Chapter 25
Digital Earth Ethics

Yola Georgiadou, Ourania Kounadi and Rolf A. de By

Abstract Digital Earth scholars have recently argued for a code of ethics to protect individuals' location privacy and human dignity. In this chapter, we contribute to the debate in two ways. First, we focus on (geo)privacy because information about an individual's location is substantially different from other personal information. The compound word (geo)privacy suggests that location can be inferred from people's interests, activities, and sociodemographics, not only from traditional geographic coordinates. (Geo)privacy is a claim of individuals to determine for themselves when, how, and to what extent location information about them is communicated to others. Second, we take an interdisciplinary perspective. We draw from (geo)computing to describe the transformation of volunteered, observed, and inferred information and suggest privacy-preserving measures. We also draw from organization studies to dissect privacy into ideal types of social relationships and privacy-preserving strategies. We take the point of view of Alice, an individual 'data subject' encountered in data protection legislation, and suggest ways to account for privacy as a sociocultural phenomenon in the future. Although most of the discussion refers to the EU and the US, we provide a brief overview of data protection legislation on the African continent and in China as well as various global and regional ethics guidelines that are of very recent vintage.

Keywords Ethics · Geoprivacy · Spatial data · Inference attacks · Privacy-preserving measures

25.1 Introduction

The previous chapters of the Manual of Digital Earth describe remarkable progress to date. Key technologies envisioned by Vice President Gore in 1998 are now in place for the first-generation and next-generation Digital Earth (DE). Similar progress in DE ethics is not yet evident despite the early ethical stirrings in the geographic community. As early as 1990, at a roundtable on *Ethical Problems in Cartography*,

Y. Georgiadou (✉) · O. Kounadi · R. A. de By
Geo-Information Science and Earth Observation, University Twente, Enschede, The Netherlands
e-mail: p.y.georgiadou@utwente.nl

H. Guo et al. (eds.), *Manual of Digital Earth*,
https://doi.org/10.1007/978-981-32-9915-3_25

Brian Harley wondered whether cartography was out of step with other disciplines. He suggested that the real ethical priority is for a map to be a socially responsible representation of the world: *"Can there be an ethically informed cartography and what should be its agenda? [S]hould we be concerned with transcendental values that go to the heart of social justice in the world at large?"* (Harley 1991, p. 9). In this chapter, we update Harley's vocabulary for the current era of datafication of everyday life (Cukier and Mayer-Schoenberger 2013) and explore the *Ethics of Where* instead of the ethics of cartography. This leads us to recent debates on data justice—fairness in the way people and their resources are made visible, represented and treated as a result of their digital data production (Taylor 2017).

In 2012, DE scholars observed that any effort to develop a next-generation Digital Earth will require a principle of privacy protection that minimally guarantees control over any individual's locational privacy and the ability to turn it on or off at will. They noted, *"there is also room for a Digital Earth code of ethics that could set standards for behavior in a complex, collaborative enterprise […] necessary to tackle the growing issues of privacy and ethics that are associated with access to fine-resolution geographic information"* (Goodchild et al. 2012, pp. 11092–3). In 2014, some of the authors of the previous paper reiterated the call for privacy and reaffirmed the need for a code of DE Ethics. They argued that *"technological advancements have to be accompanied by the development of a DE code of ethics that ensures privacy, security, and confidentiality in a world where everybody can be connected to everybody else and everything all the time. Without solving this critical dilemma and allowing people to decide whether or not they want to be connected and how much of their thoughts and emotions they want to share, the dream of a wonderful virtual future may well turn into DE nightmare"* (Ehlers et al. 2014, p. 13). They boldly suggested that Digital Earth should follow the Kantian ethics of personal autonomy and human dignity in composing its code.

An obvious source of inspiration and lessons for such a code are the practices of the Association for Computing Machinery (ACM), which represents and regulates the behavior of a global computing community of approximately 100,000 members. In 2018, the ACM updated its Code of Ethics and Professional Conduct to address the significant advances in computing technology and the growing pervasiveness of computing in all aspects of society (ACM Code Task Force 2018). The responsibility to respect privacy, one of the seven general ethical principles in the ACM Code of Ethics, applies to computing professionals in a profound way. The ACM urges computing scholars and professionals to become conversant in the various definitions and forms of privacy and understand the rights and responsibilities associated with the collection and use of personal information. The ACM appeals to all computing professionals, including current and aspiring practitioners, instructors, students, influencers, and anyone who uses computing technology in an impactful way. Given that big computing companies have a significant impact on society, we should explore how their views on privacy have diverged over time from the current ACM ideal and how they contest privacy as a concept. Some consider privacy irrelevant. As early as 1999, Scott McNealy, the founder and CEO of Sun Microsystems, declared *"you have zero privacy … get over it,"* a statement some in the privacy industry took as

