Mobility Data (knowledge discovery from)

AGNESE BONAVITA, Scuola Normale Superiore,Pisa, Italy
GIOVANNI COMANDE', Scuola Superiore Sant'Anna, Pisa, Italy

1 Introduction

The big data originating from the digital breadcrumbs of human activities, sensed as a by-product of the ICT systems, record different dimensions of human social life. These data describing human activities are valuable assets for data mining and big data analytics and their availability enables a new generation of personalized intelligent services. Most of these data are of sequential nature, such as time-stamped transactions, users' medical histories and trajectories. They describe sequences of events or users' actions where the timestamps make the temporal sequentiality of the events powerful sources of information. Unfortunately, such information often might unveil sensitive information that require protection under the legal frameworks for personal data protection. Thus, when such data has to be released to any third party for analysis, privacy-preserving mechanisms are utilized to de-link individual records from their associated users[1]. Privacy-preserving methods aim at preserving statistical properties of the data while removing the details that can help the re-identification of users. The challenge to researchers around the world is to share data without revealing private information of the users, and for that they need to protect the information using data anonymization techniques[2]. Several approaches provide a worst-case probabilistic risk of user re-identification as a measure for how safe the anonymised data is[3]. However, these solutions may work to make registered users anonymous, but they are insufficient for data combined attacks. After all, with reference to tracking apps for fighting covid-19, the EDPB clarified 'that location data thought to be anonymised may in fact not be. Mobility traces of individuals are inherently highly correlated and unique. Therefore, they can be vulnerable to re-identification attempts under certain circumstances'[4]

In this entry we discuss the nature of the mobility data in all these facets. Nowadays, mobility data include a set of data types with different origins and sources but that alone, or combined, give information on how an individual moves, where she usually goes and what activities she carries out. From a legal point of view, mobility data are not considered as such per se sensitive data (as health or political opinions data are) because they do not reveal sensitive personal information of the individual on their own as described in article 9 of the GDPR (like ethnic origin, sexual or religious preferences, political opinions, etc.). However, what we highlight is how apparently

---

1 See Anirban Basu, Anna Monreale, Juan Camilo Corena, Fosca Giannotti, Dino Pedreschi, Shinsaku Kiyomoto, Yutaka Miyake, Tadashi Yanagihara, and Roberto Trasarti. *A privacy risk model for trajectory data*. In Jianying Zhou, Nurit Gal-Oz, Jie Zhang, and Ehud Gudes, editors, Trust Management VIII, pages 125–140, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

2 See Tortelli Portela, Tarlis & Vicenzi, Francisco & Bogorny, Vania. (2019). *Trajectory Data Privacy: Research Challenges and Opportunities.*

3 See Benjamin C. M. Fung, Ke Wang, Rui Chen, and Philip S. Yu. *Privacy-preserving data publishing: A survey of recent developments.* ACM Comput. Surv., 42(4), 2010.

4 See EDPB *Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak Adopted on 21 April 2020*, quoting also Yves-Alexandre de Montjoye, Cesar A. Hidalgo Michel Verleysen & Vincent D. Blondel, Unique in the Crowd: The privacy bounds of human mobility, SCIENTIFIC REPORTS | 3 : 1376 | DOI: 10.1038/srep01376; *Apostolos Pyrgelis, Carmela Troncoso, Emiliano De Cristofaro, Knock Knock, Who's There? Membership Inference on Aggregate Location Data*, https://arxiv.org/pdf/1708.06145.pdf

unproblematic mobility data can become risky for privacy when they are combined or thoroughly analyzed with the relevant methods and/or external data. Indeed, even if they are not *-per se* sensitive personal data they may easily reveal sensitive and confidential information which need to be shielded. If data are mined appropriately, from mobility data it is possible to find out or infer not only the user behaviours and the places she visits, but also who the user is (initially anonymized), where she lives and what her health status might be. From the places regularly visited, often sensitive data can be inferred with a high degree of reliability. For instance, Sunday visits in a church or Friday ones to a mosque easily reveal religious beliefs as it does the presence at a political event for political opinions. From harmless data it is possible to build an identikit of anyone, and a deeply disturbing one both for its content and for its possible use. Note in fact that, although the GDPR does not apply to anonymous data (art. 2, 4 (1)) it is also true that the borderlines between anonymity and re-identification are progressively thinner. Indeed recital 26 clarifies that to 'determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly'.

This clarification is a key element of our journey because the 'principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable'. Such reasonableness needs to be ascertained considering 'all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments'.

2.1 Personal Data vs Sensitive Data

Personal data[5] protection legal rules represent the technical-juridical tool through which national and EU legislators protect all the rights connected to personal identity. A data is considered personal if it allows the identification of the individual (natural person) or if it describes the individual in such a way as to allow identification by acquiring other data. Both types of data are protected in the same way. With the term identification, therefore, we mean the possibility of distinguishing the person from any other subject (e.g., qualification as secretary of State) or within a category. If identification requires the acquisition of additional data for which unreasonable time and costs are required, then the person cannot be considered identifiable. Thus, data are not personal and the legal rules on personal data protection do not apply at all. However, it is not necessary to reach a high level of identification (let us think of the names that correspond to more than one person) for the data to be subject to protection. The European Union Court of Justice has developed a test for identifiability already under the EU Directive 95/46/EC, the so called Breyer test[6], clarifying that (at 43) 'it is not required that all the information enabling the

5 *Art 4.1 GDPR: 'For the purposes of this regulation: personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'*

6 See Case C-582/14 Breyer v Bundesrepublik Deutschland ECLI:EU:C:2016:779

identification of the data subject must be in the hands of one person')[7]. Thus, personal data is a dynamic concept, which must always be referred to the context, in the sense that even if an isolated information is not able to lead to the identification of an individual, such information could be used for identification through crossing with other data. This determines the nature of personal data. Hence the nature of personal data is not an absolute one, but it depends on 'all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly' (Recital 26 GDPR). What makes a 'means' reasonably expected to be used depends on many factors. As anticipated, recital 26 suggests that 'account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments'. Thus, the notion of 'personal data' depends on many variables.

Here we argue a similar path for sensitive personal data (called 'special category' by the GDPR at article 9) whose borders with non-sensitive data are fading away and apply it to mobility data. Whether or not a personal data turns into a sensitive one (belonging to the special category listed by art 9) depends on several factors. We move in this analysis along the lines theoretically set already adopting mobility data as a use case[8]. Authors defined 'quasi health data' those data useful to predict or determine the health status but that are not directly related to it[9]. We are going to see if the case of mobility data falls in the category. As a result, the notions we are elaborating upon are anonymous data, personal data, sensitive personal data and inferred data.

