Privacy Beyond Confidentiality, Data Science Beyond Spying: From Movement Data and Data Privacy Towards a Wider Fundamental Rights Discourse

Bettina Berendt(✉)

Department of Computer Science, KU Leuven, Leuven, Belgium
bettina.berendt@cs.kuleuven.be,
https://people.cs.kuleuven.be/~bettina.berendt/

Abstract. Although privacy and AI/data science are multi-faceted concepts, there is an increasing trend to focus on only a subset of their meaning: privacy as data privacy, with a focus on confidentiality, and AI/data science as a threat to autonomy and privacy, through data collection, unwanted inferences, and profiling. However, confidentiality and "invisibility" are not always constitutive of privacy as "the freedom from unreasonable constraints on the construction of one's own identity" – in some cases, visibility can be more important, and data collection, presentation, and inferences can help and extend a desired visibility. In this position paper, I will focus on a specific application around these phenomena: the analysis of vehicle/human trajectory data. I will discuss two recent examples of the analysis of such data: the New York City taxi rides dataset, and the use of data from the maritime Automatic Information System (AIS) for mapping refugee movements on the Mediterranean Sea. The goal is to encourage a discussion as to whether and how such wider fundamental-rights questions and their implications for privacy, data protection, and technology can and should be investigated in the scope of APF.

Keywords: Modelling of data protection and privacy requirements · Aspect of privacy in artificial intelligence · Privacy and other fundamental rights · Privacy of location and trajectory data

1 Introduction

"Privacy is a contested notion" used to be a stock phrase in presentations and papers throughout the nineties and noughties, and a vast number of classifications of different notions and aspects of privacy have been proposed, see [14,16,30] for just three examples. Data protection is a similarly multi-faceted concept – not only a counterpart right to the right to privacy [8], but also a means

M. Naldi et al. (Eds.): APF 2019, LNCS 11498, pp. 59–71, 2019.
https://doi.org/10.1007/978-3-030-21752-5_5

for protecting "fundamental rights and freedoms" in general (Article 1 GDPR).[1] Finally, computer technology in general and Artificial Intelligence applications in particular play many roles, ranging from the quintessential threat of profiling (an activity that even entered the title of an article in the GDPR, Article 22) to the mandate, in the GDPR, to use state-of-the-art (privacy-enhancing) technologies: Articles 25(1) and 32(1) require controllers and processors to give due regard to the state of the art when choosing the technologies.

Yet, in spite of this richness of meaning in the three concepts of "privacy", "data protection", and "AI", there is a prototype: (1) Privacy is operationalised via data privacy and therefore obtains when information is hidden from all or at least specific others; it is therefore centered around confidentiality. (2) Data protection obtains when this information is removed in appropriate ways (e.g. through anonymisation or pseudonymisation) or at least restricted to its intended recipients (e.g. through access control), cf. for example [13] or, specifically for trajectory data, [29]. (3) AI and – particularly relevant when it comes to the processing of personal data – data science[2] are conceptualised as dangers to individuals and their autonomy, by combining unwanted data collection with intransparent inferences and manipulation (as exemplified in the Cambridge Analytica media narrative).

The present paper starts from reconceptualising privacy as more than confidentiality. It goes back to the alternative of "privacy as the right to be let alone" as formulated by Agre and Rotenberg: "the freedom from unreasonable constraints on the construction of one's own identity" [1], also expanded on by Hildebrand [17]. Crucially, as feminist scholars and others have pointed out, keeping something confidential or *invisible* can serve to perpetuate oppression and therefore counteract the very liberatory effects that a private sphere is supposed to have, cf. [27]. On the contrary, it may be necessary to make certain information *visible* in order to fight and overcome oppression and oppressive structures. The standard example used to be the treatment of domestic violence as a "private matter" vs. its publication and the legal and regulatory successes this has enabled; a current example is the #metoo movement.[3]

[1] Of course, the GDPR also contains and elaborates on many other principles, including requirements on data processing related to IT security (integrity and availability in addition to confidentiality, Article 32 and Article 5(1)(f)), accountability, weighing of interests, and others.