tantamount to a declaration of war (Sprenger 1999). Others consider it an evolving social norm. In 2010, Mark Zuckerberg claimed that *"people have really gotten comfortable not only sharing more information and different kinds, but more openly and with more people," he said. "The [privacy] social norm is just something that has evolved over time"* (Johnson 2010). Others such as Apple CEO Tim Cook note that *"the poor privacy practices of some tech companies, the ills of social media and the erosion of trust in [Cook's] own industry threaten to undermine "technology's awesome potential" to address challenges such as disease and climate change"* (Romm 2018).

Privacy is a contested concept for good reasons. First, the etymology—the history of linguistic forms—reveals how privacy changed meaning from derogatory to laudatory. The ancient Greek word ἰδιώτης (pronounced idiōtēs) originally meant a private man, an ignoramus, as opposed to δημόσιος (pronounced dēmosios; meaning 'of the people'), a person of public distinction (Liddell and Scott 1940). Currently, the stem of idiōtēs forms the word idiot and dēmos is one of the two stems of democracy. The word private in Latin meant 'deprived' of public office–privacy designated a (negative) state of deprivation. For instance, a private in the army is a person with no rank or distinction and very little privacy (Glanville 2018). Second, privacy is contested because it can be portrayed in various competing ways—as a positive or negative right (Floridi 2014); as an instrument for Kantian ethics—human dignity and personal autonomy; and as an instrument for Aristotelean virtue ethics—personal development and human flourishing (van der Sloot 2014). The watershed US Supreme Court case, *Kyllo v. United States*, reported in Mulligan et al. (2016) and reproduced in the box below, is an example of how a seemingly simple case of home privacy violation was contested by the defendant, the federal government and the Supreme Court in 2001. The five to four decision of the Supreme Court eventually upheld the Fourth Amendment—the right of an individual to retreat into his own home and be free from unreasonable governmental intrusion, in this case, free from the intrusion of a thermal imaging device deployed by a federal agent to scan the outside of Kyllo's home (US Supreme Court 2001).

Kyllo v. United States involved an investigation of a marijuana cultivation and distribution operation in which a federal agent used a thermal imaging device to scan the outside of Kyllo's home. The resulting thermal image was used to obtain a warrant to search the house. Kyllo moved to suppress the evidence recovered from the search of his home, arguing that the use of the thermal imaging device to scan it was an invasion of his reasonable expectation of privacy. In a five to four decision, the Supreme Court held that 'obtaining by sense-enhancing technology any information regarding the interior of the home that could not otherwise have been obtained without physical "intrusion into a constitutionally protected area", constitutes a search—at least where (as here) the technology in question is not in general public use'.

The Kyllo case was contested at every level. The parties disagreed over the object of privacy under contention. The government argued that Kyllo had no expectation of privacy in 'the heat emitted from the home', while Kyllo argued that what privacy protected was the 'private activities' occurring within the home. The five justices who made up the majority determined that the case was about the 'use of technology to pry into our homes', the related matter of the sanctity of 'private lives', and the need to draw a not only 'firm but also bright' line to protect the sanctity of the home and the activities occurring within it. During oral argument, the justices drew attention to evidence provided to the appellate court revealing that a thermal image reading could 'show[ed] individuals moving . . . inside the building' to emphasize that what was at risk was not data, but 'what's going on in the house'.

The dissenting justices drew a distinction between 'through-the-wall surveillance that gives the observer or listener direct access to information' and 'inferences from information in the public domain' explaining that inferences drawn from 'gathered data exposed on the outside of petitioner's home' did not intrude on privacy. Justice Stevens's writing for the dissent explained, 'it would be quite absurd to characterize [the police's] thought processes'—the inference they drew from the data that seeped through the walls—as 'searches'. The majority justified its decision to prohibit the use of thermal imagers absent a warrant in order to protect the privacy of in-home activities on the basis that 'at the very core' of the Fourth Amendment 'stands the right of a man to retreat into his own home and there be free from unreasonable governmental intrusion'. The ruling was justified by the need to limit the Government's access to individuals' private lives.

Reprinted with permission from Mulligan et al. (2016, pp. 6–7). Copyright 2016 The Royal Society Publishing.