Personal data that has been rendered anonymous in such a way that the individual is not or no longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible with the caveats illustrated by recital 26. The notion it results in is dependent on, among else, security measures, the chosen architecture for ingesting and processing data, accessibility of data (connected or not to the internet). The GDPR protects personal data regardless of the technology used for processing them.

Examples of personal data are: a name and surname; a home address; an email address; an identification card number; location data. A special attention is given to the so-called sensitive data (special categories of data subject to more stringent rules[10]). These are categories of data that historically lent themselves to larger abuses against fundamental rights and freedoms (e.g., via discrimination). Their heightened protection aims at protecting the core values of our societies,

7 See also the model proposed by the WP29 acording to which the content, purpose, or result of the data processing must relate to an identifiable person either directly or indirectly (Article 29 Data Prot. Working Party, *Guidelines on the Right to Data Portability*, 16/EN, WP242rev.01, at 71).

8 See Gianclaudio Malgieri and Giovanni Comandé. Sensitive-by-distance: quasi-health data in the algorithmic era. Information Communications Technology Law, 26:1–21, 06 2017.

9 See Giovanni Comandé; 'The Rotting Meat Error: From Galileo to Aristotle in Data Mining?', in European Data Protection Law Review (2018), Volume 4, Issue 3, pages 270-277, at 271.

10 Art. 9.1 GDPR. 'Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited. Paragraph 1 shall not apply if one of the following applies: (a) the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject; (b) processing is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the field of employment and social security and social protection law in so far as it is authorized by Union or Member State law or a collective agreement pursuant to Member State law providing for appropriate safeguards for the fundamental rights and the interests of the data subject; (c) processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent'

human dignity and prevent possible discrimination. That is why society perceives them as more delicate than simple personal data and their processing is as a default rule a prohibition (art. 9 GDPR). The general principle is that their processing is prohibited unless one of the specific exceptional grounds apply. In addition, access to sensitive data should be limited through sufficient data security and information security practices designed to prevent unauthorized disclosure and data breaches. Article 9 GDPR lists the special categories of data considered 'sensitive'. They are 'personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.' Therefore, sensitive personal data is a specific set of 'special categories' inside the personal data context that must be processed with extra caution.

Our departing analytical point is that the borderlines between the two categories of personal data is more fluid than appears from the formal statutory definitions. There is data that might not be considered *prima facie* sensitive as such or even personal data, but that could produce sensitive information enabling either discovery or inference of personal and even sensitive data. These data apparently seem harmless but potentially conceal confidential information. The protection standard for this type of data is much less clear-cut. Note also that applicable legal regimes range from non-application of the GDPR to the application of its most stringent rules. Also note that in redefining the boundaries between personal and non-personal data and between personal data and special categories of personal data an important role is played by both the notion of inferred data[11] and the kind of attacks data can endure[12]. In this entry we concentrate on understanding how mobility data are classified from a legal point of view and how they should be classified according to how dangerous they can be for privacy attacks. More in details we illustrate that, although at first glance not considered personal (or at least sensitive) data, some mobility data can generate sensitive data or lead to infer sensitive data (without certainty on their accuracy and correspondence to reality) that are used to make decisions upon individuals ,impacting their rights. With reference to this last element a key point is the purpose of inferred personal data processing and the relevance of their (un)accuracy that can lead to serious violations of fundamental rights and plain violations of the core principles set in the GDPR.

2.2 Mobility data, a notion

Human mobility data is generally collected in an automatic way through electronic devices (e.g. mobile phones, GPS devices) in form of raw trajectory data. A raw trajectory of an individual is a sequence of records identifying the movements of that individual during the period of observation. Generally, there are two systems that monitor the location and movement of users; one is the location system and the other is the motion system[13]. In the following a quick overview of the most common types of mobility data is given:

---

11 *For the distinction among provided and observed data on the one hand, and derived and inferred data see Article 29 Data Prot. Working Party, Guidelines on the Right to Data Portability, 16/EN, WP242rev.01, at 9–11 (Dec. 13, 2016), https://ec.europa.eu/newsroom/document.cfm?doc_id=44099*

12 See F. McSherry, Statistical Inference considered harmful, https://github.com/frankmcsherry/blog/blob/master/posts/2016-06-14. md (2016);

13 See Choujaa, D. & Dulay, N. 2009. Activity Recognition from Mobile Phone Data: State of the Art, Prospects and Open Problems. Imperial College London, 1-32

- **Call Detail Records (CDR)**: Call Detail Records are records generated by a telephone exchange[14]. When an individual makes a call, the closest cellular network tower that routes the call is recorded, indirectly reflecting the user's geographic location.
- **Surrounding WiFi AP Records (SWFAPRs)**: Many mobile phones have embedded wireless and lots of urban areas have open Wi-Fi access points, which makes Wi-Fi positioning feasible by applying the received signal strength indication technique.
- **GPS Locations** : Global Positioning System provides geolocation information to a GPS user with at least four GPS satellites.
- **Geotagged Social Media (GTSM)**: Example GTSM data sources are Twitter and Foursquare. A geotagged tweet contains user identifier, text, and a location (latitude and longitude).

Some of the data mentioned above might not be considered personal data. However, the definition of personal data mentions also 'identifiers such as a name, an identification number, location data, an online identifier'. Thus, normally they are personal data, often perceived as non-dangerous ones and privacy-preserving. In this context we will make some considerations and comparisons between data of different nature to show how, by combining them in the right way they can allow to infer a lot of information about an individual. Above all, we want to demonstrate how some of these data can have an impact similar to sensitive data processing. Our demonstration will follow two paths. On the one hand, we illustrate the possibility to infer sensitive data from mobility data. On the other hand, we argue that, although these inferred data might be considered in a grey area in which data controllers' and data subjects' rights need to be balanced, their use as component of data controllers' activities can have a significant impact on rights and freedoms of data subjects. What matters is more the way in which personal data are used than their actual nature as sensitive data. Authors in [10] divide raw data as 'Received data' and 'Observed data'. The former are spontaneously provided by individuals (for examples when filling in a registration form or answering some questions) while the latter are collected by the data controller, after an individual's consent and through sensors, or a very simple combination of data (indirectly or passively provided by the data subject). Hence, we can start our reasoning considering raw mobility data as 'Observed data'. Note, however, that many mobility data to be observed require to be provided by individuals in a form or another. Also, unless their processing is necessary and limited to a specific aim with a legal basis (art. 6 or 9 GDPR) any further processing for different aims requires an appropriate legal basis that might not be found in further processing[15] . This qualification has several consequences, for instance triggering the specific conditions for consent and the right to portability. Conversely, complex data (data controller generated) are classified as 'Inferred data'

---

14 See Jinzhong Wang, Xiangjie Kong, Feng Xia, and Lijun Sun. Urban human mobility: Data-driven modeling and prediction. ACM SIGKDD Explorations Newsletter, 21:1–19, 05 2019.

