.: Never walk alone: Uncertainty for anonymity in moving objects databases, in: 2008 IEEE 24th International Conference on Data Engineering, pp. 376–385, IEEE, 2008.

Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., and Palamidessi, C.: Geo-indistinguishability: Differential privacy for location-based systems, in: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, pp. 901–914, 2013.

Bike Data Project: About the Bike Data Project, available at: https://www.bikedataproject.org/about, last access: 22 February 2023, 2020.

Brauer, A., Mäkinen, V., and Oksanen, J.: Characterizing cycling traffic fluency using big mobile activity tracking data, Computers, Environment and Urban Systems, 85, 101 553, 2021.

Brauer, A., Mäkinen, V., Forsch, A., Oksanen, J., and Haunert, J.-H.: My home is my secret: concealing sensitive locations by context-aware trajectory truncation, International Journal of Geographical Information Science, 36, 2496–2524, 2022.

Broach, J., Dill, J., and Gliebe, J.: Where do cyclists ride? A route choice model developed with revealed preference GPS data, Transportation Research Part A: Policy and Practice, 46, 1730–1740, 2012.

Chen, R., Acs, G., and Castelluccia, C.: Differentially private sequential data publication via variable-length n-grams, in: Proceedings of the 2012 ACM Conference on Computer and Communications Security, pp. 638–649, 2012.

Dai, Y., Shao, J., Wei, C., Zhang, D., and Shen, H. T.: Personalized semantic trajectory privacy preservation through trajectory reconstruction, World Wide Web, 21, 875–914, 2018.

De Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., and Blondel, V. D.: Unique in the crowd: The privacy bounds of human mobility, Scientific reports, 3, 1–5, 2013.

Dwork, C.: Differential privacy, in: Proceedings of the International Colloquium on Automata, Languages, and Programming, pp. 1–12, Springer, 2006.

European Comission: A Clean Planet for all - A European strategic long-term vision for prosperous, modern, competitive and climate neutral economy, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0773, last access: 21 February 2023, 2018.

Fiore, M., Katsikouli, P., Zavou, E., Cunche, M., Fessant, F., Le Hello, D., Aivodji, U. M., Olivier, B., Quertier, T., and Stanica, R.: Privacy in trajectory micro-data publishing: a survey, Transactions on Data Privacy, 13, 91–149, 2020.

Fung, B. C., Wang, K., Chen, R., and Yu, P. S.: Privacy-preserving data publishing: A survey of recent developments, ACM Computing Surveys, 42, 1–53, 2010.

Gambs, S., Killijian, M.-O., and del Prado Cortez, M. N.: De-anonymization attack on geolocated data, Journal of Computer and System Sciences, 80, 1597–1614, 2014.

Gidofalvi, G., Huang, X., and Pedersen, T. B.: Privacy-preserving data mining on moving object trajectories, in: 2007 International Conference on Mobile Data Management, pp. 60–68, IEEE, 2007.

Golle, P. and Partridge, K.: On the anonymity of home/work location pairs, in: International Conference on Pervasive Computing, pp. 390–397, Springer, 2009.

Gramaglia, M., Fiore, M., Tarable, A., and Banchs, A.: Preserving mobile subscriber privacy in open datasets of spatiotemporal trajectories, in: IEEE INFOCOM 2017-IEEE Conference on Computer Communications, pp. 1–9, IEEE, 2017.

He, X., Cormode, G., Machanavajjhala, A., Procopiuc, C. M., and Srivastava, D.: DPT: differentially private trajectory synthesis using hierarchical reference systems, Proceedings of the VLDB Endowment, 8, 1154–1165, 2015.

Hickman, R. and Banister, D.: Transport, climate change and the city, Routledge, 2014.

Hua, J., Gao, Y., and Zhong, S.: Differentially private publication of general time-serial trajectory data, in: 2015 IEEE Conference on Computer Communications (INFOCOM), pp. 549–557, IEEE, 2015.

Huo, Z., Meng, X., Hu, H., and Huang, Y.: You can walk alone: trajectory privacy-preserving through significant stays protection, in: International Conference on Database Systems for Advanced Applications, pp. 351–366, Springer, 2012.

Jokinen, V., Mäkinen, V., Brauer, A., and Oksanen, J.: Would Citizens Contribute their Personal Location Data to an Open Database? Preliminary Results from a Survey, in: 16th International Conference on Location Based Services, p. 171, 2021.

Korpilo, S., Virtanen, T., and Lehvävirta, S.: Smartphone GPS tracking—Inexpensive and efficient data collection on recreational movement, Landscape and Urban Planning, 157, 608–617, 2017.

Ma, C. Y., Yau, D. K., Yip, N. K., and Rao, N. S.: Privacy vulnerability of published anonymous mobility traces, in: Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking, pp. 185–196, 2010.

Milne, D. and Watling, D.: Big data and understanding change in the context of planning transport systems, Journal of Transport Geography, 76, 235–244, 2019.

Nergiz, M. E., Atzori, M., and Saygin, Y.: Towards trajectory anonymization: a generalization-based approach, in: Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy in GIS and LBS, pp. 52–61, 2008.

Primault, V., Boutet, A., Mokhtar, S. B., and Brunie, L.: The long road to computational location privacy: A survey, IEEE Communications Surveys & Tutorials, 21, 2772–2793, 2018.

Pucher, J. and Buehler, R.: Cycling towards a more sustainable transport future, Transport reviews, 37, 689–694, 2017.

Scott, D. M., Lu, W., and Brown, M. J.: Route choice of bike share users: Leveraging GPS data to derive choice sets, Journal of Transport Geography, 90, 102 903, 2021.

Seidl, D. E., Jankowski, P., and Tsou, M.-H.: Privacy and spatial pattern preservation in masked GPS trajectory data, International Journal of Geographical Information Science, 30, 785–800, 2016.

Shokri, R., Theodorakopoulos, G., Le Boudec, J.-Y., and Hubaux, J.-P.: Quantifying location privacy, in: 2011 IEEE Symposium on Security and Privacy, pp. 247–262, IEEE, 2011.

Song, Y., Dahlmeier, D., and Bressan, S.: Not so unique in the crowd: a simple and effective algorithm for anonymizing location data, in: PIR@SIGIR '14: Proceeding of the 1st International Workshop on Privacy-Preserving IR, pp. 19–24, 2014.

Strava: Strava announces Strava Metro, the largest active travel dataset on the planet, is now free and available to cities everywhere, available at: https://blog.strava.com/press/metro/?btn=4ngJq5kHx1SlNwIJ4IUG61&par=5WF6BYBmGyFYTCMihvXT4I, last access: 22 February 2023, 2020.

Sweeney, L.: k-anonymity: A model for protecting privacy, International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 10, 557–570, 2002.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., et al.: The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3, 1–9, 2016.