Currently, the dissenting judges' claim that inferences drawn from thermal imagery of Kyllo's home were not an intrusion of his privacy but only the *'police's thought processes'* and the government's assertion that *'the heat emitted from the home'* is not private seem normal. In the Netherlands, heat detection from police helicopters is not considered systematic government observation (Hennepadvocaat 2019) and thus constitutes legal proof. Our location and movement, tweets, emails, photos and videos, purchases, our every click, misspelled word, and page view—are routinely observed by government and big tech via mobile phones, surveillance cameras, drones, satellites, street views, and corporate and government databases to draw inferences that can control, predict and monetize our behavior. Siegel (2013) notes that an individual's data can be purchased for approximately half a cent, but the average user's value to the Internet advertising ecosystem is estimated at $1,200 per year. Wall Street values tech giants, not because of the services they provide but for the data they collect from individuals and its worth to advertisers (Halpern 2013). Ironically, these data may be emitted by millions of automated accounts, each sold

by obscure companies many times over, or celebrities, businesses or anyone desiring to exert influence online, according to a New York Times investigation (Confessore et al. 2018).

These facts have not escaped the public's attention. The Snowden revelations (Greenwald and MacAskill 2013) and the Cambridge Analytica scandal (The Guardian 2018) were probably the biggest contributors to citizens' changing perceptions of privacy, though not in the way Zuckerberg predicted in 2010. People care now more about privacy, and liberal governments responded accordingly. A 2018 survey by The Atlantic found that in the USA, *"overall, 78.8 percent of people said they were "very" or "somewhat" concerned about the privacy of their information on social media, and 82.2 percent said they self-censor on social media"* (Beck 2018). In 2018, legislation was passed in Vermont to regulate data brokers and California gave its residents the right to be informed about the kinds of personal information companies have collected about them, as well as the right to request that their personal information be deleted. Colorado-based companies will be required to, among other things, dispose of certain kinds of personally identifying information. The different types of information prone to compromise individual privacy are explained in detail in Sect. 25.2. Overall, two thirds of Americans are now eager to see stricter privacy laws (Halpern 2018). On May 25, 2018 the General Data Protection Regulation (GDPR) came in force to protect individuals in the 28 member countries of the European Union, even if their data is processed elsewhere. The GDPR applies to publishers, banks, universities, most Fortune 500 companies, ad-tech companies and the Silicon Valley tech giants. With the GDPR, *"companies must be clear and concise about their collection and use of personal data like full name, home address, location data, IP address, or the identifier that tracks web and app use on smartphones. Companies have to spell out why the data is being collected and whether it will be used to create profiles of people's actions and habits. Moreover, consumers will gain the right to access data companies store about them, the right to correct inaccurate information, and the right to limit the use of decisions made by algorithms"* (Tiku 2018).

Ethical issues arising in studies of our planet, as a system involving natural, man-made and hybrid processes, are enmeshed with scientific or industrial practices. Professional codes of ethics safeguard the public good by requiring honesty, trust and fairness, and the avoidance of harm. Respect for privacy and other people's work addresses concerns of intrusion and intellectual property. Studies involving geospatial information may be riddled with ethical ambiguity because professional responsibility requires acknowledging that the proposed methods may not travel well to other geographies. In short, location is burdened with contextual specifics. If such specifics are not parameterized, the earth sciences are vulnerable to the reproducibility crisis (Baker 2016). Ethics in Digital Earth methods are thus fundamentally important to study, and we expect open science approaches (Vicente-Saez and Martinez-Fuentes 2018) to mature in coming years and allow improvement of their methodical robustness.

In this chapter, we contribute to the *Ethics of Where* in two ways. First, we focus on information privacy, and location privacy, or (geo)privacy. This is necessary because

information about an individual's location is substantially different from other kinds of personal information. The reasons for this include the ease of capturing an individual's location, the improvement of a service when the user shares their location with a service provider, and the potential to infer sensitive information about social, economic or political behavior from location history (Keßler and McKenzie 2018). Data inferred from an individual's location are socially constructed. If a society considers a given mode of personal behavior—e.g., political opinion, sexual orientation, religious or philosophical beliefs, trade union membership—to be socially legitimate, then these data are deemed personal and worthy of protection. We define privacy as a positive right concerning *"the claim of individuals to determine for themselves when, how, and to what extent location information about them is communicated to others"* because control of location information is the central issue in location privacy (Duckham and Kulik 2006, p. 36). Second, we complement other studies that describe the current state of the art and formulate challenges (Keßler and McKenzie 2018; Zook et al. 2017) or describe different scenarios concerning the development of geoprivacy (Wegener and Masser 1996) and revisit them (Masser and Wegener 2016) by taking an interdisciplinary perspective. We draw from the field of (geo)computing to describe the transformation of volunteered, observed, and inferred information (Sect. 25.2) and suggest privacy-preserving measures (Sect. 25.4). We draw from organization studies to dissect privacy into some ideal types of social relationships and strategies (Sect. 25.3), and draw from cultural theory to suggest future research (Sect. 25.5). The final section provides a brief overview of data protection legislation on the African continent and in China as well as various global and regional ethics guidelines.