15 Art.6.4 GDPR: 'Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives, the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia: (a) any link between the purposes for which the personal data have been collected and the purposes of the intended further processing; (b) the context in which the personal data have been collected, in particular regarding the relationship between data subjects and the controller; (c) the nature of the personal data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10; (d) the possible consequences of the intended further processing for data subjects; (e) the existence of appropriate safeguards, which may include encryption or pseudonymisation.

and 'Predicted data'. The first ones are descriptive data about the individual's past or present life inferred by an app via data mining or combination of raw data. The predicted ones are information produced from other data and referred to an individual's future behaviour applying a model developed on personal data eventually not referring to the same data subjects. For instance, a mobility data analysis revealing an individual has moved too fast on a trajectory to respect speed limits infers a traffic rule violation and from that a predisposition to violate –at least- traffic or safety rules. This information in building a risk profile applied to forecast the risk propensity of the same individual generates predicted data. If a number of individuals sharing the same living area are found with such a predisposition, a model can be created applying the same risk prediction to anyone sharing the same living area (e.g. by using a zip code as a trigger for the model). All of these categories present different levels of interplay with data protection, contract, tort and criminal law [10]. Their interplay also has different impacts on individual rights.

3. What can I infer with mobility data?

The quick evolution and wide diffusion of technologies for the localization of devices (especially smartphones and vehicles' GPS) as well as location-based services, is leading to the production and collection of large and diversified traces of human mobility, every day more detailed and pervasive. These traces potentially contain a huge amount of information that might allow inferring models of human mobility spaces at unprecedented levels of precision and depth. They would be key enablers of many applications, ranging from monitoring urban traffic features to reconstruct inter-city mobility demands and region scale structures, which could help in making modern urban spaces more sustainable, efficient and comfortable for citizens. They can also enable the monitoring of epidemics like the COVID-19. Starting from trajectory reconstruction (translating sequences of single location fixes into a full movement trajectory, possibly including map-matching) is possible to develop several methods for processing and analysing mobility data. In this section we would like to give an overview about mobility data and their potential to 'generate' sensitive personal data. Based on the level of data enrichment, it is possible to infer more and more information about individual users. Furthermore, the addition of semantics or external information (road conditions, weather conditions, etc.) makes it even easier to make predictions. We will start considering the so called 'Observed data' to see what happens enriching them with more information step by step, until arriving to have 'Inferred' or 'Predicted' data. There is a long way to go from raw data to useful representations of mobility behaviours: we can call it a mobility knowledge discovery process[16].

Raw data

Once you have an available mobility dataset what you really have is a sequence of points with a different sensitivity depending on the data type (if we have CDRs the spatial sensitivity is lower than the GPS one). Let us consider a dataset compound of GPS points which are spatio temporal points (longitude, latitude, and timestamp). The first step is to analyze the data to recognize and build the trajectories and paths taken by users. A strict definition of movements relates this notion to change in the physical position of an entity with respect to some reference

16 See Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Fabio Pinelli, Chiara Renso,Salvatore Rinzivillo, and Roberto Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. The VLDB Journal, 20:695–719, 2011.

system within which one can assess positions. A trajectory is a path made by the moving entity through the space where it moves. In studying movements, an analyst attends to several characteristics, which can be grouped depending on whether they refer to states at individual moments or to movements over time intervals. Moment related features include position in a particular moment, position of the entity in space, direction of the entity's movement, change of direction, speed of the movement[17]. Several mobility data sources also provide information about events of various types, detected by the device. They are usually related to acceleration and direction, or to events happening within the device: harsh acceleration, harsh braking, harsh cornering, multiple cornering, vehicle switch-on (start) and switch-off (stop). In some cases, the acceleration magnitude, the maximum acceleration, angle and duration are available too. Now suppose we can add semantic data to our dataset. For instance, if we consult a road map we could overlap the GPS path with the real streets in order to discover the geographic movements of the users. A road map is enough to start inferring knowledge: which are the most frequent routes, which are routine paths and which just occasional ones. Moreover, if we supposed to have also information about road conditions, the speed limits and the synchronization of the traffic lights we can define how a user drives or how a pedestrian moves in the city. An analogous reasoning could apply to wearable devices. The terms 'wearable devices' and 'wearables' all refer to electronic technologies or computers that are incorporated into items of clothing and accessories which can comfortably be worn on the body[18]. A wearable should have sensors for the physical environment such as location (for example GPS), cameras, microphones, temperature, humidity, movement, etc. A plethora of devices can be found in the market fitting in the previous definitions, but, despite all these options, the more adopted wearables today are wrist wearables, namely smartwatches. During the past decade, rapid progress in wearable sensor technologies eased long-term physical activity behaviour monitoring in real-life conditions. Among the existing sensors included in the wearable devices, three-dimensional (3D) accelerometers have gained the most attention. A 3D accelerometer measures acceleration forces in x, y and z dimensions, and therefore can sense the status of a body's motion or postures[19]. Combining GPS and accelerometer sensors has been useful in improving movement monitoring of humans, particularly in daily life. In the transport mode detection domain, the combination of GPS and accelerometer sensors is more useful than using each sensor individually, specifically in differentiating transport related activities such as walking, cycling and running.

We can categorize the use of GPS sensors into two broad applications. The first application mainly focuses on utilizing GPS spatial coordinates to link mobility behaviour derived from accelerometer data to the location and relevant spatial data such as land use, walkability, green spaces, neighbourhood and exposure in a geographic information systems environment[20]. These links enhance our contextual knowledge of the relationship between objectively measured physical activities and social environments[21]. The second application uses features such as time, distance, altitude, and speed derived from GPS data to inform classifiers in mobility detection. So, following

---

17 See F. Giannotti and D. Pedreschi. Mobility, Data Mining and Privacy: A Vision of Convergence, pages 1–11. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

18 See Tehrani, K. and Michael, A., 2014. Wearable technology and wearable devices: Everything you need to know. Wearable Devices Magazine.