[2] A field situated in the intersection of machine learning (as a part of AI) and substantive expertise [7].

[3] This movement started as the encouragement of victims of sexual harassment (especially in, but not limited to, the workplace) to tweet about their experiences and give people a sense of the magnitude of the problem. The Twitter hashtag #metoo simplifies the retrievability of these reports, such that specific incidents as well as patterns of sexual harassment become a public and visible phenomenon, rather than remain "private" singular experiences. As in the case of calling out domestic violence, the hope is that "this 'mainstreaming' of feminist activism is laying the foundation for a collective shift towards a more just society" [24, p. 239].

This reconceptualisation will be done by contrasting two case studies that on the surface share many commonalities: human trajectories that can be derived from observed vehicle movement data and data science studies that reconstruct trajectories from these data. The case studies are (1) the New York City taxi rides dataset, and (2) the use of data from the maritime Automatic Information System (AIS) for mapping refugee movements on the Mediterranean Sea. In both cases, the data amount to "holistic trajectories", spatiotemporal data enriched with semantic information about the vehicles, the space, the voyage, and enrichable (through data-science inferences) with information about the people on that trajectory.

The first contribution of the paper is to investigate claims that have been made with regard to privacy protection in the second case study, and to argue that, unlike in the first case study, invisibility is often *not* what the affected individuals want. In their case, rather, *visibility* becomes a precondition for having rights and often life at all. Data science projects that support this goal and a counter-narrative to politically prevalent narratives, can then become tools that may further fundamental rights (rather than threaten them, as in the default narrative). Data protection law, in turn, may or may not be applicable, and in any case probably not conceptualised as in standard GDPR-related discussions. The second contribution is to highlight some possible questions that can be asked of the data and their presentation.

The paper is a position paper, a question and a proposal. It asks the question whether and how the APF community wants to engage with the highly politically charged topics around migration, data, and fundamental rights. It proposes a number of (technical and social) questions as a starting point to such an engagement. Lastly, it is (obviously) the opinion of the current author that this is a discussion worth having at APF.

2 Case Study 1: New York City Tax Rides Dataset

In 2014, the City of New York released, in response to a Freedom of Information request, data about all 173 million taxi rides in New York in 2013, with the taxi identifiers pseudonymised, and exact spatiotemporal data about start- and endpoints, as well as fares, given. This dataset provided a rich real-life dataset for a wide range of data mining studies, such as "optimization of the revenue of NYC Taxi Service using Markov Decision Processes" [21]. At the same time, the publication of the dataset was soon criticized on privacy grounds. For example, the taxi pseudonyms could easily be re-identified to their actual medallion numbers [26]. It was also argued that the data allowed inferences towards sensitive attributes of the taxi drivers, such as the patterns of breaks during the day indicating that someone is a devout Muslim [35]. Finally, with some background knowledge, inferences can be made towards the identity of taxi customers, and based on that, details about their whereabouts learned [3]. The futility even of better pseudonymisation/anonymisation approaches was demonstrated by Douriez et al. [12]. Medallion and driver license IDs were removed from NYC's taxi datasets released in subsequent years.[4]

The taxi rides represent a typical case of personal data in the sense of the GDPR. Personal data are "any information relating to an identified or identifiable natural person ('data subject')" (Article 4). Since at least some, and likely many, taxi drivers and taxi customers are easily identifiable, the dataset contains personal data. Taxi customers (and conceivably also taxi drivers) had not been asked to give their consent to these data being published online for unspecified purposes, nor are other grounds for such processing (Article 6 GDPR) present. This is textbook privacy violation by data[5] (more accurately in the EU context: a violation of data protection law). While the GDPR defines a number of exemptions for research, it does so under conditions [19,22], such that currently ethics boards in EU universities are cautious and therefore discourage the use of this dataset for any kind of data mining.[6]

This perception of a dataset assumes that the population of data subjects consists of informed individuals, who exercise their autonomy among other things by travelling in vehicle passages they pay for, and who have a reasonable expectation of privacy in doing so that requires that the data about their movements remain confidential. The main question for the responsible data scientist appears to follow from the observation that the removal of taxi identifiers "would adversely impact certain types of analysis on the data" [12, p. 148] and the need to find different analysis types.