We use the compound word (geo)privacy to suggest that, although control of location information is the central issue, location can be inferred from people's interests, activities, and sociodemographics, not only from 'traditional' location information, e.g., geographic coordinates (Keßler and McKenzie 2018). Further, we emphasize the distinction between privacy as a negative right (freedom from interference) and privacy as a positive right (freedom to control). This is because old, predigital technologies—such as the instantaneous photographs and newspaper tabloids in Brandeis and Warren's time—restricted individuals to claiming privacy as a negative right only, as freedom from interference or '*the right to be left alone*' (Warren and Brandeis 1890). New digital technologies can reduce or significantly enhance privacy as a positive right, i.e., the freedom to control (Floridi 2014), often in combination with social and/or organizational and/or legal measures/strategies (Mulligan et al. 2016).

25.2 Transforming Volunteered and Observed Data to Inferred Data

We distinguish three types of personal data: volunteered, observed and inferred data. These new types replace the old, 'personal, nonpersonal' data distinction, which has outlived its usefulness in the era of datafication. We define the three data types as

suggested by the World Economic Forum (2011, p. 7): "*Volunteered data are created and explicitly shared by individuals, e.g. social network profiles. Observed data are captured by recording the actions of individuals, e.g. location data when using cell phones. Inferred data are data about individuals based on analysis of volunteered or observed information, e.g. credit scores.*" These three types involve both spatial and nonspatial data. We define spatial data as data that includes explicit coordinates interpretable in an open, well-known system. Examples are map coordinates, postal codes and street addresses. We do not think of mobile cell tower numbers as spatial data because special insight into the coding mechanism is required to understand their location.

To explain how volunteered and/or observed spatial data can be transformed into inferred data, we describe spatial data types with private or confidential components and provide examples of possible inference attacks on them. In principle, the subjects of these data types can be humans, organizations, groups of people, animals, nonliving physical objects such as buildings, or other confidential information with location attributes. Here, we focus on individual humans as data subjects. Hence, we drop the term 'confidentiality' and focus on 'privacy' because data classified as confidential (e.g., health records) is also private at an individual level. Similarly, inferences or inference attacks refer to private data that can be derived for each individual included in a spatial dataset.

We define a key identifier as an attribute that can be exploited with minimal effort to identify a subject. According to the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, some common key identifiers are a person's name, telephone number, fax number, street address, electronic mail address, social security number, vehicle license plate, device identifier, biometric identifier, facial image, Internet protocol (IP) address, and web universal resource locator (URL) (U.S. Government Publishing Office 2009). Other potential key identifiers are account names on Internet platforms (e.g., in social media applications) and coordinate pairs of private information (e.g., location of households). In some cases, a key identifier links private information to a single individual only for a subset of the data. For example, in a dataset with locations of households, a small percentage corresponds to single-family houses (or detached houses) with only one occupant. The key identifier is a direct identifier for this subset. In other cases, a key identifier links private information to a small group of people closely related to the subject. This group may be family members who become emotionally traumatized if their private information is released or may be other house occupants that are incorrectly identified as the subjects. In addition, we define a quasi-identifier as an attribute that pinpoints a subject uniquely or almost uniquely, when combined with at least one other quasi-identifier attribute. A unique identifier (UID) is an attribute that allows for uniquely identifying single subjects. In some cases, a UID can be a key identifier (e.g., social security number, which identifies a subject), in others, its value may not be subject-specific, for instance, if it identifies a drug brand or a pharmaceutical factory process number, which cannot be used to disclose private information. Finally, a private attribute is any attribute that is not a key identifier, a quasi-identifier, or a UID, and contains

other information about the subject from which inferences regarding privacy can be drawn.

The above data typology focuses on the usefulness of spatial or non-spatial data in inferences that affect privacy. Below, we discuss a second data typology that characterizes the roles of spatial and temporal attributes.

The simplest spatial data type is *'discrete location data'* (abbreviated *Dd*); it is a collection of one or more key spatial identifiers. The disclosure of this data type implies disclosure of subjects linked to the private information or to a small circle of possible subjects for each key identifier. Examples of *Dd* are the locations of domestic violence events and addresses of cancer patients. In both these cases, subjects can be identified as a person living at the disclosed location. As with all the data types discussed here, we assume that the data holder can interpret the data because they are aware of the contextual information that defines the search (e.g., "this is a collection of addresses of cancer patients").