19 Allahbakhshi, H.; Conrow, L.; Naimi, B.; Weibel, R. Using Accelerometer and GPS Data for Real-Life Physical Activity Type Detection. *Sensors* **2020**, *20*, 588.

20 See Maddison, R.; Jiang, Y.; Hoorn, S.V.; Exeter, D.; Ni Mhurchu, C.; Dorey, E. Describing patterns of physical activity in adolescents using global positioning systems and accelerometry. Pediatr. Exerc. Sci. 2010, 22, 392–407.

21 See Cebrecos, A.; Díez, J.; Gullón, P.; Bilal, U.; Franco, M.; Escobar, F. Characterizing physical activity and food urban environments: A GIS-based multicomponent proposal. Int. J. Health Geogr. 2016, 15, 35.

the examples mentioned above, it is easy to recognize how with few accessible information it is possible to find out a lot about an individual, entering his/her privacy area.

Trajectories and POIs

As mentioned, people perform movements in specific areas and time instants. These people are called users and each movement is composed by a sequence of spatio-temporal points *(x; y; t)* where *x* and *y* are the coordinates, while *t* is the timestamp. We call trajectory the sequence of spatio-temporal points which describe a movement[22]:

Definition 3.1 *A trajectory m is a sequence of spatio-temporal points* $m = \{(x_1, y_1, t_1), ..., (x_n, y_n, t_n)\}$ *where the spatial points* (x_i, y_i) *are sorted by increasing time* t_i , i.e., $\forall 1 \leq i \leq k$ we have $t_i < t_{i+1}$.

The set of trajectories travelled by a user makes her individual history:

Definition 3.2 *Given a user u, we define the individual history of the user as the set of trajectories travelled by him and denoted by* $M_u = \{m_1, .., m_k\}$.

Thanks to these two elements it is possible to enrich mobility data with annotations about human activities. These approaches are focused either on places of general interest (like restaurants, shopping centres) or on individual based destinations (like home or work) and yet they might lead to discover other individual destinations (e.g. clandestine meeting points for mistresses, political activities,...). The mobility history of a driver may enable many services such as location recommendation or sales promotion. In[23], by considering users' travel experience and the subsequent locations visited, the authors learn the location correlation from GPS trajectories useful to construct a personalized location recommendation system. In[24], the authors analyse urban mobility trying to feature the places in a city according to how people move among them. The authors build a network of points of interests by connecting places by the individual trajectories passing through them. An interesting analysis on mobility data[25] discovered two distinct classes of individuals: returners, whose mobility is produced by the commuting between home location and work location, and explorers, whose mobility is generated by travels performed toward locations different from home and work and far from them. This work shows that returners and explorers play a distinct quantifiable role in spreading phenomena and that there exists a correlation between their mobility patterns and social interactions. Hence, analysing the trajectories of individuals, it is possible to obtain a great deal of information. For every user, a data scientist can create a mobility profile that describes an abstraction in space and time of her systematic movements, ignoring exceptional paths. Thus, the systematic behaviour of every driver can be modelled with her mobility profile and the daily mobility of each user is characterized by her routines[26]. To give a

---

22 See R. Trasarti, Riccardo Guidotti, Anna Monreale, and Fosca Giannotti. Myway: Location prediction via mobility profiling. Information Systems, 64, 11 2015.

23 See Yu Zheng and Xing Xie. Learning location correlation from gps trajectories. In Proceedings of the 11th International Conference on Mobile Data Management, May 2010.

24 See I. R. Brilhante, M. Berlingerio, R. Trasarti, C. Renso, J. A. F. d. Macedo, and M. A. Casanova. Cometogether: Discovering communities of placesin mobility data. In 2012 IEEE 13th International Conference on MobileData Management, pages 268–273, July 2012.

25 See Luca Pappalardo, Filippo Simini, S Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-Laszlo Barabasi. Returners and explorers dichotomy in human mobility. Nat Commun, 6, 09 2015.

26 See Riccardo Guidotti, Anna Monreale, Salvatore Rinzivillo, Dino Pedreschi,and Fosca Giannotti. Unveiling mobility complexity through complex network analysis. Social Network Analysis and Mining, 6:1–21, 2016.

concrete example, in[27] authors used the mobility data to extract information about the stops nature of the drivers. For this purpose, they built a new method of trajectory segmentation able to recognize all the significant stops made by private vehicles. Indeed, most existing segmentation works use fixed thresholds that are global, i.e., the same threshold value applies to all the moving individuals, irrespective of any distinctive characteristics they might have. The authors tried to overcome these limitations providing a general methodology called Self-Adapting Trajectory Segmentation (ATS) that inspects the mobility of the individual and identifies segmentation thresholds that match her mobility features. With the right semantics it becomes easy to discover the purpose of each move and from that to trace even more private aspects of the targeted individual. In Figure 3.1 it is possible to see a clear example of it: we can see a common trip from South to North Italy. With the ATS method it is possible to recognize even very short stops (few tens of minutes) and by adding some geographical information it is easy to discover that it is a stop in a service area. The picture represents just a simple example, but it can be of use in many other situations attacking privacy of the users. Furthermore, with this method it is also possible, by studying the nature and the duration of the stops, to recognize the main locations of a user and the reason behind each stop.