3 Case Study 2: AIS Data for Describing Migrant Rescue Operations

The second case study is based on the Hoffmann et al. study published in a 2017 report by the IOM [18], the UN International Organization for Migration, and illustrated in an interactive and multimedia online presentation[7]. As in case

[4] https://data.cityofnewyork.us/browse?q=taxi.

[5] It is debatable where/when the violation occurs. Opinions differ as to whether the existence of knowledge about individuals per se represents a privacy violation, whether this only occurs when this knowledge is acted upon, or whether the publication of data as an enabler of such consequences already forms a privacy violation [4].

[6] This information was given to me under conditions of confidentiality, as was the assessment that university ethics boards tend to be conservative in their interpretation of the GDPR. The publicly available university documents that I have seen on what is and what is not allowed regarding the re-use of public datasets, do not address specific questions such as "is it allowed to re-use public datasets", rather, they refer to the general principle that GDPR compliance always also depends on the whole context of research – which is of course a correct rendering of a law that requires interpretation in context. Even if I therefore cannot provide a reference for my claim, I consider it worthwhile to mention it, for example to encourage discussion among researchers about their respective institutions' GDPR handling.

[7] http://rescuesignatures.unglobalpulse.net/mediterranean/.

study 1, the base data are in principle publicly accessible. They are data from the Automatic Information System (AIS), a maritime communications system through which vessels regularly broadcast information, including their identifier, vessel type, latitude and longitude, speed, course and destination. The information is used by maritime authorities and ships to locate nearby vessels and avoid collisions. Based on these spatiotemporal data and enriched with textual and pictorial data from other sources[8], the authors generate a type of holistic trajectories, manually label them as representing (or not) a rescue operation, and use clustering and machine learning with a view to classification and prediction. Other researchers have investigated how to model and detect such trajectories. Based on AIS data, complex events, including but not limited to SAR (search and rescue) missions, and involving one or several vessels, can be modelled and detected efficiently and in real time using combinations of exploratory, machine learning, and logics-based (event calculus) techniques [28,36].

Hoffmann et al. mention several limitations of their method, mainly with regard to data quality, including the fact that as circumstances change, so do the data and patterns (thus, the analysis of timely data is crucial).

In a section on "privacy", the authors raise several points. The first is a reference to concerns over port security as a consequence of AIS data public availability. The second is the possibility that rescue organisations may not want the full details of their operations to be publicly known, because they are facing opposition and threats (a European far-right group threatening to attack rescue vessels is mentioned). Both concerns are not privacy concerns in the sense of European law (in particular because the agent requesting the confidentiality is not a natural person). As a third reason, the authors mention that "adversarial users could take advantage of the data to track the location of individual refugees [identified by record linkage with data such as photos or statements, or other background knowledge], attack rescue boats or guide piracy operations" (p. 40). Presumably, the attacks and piracy operations are security/safety concerns for the rescue vessels, their crews, and the rescued persons, and these concerns could arise from the public availability of the data as well as from possible predictors learned from them, i.e. the data scientists' work.

It also appears, from the sentence, that the possible tracking of individuals is considered a security/safety risk (because it could lead to attacks) rather than a typical privacy risk (by which an individual migrant would want to keep their identity or properties hidden). It is difficult to say what role such expectations of, or wishes for, privacy in our usual sense, play in this extreme situation. Also, it has been observed increasingly over the past years that rather than trying to hide their voyage, "migrants from Libya facilitated their traceability by national authorities and monitoring systems, anticipating in space and time border patrols by sending an SOS as soon as they entered the international waters"

[8] The authors enrich the data with broadcast warning data produced by WWNWS, a global service managed by the UN Maritime Organization IMO (data that appear to cover only a small fraction of vessels in distress, p. 37), and other data such as the tweets issued by NGO vessels.

[33, p. 576]. In other words, along their journey, migrants deal strategically with visibility and invisibility, with information disclosure and hiding/confidentiality. This is quite probably a very rational strategy given the fact that a successful and invisible journey to Europe is by now nearly impossible for many reasons, including that traffickers severely overload and under-equip their vessels, and that due to the high-resolution sensors employed in the European Border Surveillance System EUROSUR [9], even very small vessels are likely to be spotted and monitored. Strategic information disclosure (in addition to strategic information hiding) by individuals can also be observed in many other contexts that are less dramatic than the life-or-death situations faced by migrants on the Mediterranean, and it has been pointed out that strategic information disclosures too can be privacy-related behaviour [15].

A second question related to privacy is related to the referent of the data. Technically a ship's trajectory could be considered personal data in the same sense as a taxi's trajectory. (This concerns the ships provided by the traffickers as well as rescuing ships once they have been boarded by migrants.) As for taxis, the trajectory is a trajectory both of the "driver" and of the "passenger(s)".

In the NGO vessel case, the "driver" individuals are the captain and crew members. To the extent that they can be re-identified using public (or otherwise procured) records, the AIS-based trajectory data form personal data. However, their personal and professional mission is to carry out rescue tasks, and to do so in a transparent manner, and they in fact often seek visibility and publicity (for their funders as well as a political statement). It thus appears less likely that these individuals would regard the publication of the information that they were at some location at some time as a violation of their privacy (even if for security/safety reasons, they may prefer some degree of invisibility, see above). "Drivers" of non-NGO vessels such as cargo ships are likely to have other motivations, since their original task is not related to sea rescue, which may make them regard their location data differently.

As regards the "passengers", with appropriate background knowledge, similar re-identification attacks could in principle be mounted to identify individuals. These could for example be based on photos taken of individuals while on-board or disembarking, matched with named photos as background knowledge [18]. It is also conceivable that data regarding the captain or crew members and data regarding migrants are combined, and that this may result in undesired consequences. It is an open question whether such attacks are likely.[9] If such a re-identification link is not made, or is very unlikely to be made, AIS-based trajectories of rescuing ships may not count as personal data.

However, even if *individuals* may not be exposed in a traditional privacy-violation sense, there is a much more likely sense in which migrants are exposed by AIS data: as a *group*. In fact, as has been argued in this context [33] as well as in connection with other applications of big data analyses to humanitarian

[9] How likely they are will depend on the existence of background knowledge and the existence of and incentives for "attackers" (paparazzi, celebrity fans, law enforcement, criminals, ...).

causes [31], there is a temptation to focus on migrants as a group defined only by one feature (here: to be in need of rescue).

The fact that big data constitute new risks in the profiling of groups has been lamented often in connection with data protection laws such as the GDPR (which focus on the protection of individuals' rights and freedoms); in the humanitarian realm, it creates additional and different challenges [32].

For the data scientist, this means that also the response to these risks and threats may need to be very different, because traditional approaches to (for example) anonymisation are focussed on the protection of individuals from threats against these persons as individuals. It is an open research question what could constitute effective measures of group protection.

Data privacy, viewed technically, does not need to make a clear distinction between protecting information and control over it related to individuals (a concept rooted in human rights) and protecting information and control over it related to other entities (such as organisations, the NGOs in the current example and in the argument made by Hoffmann et al.) [11]. In data privacy, a different and independent dimension becomes relevant when one asks "whose privacy" should be protected. A useful distinction is that between data owner, data respondent (the data subject, although not always in its legal sense of an individual person), and data user; and this distinction has implications for the choice of data-privacy protection methods [11]. In the present example, one assignment of these roles that follows the argument about risks above could be: the NGO as the data owner, the migrant (or migrant group) as the data respondent, and various (potential) data users: the public, politicians, pirates, ...

Moving beyond privacy and data privacy, many other questions, technical as well as ethical, arise about information disclosure and hiding. The study and visualisation of "rescue patterns" can have different objectives. Hoffmann et al. mention operational objectives (e.g., supporting coordination of rescue operations), analytic objectives (e.g., determining conditions under which rescues are most effective), and reporting objectives. The latter are described as follows: "supplement the large amount of qualitative, descriptive coverage already produced by NGOs and the news media", "help external observers ... obtain a high-level picture of what is happening in the region over time. An overview of these patterns is critical for coordination and advocacy purposes; it enables stakeholders to see the true magnitude of rescue operations, and to quantify costs, shortcomings and future needs." [18, p. 30].