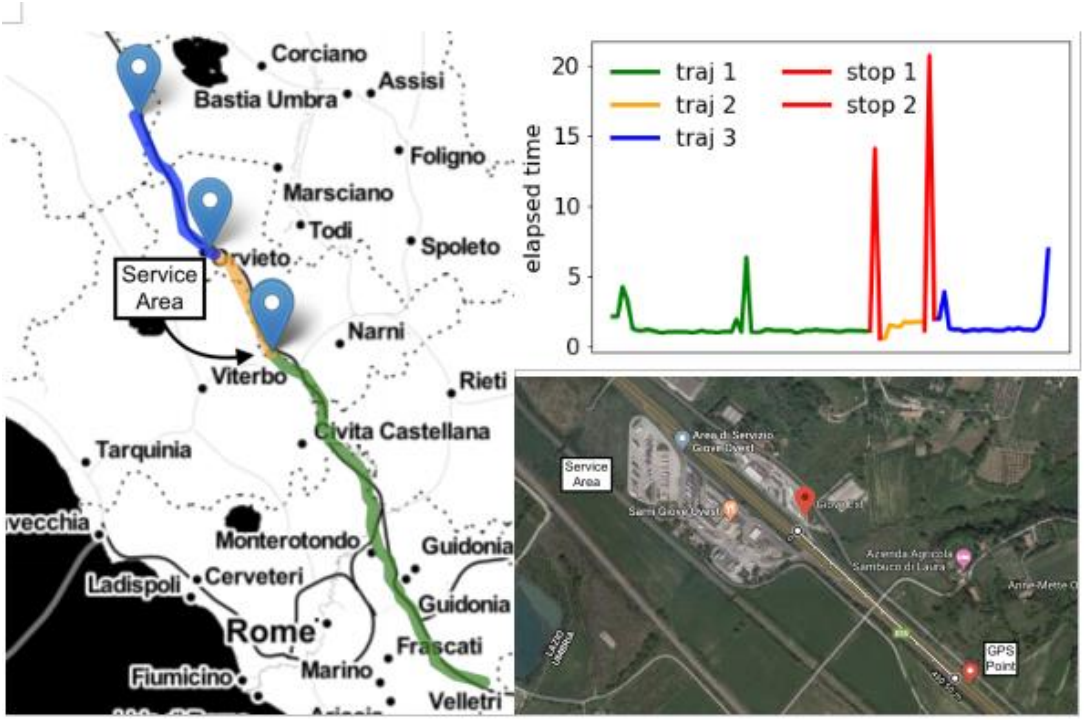

Figure 3.1 Trajectory segmentation returned by ats method. The user is traveling from South to North Italy. Top: spatial representation showing the trajectory segments. Center: temporal segmentation showing the interleaving time between GPS points. Bottom: zoom on the service area

[27] See Bonavita Agnese, Guidotti Riccardo, Nanni Mirco 'Self-Adapting Trajectory Segmentation.' EDBT/ICDT Workshops (2020).

highlighted in the top maps where the user probably stops for ~ 15 minutes. Best view in colour.

Instead, if we focus the attention not on the single user but we consider the collective aspect, having the trajectories it is possible to trace even the relationships among users. Indeed, using raw trajectories, we can first compute flocks and encounters, then from these encounters find a method to infer relationships. The basic idea is the following: if two users travel the same roads at the same time or attend the same places then they are likely to know each other. Moreover, counting how many times two people stay together (according to how many times their trajectories coincide) it is possible to build a hierarchy of relationship in order to understand the degree of relationship between two users. A similar work is found in[28]. This illustrates that already by considering the trajectories instead of the raw data allows us to reach a much higher level of inference and deduction turning mobility data into potential sensitive data. A conclusion already reached by the WP29[29]. The same WP29 in an earlier opinion concluded that special categories cover 'not only data which by its nature contains sensitive information . . . but also data from which sensitive information about an individual can be concluded.'[30]

Social Media Data

The introduction of location-based services in social media applications of smartphones has enabled people to share their activity related choices (check-in) in their virtual social networks (e.g. Facebook, Foursquare, Twitter etc.) providing unprecedented amounts of user-generated data on human movement and activity participation. This data contains detailed geo-location information, which reflects extensive knowledge about human movement behaviour. In addition, the venue category information for each check-in is recorded from which user activities can be inferred. If analysed properly, such data can help to better understand how citizens experience the cities they live in. Note that all these data are already from the outset personal data since they are linked to specific profiles. Also they can help identify mobility data which are not related to individuals by allowing the association of devices to individuals and to run cross-device associations[31]. Compared with other data sources, social media data has its unique characteristics such as more social information, which provides a multidimensional view of studying human mobility patterns. A direction to obtain accurate estimates of people's activities is to combine data from different sources, for example combining GPS data with geo-tagged social network data could be very useful to improve the data mining process knowledge[32]. The former data provide a sample of a user's whereabouts but are noisy and lack semantics, the latter provide visits to venues of exact locations, but they are not able to give information about the paths. There have been

---

28 See Andre Furtado, Areli Santos, Luis Alvares, Nikos Pelekis, and Vania Bogorny. Inferring relationships from trajectory data. 01 2015.

29 See 57Article 29 Data Prot. Working Party, *Opinion 03/2013 on Purpose Limitation*, at 47, 00569/13/EN, WP203 (Apr. 2, 2013), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf : '[m]ore often than not, it is not the information collected in itself that is sensitive, but rather, the inferences that are drawn from it and the way in which those inferences are drawn, that could give cause for concern.'. See also European Data Prot. Supervisor, *EDPS Opinion on Online Manipulation and Personal Data* at 5, 8–16, Opinion 3/2018 (Mar. 19, 2018), https://edps.europa.eu/sites/edp/files/publication/18-03-19_online_manipulation_en.pdf

30 Article 29 Data Prot. Working Party, Advice Paper on Special Categories of Data ('Sensitive Data'), at 6. See also Article 29 Data Prot. Working Party, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, 17/EN, WP251rev

31 Flavio Bertini, Stefano Giovanni Rizzo, Danilo Montesi, Can Information Hiding in Social Media Posts Represent a threat?, IEEE

32 See Riccardo Guidotti, Anna Monreale, Salvatore Rinzivillo, Dino Pedreschi,and Fosca Giannotti. Unveiling mobility complexity through complex network analysis. Social Network Analysis and Mining, 6:1–21, 2016.

extensive studies in mining geo-tagged social media data. For example[33] the authors analysed urban human mobility and activity patterns using location-based data collected from social media applications also exploring the frequency of visiting a place with respect to the rank of the place in individual's visitation records. Therefore, if there is the possibility to collect different kinds of mobility data, from raw data to social networks ones, we not only could predict individuals future behaviour but also reconstruct their personal information and relationships. Attackers may combine the data to identify the anonymized users invading the privacy area of everyone [34]. By following all these crumbs and connecting the dots any attacker could reconstruct the personal's file of everyone harnessing inferential analytics[35]. With wearable sensors data it is possible to make a similar reasoning since they are another type of data that is interesting to recall. As mentioned in the previous paragraphs wearable devices offer new opportunities to monitor human mobility activity continuously with the miniature wearable sensors embedded. However, there are few challenges faced on smartwatches about security issues which put users' safety and privacy at risk[36]. For instance, sensors as accelerometers, which is used to measure linear acceleration and it can determine whether the device is horizontal or vertical, and whether it is moving or not counting the steps a user takes, may hide several other functionalities. GPS sensors are integrated in wearable devices too, to locate a person's location and create a whole picture of her own mobility history. But this kind of sensor allows also to go beyond their primary purpose: for example, using accelerometers is possible to detect a range of activities including step counts, worn/not worn state, overall physical activity levels, eating behaviour, pill bottle opening movements, scratching, cardiopulmonary resuscitation (CPR) compression depth and frequency. A study[37] used a smartwatch accelerometer as compared to ground truth video to identify eating moments in 7 participants for a single day with 66.7% precision and 1 participant for 31 days with 65.2% precision. Besides that, in[38] authors used accelerometry to detect seizures in epilepsy patients and tremors. That is not all, since wearable devices also include sophisticated sensors specially designed to monitor health parameters which provide human activity measurement such as sleep quality, burned calories and other personal health metrics like heart-rate, body temperature, stress and hydration levels[39]. Wearables are collectors of a large set of confidential information in a way that allows to infer a lot about people's lifestyle and their own health status. Just to give a current example: in these days researchers and experts are fielding a new app that aims to exploit data extracted from smartwatches to prevent Covid-19 cases. In short, researchers want to develop a new remote computer model capable of carrying out a first screening in the monitoring activity of people positive to the Covid-19 virus of a large portion of the population. The most sophisticated smartwatches can measure oxygen saturation, heart rate and blood pressure, all important

---

33 See Samiul Hasan, Xianyuan Zhan, and Satish V. Ukkusuri. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, UrbComp '13, New York, NY, USA, 2013. Association for Computing Machinery.

34 See Giovanni Comandé, Regulating algorithms regulation? First ethico-legal principles, problems and opportunities of algorithms, in Tania Cerquitelli, Daniele Quercia, Frank Pasquale (eds), Towards glass-box data mining for Big and Small Data, Springer International, 2017, 169-207; Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter & Luciano Floridi, The Ethics of Algorithms: Mapping the Debate, BIG DATA & SOC'Y, July–Dec. 2016, at 1–2.

35 See Furnas,A.(2012). Everything you wanted to know about data mining but were afraid to ask. Retrieved from https://www.theatlantic.com/technology/archive/2012/04/everything-you-wanted-to-know-about-data-mining-but-were-afraid-to-ask/258538/;

36 See Ke Ching and Manmeet (Mandy) Mahinderjit Singh. Wearable technology devices security and privacy vulnerability analysis. International Journal of Network Security Its Applications, 8:19–30, 05 2016.

37 SeeThomaz, Edison & Essa, Irfan & Abowd, Gregory. (2015). A Practical Approach for Recognizing Eating Moments with Wrist-Mounted Inertial Sensing. Proceedings of the ... ACM International Conference on Ubiquitous Computing . UbiComp (Conference). 2015. 1029-1040. 10.1145/2750858.2807545.

38 See Lockman, Juliana & Fisher, Robert & Olson, Donald. (2011). Detection of seizure-like movements using a wrist accelerometer. Epilepsy & behavior : E&B. 20. 638-41. 10.1016/j.yebeh.2011.01.019.

39 See Kalantarian, Haik & Alshurafa, Nabil & Sarrafzadeh, Majid. (2015). Detection of Gestures Associated With Medication Adherence Using Smartwatch-Based Inertial Sensors. IEEE Sensors Journal. 16. 1-1. 10.1109/JSEN.2015.2497279.

parameters to be included in an Artificial Intelligence engine to build the risk profiles of the individual citizens. However, even in these moments, when the end seems to be able to justify any means, it is essential that the privacy rights of each person are preserved[40]. After considering all these data types, it is evident how the manipulation and the combination of this information can lead to obtain a whole picture of an individual's mobility. Starting from the raw data, which only supply the position of the subject, some singularities of the individual user can already be identified. With the adequate computational capacity, it is possible to analyse the data and recognize significant stops within the same trip for instance. After that, by adding semantics and recognizing the geographical areas, it is possible to understand the reason for the stops (a supermarket, take the children to school, go to the swimming pool, visiting a healthcare facility, etc.)[41].

We could offer many other examples to show the ease with which everyone could deal with this kind of data inferring sensitive personal data. Focusing the attention on mobility paths, trajectories and semantics of the territory enable to identify daily travel routes. Attackers may use trajectory data to deduce individual's mobility patterns and identify their home and workplaces or other 'special' ones. But that is not all: it is possible to reconstruct the individual behaviour and understand the relationships between users who travel the same roads or frequent the same points of interest. Even if the data are originally anonymized, if we know how a user moves, what places she attends, where she lives and works, it becomes immediate to go back to her identity. Anonymizing user identities is not enough to protect people's privacy. Then using social networks data any attacker could use the location tags (or hashtags) to verify the visit frequency of a given point of interest to correlate by matching people profiles and trajectory data to identify the users. They could also infer users' preferences, relationship, and personal habits. In conclusion, adding knowledge from wearable devices one can map the user, recognize the locations where she goes, the speed of her movements (how many steps, how many calories burned) and at the same time the heart rate, the percentage of oxygen usage and the hydration level.

Leaving aside for a moment the specific sensors that collect health data within smartwatches it is important to underline again that is possible to infer health information, or possible risks related to that, only from tracking and mobility sensors. Only by using mobility data we get to define the health status of any person. Thus, we could say that mobility data become quasi-health data ([10]) since we are able to infer users' health conditions from studying their movements. Even if mobility data are not inherently medical data, without the right protection level any attackers could easily lead to conclusions about individual's health conditions. Note also that the 'attacker' could be the legitimate data processor gathering the mobility data if she has a legal basis for such a further processing. We showed how the label 'sensitive data'' does not guarantee that there can be no privacy attacks using other data not tagged as such. It is necessary to identify the non-sensitive data that provide information with a high degree of confidentiality and which are equally risky for privacy protection. All this, if we consider a third-party attack that might lead to reidentification.

---

40 See EDPB Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 quoted

41 See Giovanni Comandé; 'The Rotting Meat Error: From Galileo to Aristotle in Data Mining?', in European Data Protection Law Review (2018), Volume 4, Issue 3, pages 270-277, at 271.