Concentrating on the reporting objective, it can be argued that rescue patterns constitute a counter-mapping practice: in the EUROSUR monitoring system, selected migratory events are produced from the sensed data and mapped in time and space [33]. The website watchthemed.net, initiated and run by a network of NGOs, activists and researchers, maps events to monitor deaths and violations of migrants' rights. In the SAR-centric applications described here, rescue events are produced and mapped. EUROSUR is run by Frontex, and its data and analytics are not available to NGOs and other external partners, whereas the rescue patterns are mined from data available publicly (AIS data) or available to partners of the research (the broadcast warnings), and enriched with further aspects from public data (such as tweets).

Mapping practices generate a narrative around their real-world phenomenon. The current data models and visualizations of rescue patterns, maybe for technical reasons (because the EUROSUR data are not available), maybe to avoid visual clutter, display these patterns in an otherwise "empty" space. Is it possible, and is it advisable, to at least represent that far more data exist (even if one does not have access to them)? In other words, should the "known unknown" data be modelled and represented too, and if so, how? These data are important for technical reasons as much as for narrative reasons – how can and should these two motivations be addressed, and how can the choices made be made in a transparent and accountable way? In the following paragraphs, I will illustrate three examples of these considerations.

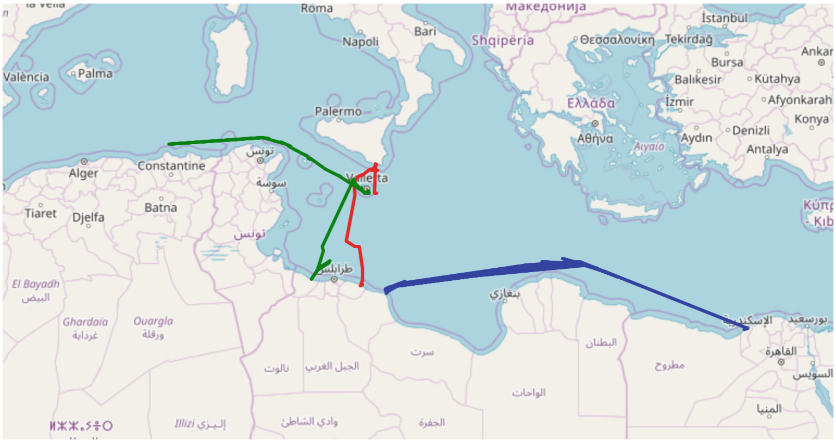

Fig. 1. The Alexander Maersk's June 2018 trajectory (in red, via Valletta). (Color figure online)

First, sometimes trajectory data illustrate very directly the influences of context and the uncertainty and the "unknowns" of vessel operators. As an example, consider the recent case of a commercial cargo ship that took on 113 people saved by an NGO rescue ship and then spent four days in a political stand-off on a zig-zag trajectory between ports before being allowed to dock in Sicily [2,5], see Fig. 1.[10] Can and should holistic trajectories measure and visualize the enormous costs caused by such decisions, as well as the incentives and influences this may have on further behaviour by vessel operators? What about similar odysseys that have since taken place in a politically more and more charged climate, such

[10] I thank Konstantinos Tserpes for mentioning this example and making available a visualization of the trajectory.

as those involving a coastguard and a sea-rescue NGO ship respectively [20,38]? What about, reversely, the trajectories of ships that could not and were not 'doing anything anymore' under these circumstances, with trajectories (enforced by the political context) so dis-incentivizing that it contributed to Germany's withdrawal from Sophia, the EU naval mission targeting human trafficking in the Mediterranean [10]? Could and (how) should a visualisation illustrate a progressive emptying of the knowable in the space, caused by the reduction of official and NGO sea-rescue vessels active in the area?

Second, many of the existing, but not accessible data have strong effects on the rescue events modelled. For example, the Libyan coastguard now has indirect access to EUROSUR data [25]; thus, their rescue actions, including those in cooperation or competition with European actors, may be planned based on data that are not modelled in the rescue patterns system, and which therefore can co-determine the "coordination" and "effectiveness" of a rescue. Can and should these data (or at least the fact of their existence and possible influence) be modelled?

Third, further questions concern which aspects are important to judge the legal and ethical dimensions of a rescue operation. An example is provided by [23] resp. [6]: in a case in which a commercial towboat under Italian flag rescued migrants and then handed them over to the Libyan coastguard, a key legal question revolved not around the spatiotemporal data of the rescue operation, but around whether it was instructed by the Italian or the Libyan authorities [37]. Can these aspects be modelled as part of holistic trajectories, and how could this be done if the datum itself is still being contested?

4 Towards a Comparative Analysis

The preceding sections have shown that vehicle trajectory data are often rich sources of personal data, of individuals as well as of groups. However, even if very similar in technical aspects, such data can present very different challenges in different contexts. In both examples analysed in the present paper, concerns of different stakeholders need to be weighed. Even if we only regard stakeholder groups' interests with regard to invisibility (confidentiality of the data) or visibility, further differentiation becomes apparent. For reasons of space, I cannot present a worked-out comparative analysis of the two case studies, or provide a weighing of the different interests in a GDPR sense. Instead, I will sketch some further subdivisions that arise within stakeholder groups, and argue that a weighing of these interests is a more far-reaching political decision.

As regards the stakeholder group "drivers", it appears that those in the taxi case study probably have an interest in invisibility, whereas those in the vessel case study may seek invisibility, be indifferent, have an interest in visibility, or regard being located as a security/safety risk more than as a privacy violation. Their views may also depend on whether they are in the subgroup of "NGO vessel driver" or "other vessel driver".

For the "passengers", invisibility appears a strong interest in the first case study, whereas visibility may be strongly preferred as a prerequisite for surviving by the vulnerable people in the second case study, and different protection needs of individuals and groups become apparent.

"The public" also consists of subgroups with different interests. In the taxi case study, these include citizens interested in the visibility of public city data (as the motivation of the FOI request), scientists interested in publicly available datasets, celebrity spotters interested in disclosures, and privacy activists and data scientists interested in highlighting and preventing data-privacy attacks. In the AIS data case study, different subgroups of the public are even interested in creating different overall narratives, including (a) "there is an invasion of migrants", (b) "the migration crisis is over", and (c) "people keep dying". It may be argued that these narratives [34] induce preferences for visibility (a, c) and invisibility (b), for different reasons, with different finalities, and therefore with foci on different data.

5 Conclusion

In sum, a consideration of the modelling and reporting of vehicle data and patterns, even if restricted to what data are to be included and how, what information is to be kept confidential or disclosed, can reach far beyond the traditional questions discussed under data protection and data privacy, and data science can assume the importance and responsibility normally associated with PETs. This use of modelling and AI requires a critical examination of the sociopolitical background of the mobility that these vehicles afford, support, or impede, and of the goals of the data-science project undertaken. And although "data can help citizens demand accountability", "ultimately, the inferences that can be drawn from the data are only as valuable as the actions they induce. There is a need for political momentum to address the situation in the Mediterranean, and this problem will not be solved with data alone." [18, p. 42].

Acknowledgements. I thank the members of the EU H2020 Marie-Skłodowska-Curie programme project MASTER for the inspiration to deal with the privacy aspects of AIS data and migrants, and for many valuable discussions of previous versions of this article.

References

1. Agre, P.E., Rotenberg, M.: Technology and Privacy: The New Landscape. MIT Press, Cambridge (2001)
2. Al Jazeera News: Danish cargo ship carrying refugees allowed to dock in Italy, 26 June 2018. https://www.aljazeera.com/news/2018/06/danish-cargo-ship-carrying-refugees-allowed-dock-italy-180626081632471.html
3. Atockar: Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset (2014). https://research.neustar.biz/2014/09/15/riding-with-the-stars-passenger-privacy-in-the-nyc-taxicab-dataset/
4. Berendt, B.: More than modelling and hiding: towards a comprehensive view of web mining and privacy. Data Min. Knowl. Disc. **24**(3), 697–737 (2012)