The situation gets more problematic once we consider the actual/potential uses of these inferred (sensitive) data by data controllers themselves. As we noted, location and mobility data are collected by different sources and often transferred to third parties who have other datasets of information enabling both reidentification and further data inferences by applying models and correlation patterns to the enriched mobility data. Again, regular presence on Fridays at a location corresponding to the address of a mosque easily leads to infer religious belief without the need to apply a complex model. Mobility data described above can trigger the application of models 'qualifying' the individual for specific features, health risks, for instance or sexual habits (e.g. recurrent passage and stops in an area of prostitution), religious habits. Once this qualification is obtained, what is relevant is its use, that is the actual application of the model with all the obvious implications in terms of decision-making. If a decision maker has to act upon a large number of individuals it might be satisfied with a certain degree of accuracy in the application of the model to the dataset triggering it. Recalling the previous example of the risk prone behavior, a zip code might become the data triggering the application of the model. In other words, 'users of data mining outputs could be willing to use these results although aware that the output might not be correct' [32]. Note that the WP29 has clearly identified as personal data those 'likely to have an impact on a certain person's rights and interests'[42]. Once the model applied to mobility data suggests a certain degree of health risk (e.g. developing diabetes, a risk prone driving attitude, what matters is not the fact that the suggestion can be considered 'data concerning health' but the actual use of this inferred information as such.

The emerging issues here can be characterized both in terms of ownership of the inferred information (to the data controller or the data subject) and in terms of accuracy of the information itself. On the latter, personal data needs to be accurate. The accuracy principle provided for by art. 5 .1.d requires that personal data are 'accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay'[43]. On the former, it has been questioned that inferred data are personal data at all[44]. However, both issues rely on the fact that such information is considered as personal data and must cope also with the impact on groups, not only individuals. Indeed, the ability to challenge conclusions deriving from inferred data is problematic for individuals, it is even more so when the application of inferences does not directly reach the individual level[45]. Noteworthy is the fact that individuals have little or no power on data made anonymous before creating the models applied to them unless specific legislation is triggered (e.g. antidiscrimination rules). For this reason the call to establish a 'right to reasonable inferences'' as a normative goal *de iure condendo*, although acceptable, lacks of bite. Once it is accepted that mobility data, although originally anonymous, can lead to identification and to reveal special categories of personal data pursuant to art. 9 GDPR ('data concerning health' for

42 *Article 29 Data Prot. Working Party, Opinion 4/2007 on the Concept of Personal Data, 01248/07/EN WP136, at 8 (June 20, 2007) http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf at 10–11 according to which derived and inferred data enjoy enjoy the full protection of individual rights enshrined in Articles 15–18 and Article 21 of the GDPR. See, however, for a more nuanced conclusion on inferred data as personal data Cases C-141/12 & 372/12, YS, M and S v. Minister voor Immigratie, Integratie en Asiel, 2013 E.C.R. I-838. Yet see the later Novak case (169Case C-434/16, Peter Nowak v. Data Prot. Comm'r, 2017 E.C.R. I-994, par. 34) according to which personal data include those 'in the form of opinions and assessments, provided that it 'relates' to the data subject.'*

43 See Omer Tene & Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 11 NW. J. TECH. & INTELL. PROP. 239, 270 (2013)

44 See Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Columbia Business Law Review., passim.

45 See Wim Schreurs. Mireille Hildebrandt, Els Kindt, Michaël Vanfleteren, Cogitas, Ergo Sum. The Role of Data Protection Law and Non-Discrimination Law in Group Profiling in the Private Sector, in PROFILING THE EUROPEAN CITIZEN 241, 246 (Mireille Hildebrandt & Serge Gutwirth eds., 2008).

instance) a higher level of protection can be channelled by art, 35 GDPR. It imposes a data protection assessment ('DPIA'), with consequent actions, every time 'a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons'. A DPIA is especially required in case of ‘a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person'.

Pursuant to the previous analysis, every time inferred data triggers the application of a model producing 'legal effects concerning the natural person or similarly significantly affect the natural person', especially when based on automated processing would impose a DPIA with its characteristics. As described by art 35.7 this must include '(a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller; (b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes; (c) an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and (d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.' Once the implication of inferred data leads to generalize a duty to perform a DPIA, the recommended publicity of the DPIA forces upon data controllers the adoption of appropriate safeguards and information duties, expanding the protection potentials of the GDPR even before a formal recognition of a right to reasonable inferences.

4. Summary remarks

In this entry we discussed the risk coming from mobility data to people’s privacy and to give further content to the claim that in digital society 'anonymity' is a computational concept as stressed by some authors and clearly endorsed by recital 26 of the GDPR. Oftentimes mobility data are not considered as sensitive data, so they apparently do not fall under art.9 of GDPR nor they entails heightened safeguards. However, as we demonstrated in certain situations they easily become as risky as those considered sensitive or even more since there is much little awareness of the inferences that can be drawn from them.

We explained why more attention is needed on those data that are not considered a priori sensitive but that instead can easily entail serious consequences for data subjects if processed in the appropriate way. Adding together information, from different sources, about people movements (routine paths, points of interest, social network tagged information) any attacker can trace the profile of an individual and re-individualize the dataset. An attacker, who might not necessarily be a criminal but just an interested player on the market, could also map data subjects’ movements, recognize their habits and preferences (social, religious, etc.) and even get to infer about their present and future state of health. It is true that those would be inferences or at best predictions based on models developed with machine learning and questioned, as such, as real personal data. Nevertheless, would it make a difference if market players or malevolent attackers act upon them

anyway? For these reasons, we argued that a first layer of protection is offered by a clear application of art. 35 GDPR. Intuitively, it applies to most instances of massive inference of (sensitive and non sensitive) data from mobility data and to the application of the models they enable to generate irrelevant any issue of truthfulness, verifiability or intentionality in their use [36]. Indeed, this is a first step in need of further research.

Selected Bibliography

[1] Anirban Basu, Anna Monreale, Juan Camilo Corena, Fosca Giannotti, Dino Pedreschi, Shinsaku Kiyomoto, Yutaka Miyake, Tadashi Yanagihara, and Roberto Trasarti. A privacy risk model for trajectory data. In Jianying Zhou, Nurit Gal-Oz, Jie Zhang, and Ehud Gudes, editors, Trust Management VIII, pages 125–140, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[2] I. R. Brilhante, M. Berlingerio, R. Trasarti, C. Renso, J. A. F. d. Macedo, and M. A. Casanova. Cometogether: Discovering communities of placesin mobility data. In 2012 IEEE 13th International Conference on MobileData Management, pages 268–273, July 2012.

[3] Ke Ching and Manmeet (Mandy) Mahinderjit Singh. Wearable technology devices security and privacy vulnerability analysis. International Journal of Network Security Its Applications, 8:19–30, 05 2016.