5. Borghese, L., Vandoorne, S., Vonberg, J.: Migrant rescue ship Lifeline to dock in Malta after being stranded for five days in the Mediterranean. CNN News, 26 June 2018. https://edition.cnn.com/2018/06/26/europe/migrant-ships-maersk-lifeline-intl/index.html
6. Cancellato, F.: Poche palle, i migranti della Asso 28 li abbiamo respinti noi: e la Libia è solo la foglia di fico della nostra ipocrisia. Linkiesta, 1 August 2018. https://www.linkiesta.it/it/article/2018/08/01/poche-palle-i-migranti-della-asso-28-li-abbiamo-respinti-noi-e-la-libi/39019/
7. Conway, D.: The Data Science Venn Diagram (n.d.). http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram
8. De Hert, P., Gutwirth, S.: Privacy, data protection and law enforcement. Opacity of the individual and transparency and power. In: Claes, E., Duff, A., Gutwirth, S. (eds.), Privacy and the Criminal Law, pp. 61–104. Antwerp/Oxford: Intersentia (2006)
9. Deibler, D.: EUROSUR - A Sci-fi border zone patrolled by drones? In: Camenisch, J., Fischer-Hübner, S., Hansen, M. (eds.) Privacy and Identity Management for the Future Internet in the Age of Globalisation, pp. 87–109. Springer, Berlin etc. (2015). https://doi.org/10.1007/978-3-319-18621-4
10. Welle, D.: Germany pulls out of Mediterranean migrant mission Sophia, 23 January 2019. https://www.dw.com/en/germany-pulls-out-of-mediterranean-migrant-mission-sophia/a-47189097
11. Domingo-Ferrer, J.: A three-dimensional conceptual framework for database privacy. In: Jonker, W., Petković, M. (eds.) SDM 2007. LNCS, vol. 4721, pp. 193–202. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75248-6_14
12. Douriez, M., Doraiswamy, H., Freire, J., Silva, C.T.: Anonymizing NYC taxi data: does it matter? In: Proceedings of 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, 2016, pp. 140–148 (2016)
13. Elliot, M., Mackey, E., O'Hary, K., Tudor, C.: The Anonymisation Decision-Making Framework. UKAN, Manchester (2016). http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf
14. Friedewald, M., Finn, R., Wright, D.: Seven types of privacy. In: Gutwirth, S., Leens, R., De Hert, P., Poullet, Y. (eds.) European Data Protection: Coming of Age, pp. 3–32. Springer, Heidelberg (2013). https://doi.org/10.1007/978-94-007-5170-5_1
15. Gürses, S.F., Berendt, B.: The social web and privacy. In: Ferrari, E., Bonchi, F. (eds.), Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques. Data Mining and Knowledge Discovery Series. Chapman & Hall/CRC Press, Boca Raton (2010). https://www.esat.kuleuven.be/cosic/publications/article-1304.pdf
16. Hansen, M., Jensen, M., Rost, M.: Protection goals for privacy engineering. In: Proceedings of 2015 IEEE CS Security and Privacy Workshops (2015). https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7163220
17. Hildebrandt, M.: Privacy and identity. In: Claes, E., Duff, A., Gutwirth, S. (eds.) Privacy and the Criminal Law, pp. 43–58. Intersentia, Antwerp (2006)
18. Hoffmann, K., Boy, J., Leon-Dufour, J., Breen, D. Earney, C., Luengo-Oroz, M. Using big data to study rescue patterns in the Mediterranean. In: Fatal Journeys. vol. 3, Part 1: Improving Data on Missing Migrants, pp. 24–46. International Organization for Migration, Geneva (2017). https://publications.iom.int/system/files/pdf/fatal_journeys_volume_3_part_1.pdf

19. Knapton, J.: General Data Protection Regulation: academic research (n.d.). https://www.information-compliance.admin.cam.ac.uk/files/gdpr_and_academic_research_v1.pdf
20. La Repubblica: Diciotti, dopo dieci giorni i migranti sbarcano dalla nave, 26 August 2018. https://www.repubblica.it/cronaca/2018/08/26/news/migranti_diciotti_sbarco-204935293
21. Li, J.P.K., Bhulai, S., van Essen, T.: Optimization of the revenue of the New York City taxi service using Markov decision processes. In: Proceedings of DATA ANALYTICS 2017: The Sixth International Conference on Data Analytics (2017). https://www.thinkmind.org/download.php?articleid=data_analytics_2017_4_10_68005
22. Maldoff, G. How GDPR changes the rules for research. IAPP News (2016). https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/
23. Medina, J.: U.N. says migrants' return to Libya by Italian boat could be illegal. Reuters, 31 July 2018. https://www.reuters.com/article/us-europe-migrants-libya/migrants-return-to-libya-by-italian-boat-could-breach-international-law-u-n-idUSKBN1KL1K4
24. Mendes, K., Ringrose, J., Keller, J.: #MeToo and the promise and pitfalls of challenging rape culture through digital feminist activism. Eur. J. Women's Stud. **25**(2), 236–246 (2018). https://doi.org/10.1177/1350506818765318
25. Monroy, M.: Durch die Hintertür: Anschluss Libyens an europäische Überwachungssysteme. CILIP Blog, 19 January 2018. https://www.cilip.de/2018/01/19/durch-die-hintertuer-anschluss-libyens-an-europaeische-ueberwachungssysteme/
26. Pandurangan, V.: On Taxis and Rainbows: Lessons from NYC's improperly anonymized taxi logs (2014). https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1
27. Phillips, D.: Privacy policy and PETs: the influence of policy regimes on the development and social implications of privacy enhancing technologies. New Media Soc. **6**(6), 691–706 (2004)
28. Patroumpas, K., Alevizos, E., Artikis, A., Vodas, M., Pelekis, N., Theodoridis, Y.: Online event recognition from moving vessel trajectories. GeoInformatica **21**(2), 389–427 (2017). https://doi.org/10.1007/s10707-016-0266-x
29. Pratesi, F., Monreale, A., Trasarti, R., Giannotti, R., Pedreschi, D., Yanagihara, T.: PRUDEnce: a system for assessing privacy risk vs utility in data sharing ecosystems. Trans. Data Priv. **11**(2), 139–167 (2018)
30. Solove, D.J.: Understanding Privacy. Harvard University Press, Cambridge (2008)
31. Taylor, L.: Safety in numbers? Group privacy and big data analytics in the developing world. In: Taylor, L., Floridi, L., van der Sloot, B. (eds.) Group Privacy. PSS, vol. 126, pp. 13–36. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-46608-8_2
32. Taylor, L., Floridi, L., van der Sloot, B. (eds.): Group Privacy. PSS, vol. 126. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-46608-8
33. Tazzioli, M.: Eurosur, humanitarian visibility, and (nearly) real-time mapping in the Mediterranean. ACME **15**(3), 561–579 (2016). https://acme-journal.org/index.php/acme/article/view/1223/1201
34. The Guardian: EU declares migration crisis over as it hits out at 'fake news', 6 March 2019. https://www.theguardian.com/world/2019/mar/06/eu-declares-migration-crisis-over-hits-out-fake-news-european-commission

35. uluman: Identifying Muslim cabbies from trip data and prayer times (2015). https://www.reddit.com/r/dataisbeautiful/comments/2t201h/identifying_muslim_cabbies_from_trip_data_and/
36. Varlamis, I., Tserpes, K., Sardianos, C.: Detecting search and rescue missions from AIS data. In: 2018 IEEE 34th International Conference on Data Engineering Workshops (ICDEW), Paris, 2018, pp. 60–65 (2018). http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8402020&isnumber=8402003
37. Ziniti, A.: Migranti, inchiesta sul comportamento del rimorchiatore italiano Asso 28. La Repubblica, 8 August 2018. http://www.repubblica.it/cronaca/2018/08/08/news/migranti_presentato_un_esposto_sul_comportamento_del_rimorchiatore_italiano_asso_28-203654671/
38. Ziniti, A.: Migranti Sea Watch anche in Italia, accordo europeo raggiunto con Malta. La Repubblica, 9 January 2019. https://www.repubblica.it/cronaca/2019/01/09/news/migranti_accordo-216163365/