[4] Andre Furtado, Areli Santos, Luis Alvares, Nikos Pelekis, and Vania Bogorny. Inferring relationships from trajectory data. 01 2015.

[5] F. Giannotti and D. Pedreschi. Mobility, Data Mining and Privacy: A Vision of Convergence, pages 1–11. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[6] Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Fabio Pinelli, Chiara Renso,Salvatore Rinzivillo, and Roberto Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. The VLDB Journal, 20:695–719, 2011.

[8] Riccardo Guidotti, Anna Monreale, Salvatore Rinzivillo, Dino Pedreschi,and Fosca Giannotti. Unveiling mobility complexity through complex network analysis. Social Network Analysis and Mining, 6:1–21, 2016.

[9] Samiul Hasan, Xianyuan Zhan, and Satish V. Ukkusuri. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, UrbComp '13, New York, NY, USA, 2013. Association for Computing Machinery.

[10] Gianclaudio Malgieri and Giovanni Comandé. Sensitive-by-distance: quasi-health data in the algorithmic era. Information Communications Technology Law, 26:1–21, 06 2017.

[11] Luca Pappalardo, Filippo Simini, S Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-Laszlo Barabasi. Returners and explorers dichotomy in human mobility. Nat Commun, 6, 09 2015.

[12] R. Trasarti, Riccardo Guidotti, Anna Monreale, and Fosca Giannotti. Myway: Location prediction via mobility profiling. Information Systems, 64, 11 2015.

[13] Jinzhong Wang, Xiangjie Kong, Feng Xia, and Lijun Sun. Urban human mobility: Data-driven modeling and prediction. ACM SIGKDD Explorations Newsletter, 21:1–19, 05 2019.

[14] Yu Zheng and Xing Xie. Learning location correlation from gps trajectories. In Proceedings of the 11th International Conference on Mobile Data Management, May 2010.

[15]Choujaa, D. & Dulay, N. 2009. Activity Recognition from Mobile Phone Data: State of the Art, Prospects and Open Problems. Imperial College London, 1-32

[16] Tehrani, K. and Michael, A., 2014. Wearable technology and wearable devices: Everything you need to know. Wearable Devices Magazine.

[17] Allahbakhshi, H.; Conrow, L.; Naimi, B.; Weibel, R. Using Accelerometer and GPS Data for Real-Life Physical Activity Type Detection. *Sensors* **2020**, *20*, 588.

[18] Maddison, R.; Jiang, Y.; Hoorn, S.V.; Exeter, D.; Ni Mhurchu, C.; Dorey, E. Describing patterns of physical activity in adolescents using global positioning systems and accelerometry. Pediatr. Exerc. Sci. 2010, 22, 392–407.

[19] Cebrecos, A.; Díez, J.; Gullón, P.; Bilal, U.; Franco, M.; Escobar, F. Characterizing physical activity and food urban environments: A GIS-based multicomponent proposal. Int. J. Health Geogr. 2016, 15, 35.

[20] Thomaz, Edison & Essa, Irfan & Abowd, Gregory. (2015). A Practical Approach for Recognizing Eating Moments with Wrist-Mounted Inertial Sensing. Proceedings of the ... ACM International Conference on Ubiquitous Computing . UbiComp (Conference). 2015. 1029-1040. 10.1145/2750858.2807545.

[21] Lockman, Juliana & Fisher, Robert & Olson, Donald. (2011). Detection of seizure-like movements using a wrist accelerometer. Epilepsy & behavior : E&B. 20. 638-41. 10.1016/j.yebeh.2011.01.019.

[22] Kalantarian, Haik & Alshurafa, Nabil & Sarrafzadeh, Majid. (2015). Detection of Gestures Associated With Medication Adherence Using Smartwatch-Based Inertial Sensors. IEEE Sensors Journal. 16. 1-1. 10.1109/JSEN.2015.2497279.

[23] Tortelli Portela, Tarlis & Vicenzi, Francisco & Bogorny, Vania. (2019). Trajectory Data Privacy: Research Challenges and Opportunities.

[24] Bonavita Agnese, Guidotti Riccardo, Nanni Mirco 'Self-Adapting Trajectory Segmentation.' EDBT/ICDT Workshops (2020).

[25] Benjamin C. M. Fung, Ke Wang, Rui Chen, and Philip S. Yu. Privacy-preserving data publishing: A survey of recent developments. ACM Comput. Surv., 42(4), 2010.

[26] Case C-582/14 Breyer v Bundesrepublik Deutschland ECLI:EU:C:2016:779.

[27] F. McSherry, Statistical Inference considered harmful, Seehttps://github.com/frankmcsherry/blog/blob/master/posts/2016-06-14. md (2016);

[28] Flavio Bertini, Stefano Giovanni Rizzo, Danilo Montesi, Can Information Hiding in Social Media Posts Represent a threat?, IEEE

[29] Giovanni Comandé, Regulating algorithms regulation? First ethico-legal principles, problems and opportunities of algorithms, in Tania Cerquitelli, Daniele Quercia, Frank Pasquale (eds), Towards glass-box data mining for Big and Small Data, Springer International, 2017, 169-207; Brent Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter & Luciano Floridi, The Ethics of Algorithms: Mapping the Debate, BIG DATA & SOC'Y, July–Dec. 2016, at 1–2.

[30] Furnas,A.(2012). Everything you wanted to know about data mining but were afraid to ask. Retrieved from https://www.theatlantic.com/technology/archive/2012/04/everything-you-wanted-to-know-about-data-mining-but-were-afraid-to-ask/258538/;

[31] EDPB Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 quoted

[32] Giovanni Comandé; 'The Rotting Meat Error: From Galileo to Aristotle in Data Mining?', in European Data Protection Law Review (2018), Volume 4, Issue 3, pages 270-277, at 271.

[33] Omer Tene & Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 11 NW. J. TECH. & INTELL. PROP. 239, 270 (2013)

[34] Wim Schreurs. Mireille Hildebrandt, Els Kindt, Michaël Vanfleteren, Cogitas, Ergo Sum. The Role of Data Protection Law and Non-Discrimination Law in Group Profiling in the Private Sector, in PROFILING THE EUROPEAN CITIZEN 241, 246 (Mireille Hildebrandt & Serge Gutwirth eds., 2008).

[35]Giovanni Comandé; 'The Rotting Meat Error: From Galileo to Aristotle in Data Mining?', in European Data Protection Law Review (2018), Volume 4, Issue 3, pages 270-277, at 271.

[36] Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Columbia Business Law Review., passim.