A second data type is *'discrete location data with covariates,'* hereafter referred to as *Dd* + . The "+" symbol extends the notion of *Dd* by including additional attributes. The additional attributes are one or more quasi-identifiers. Quasi-identifiers are demographic, social, or economic attributes. A private attribute may or may not be present. An example of *Dd* + is a crime dataset of locations of offences (key identifier), the age of the victim (quasi-identifier), the ethnicity of the victim (quasi-identifier), and the type of the offence (private attribute). The location of offence is a key identifier, at least for that subset of the data collection where the type of offence occurs predominantly in residential addresses.

An inference attack on *Dd* and *Dd* + data types aims to identify (or re-engineer) the location of some subject(s). The data may not be disclosed but presented as a printed or a digital map. Such media can be geoprocessed to re-engineer the locations with considerable accuracy (Brownstein et al. 2006; Leitner et al. 2007). Multiple anonymized copies of the data can be disclosed, accompanied by specifications of the anonymization technique, for instance, for scientific transparency and reproducibility. This can provide hints to the attacker and, depending on the strength of the technique, locations can be re-engineered with the Gaussian blurring algorithm (Cassa et al. 2008).

A third data type is *'space-time data,'* hereafter referred to as *STd.* Data of this type contains location and timestamps for one or more subjects, which can be distinguished with a UID. Each location represents or approximates where a subject was at a particular time. Typical examples are call data records (CDR) and data used in location-based services (LBS). Unless the identity of the subject is known (e.g., when UIDs are real names), there is no key identifier or quasi-identifier. Nevertheless, the subjects' spatiotemporal whereabouts can be analyzed to draw a plethora of inferences such as their home address, work address, time spent away from work, and places visited during weekends (Alrayes and Abdelmoty 2014).

Gambs et al. (2010) analyzed GPS mobility traces of 90 taxi trails in San Francisco, US. They attempted to infer the home location of the drivers using a heuristic approach of the first and last recorded locations during working days. However, they did not have validation data to assess the accuracy of their approach. De Montjoye

et al. (2013) focused on the uniqueness of spatiotemporal trajectories and analyzed mobility data from mobile phone interactions (calls and messengers) for approximately 1.5 million people. They found that four random locations are enough to uniquely characterize 95% of mobile users for a sample in which the location of a user is specified hourly, with a spatial resolution equal to that determined by the carrier's antennas.

The fourth and last data type is the *'space-time-attribute'* data, hereafter referred to as *STd* + . As with *Dd* + , the "+" symbol denotes an extended version of *STd*, which includes additional attributes that can be quasi-identifiers or private attributes. An example of *STd* + is the georeferenced data of a Twitter user. Twitter data contains spatial and temporal information as well as the short message text posted by the user. Inferences can be made similar to those for *STd*. Additionally, the textual or otherwise semantic information may reveal private matters about the sender such as interests, beliefs, and attitudes. For instance, Preoţiuc-Pietro and Cohn (2013) exploited the primary venue type in Foursquare check-ins (i.e., professional and other, travel and transport, residence, food, nightlife spots, university, outdoors, arts and entertainment, and shop and service) to cluster users by behavioral patterns and estimate their next activity based on the history of past venue visits. In another real-word but small-scale study, LBS network data of university volunteers was analyzed based on location similarity. Inferences were made to predict the users' demographics such as education level and gender (Li et al. 2018).

Participatory sensing data are data collected by volunteers, mainly for research, using mobile sensors such as biometric bracelets, smartwatches or smartphones. They include data from mobile devices such as sensors carried by 'humans as sensor operators,' sensors carried by 'humans as objective sensors,' and devices carried by 'humans as subjective sensors' (Kounadi and Resch 2018). Participatory sensing data are the *STd* + type. For example, participants in a participatory research campaign may use mobile apps that track their space-time information and report their level of stress (i.e., sensitive information) throughout their activity spaces (Zeile et al. 2011). In participatory sensing data, private attributes are observed or volunteered geoinformation whereas private attributes are also inferred geoinformation in LBS network data. Thus, due to the error of the inference process, the disclosure risk of LBS network data may be lower than that of participatory sensing data.

The four types of spatial data are illustrated in Fig. 25.1, where:

- $S_{1...k}$ is the spatial attribute such as the coordinates, postal codes, or street addresses;
- $T_{1...n}$ is the temporal attribute such as hour, date, or month; and
- $A_{1...m}$ are quasi-identifiers and/or private attributes.

All spatial data types include a spatial attribute. Two of the data types contain a temporal attribute (*STd* and *STd* +) and two contain additional attributes (*Dd* + and *ST* +).

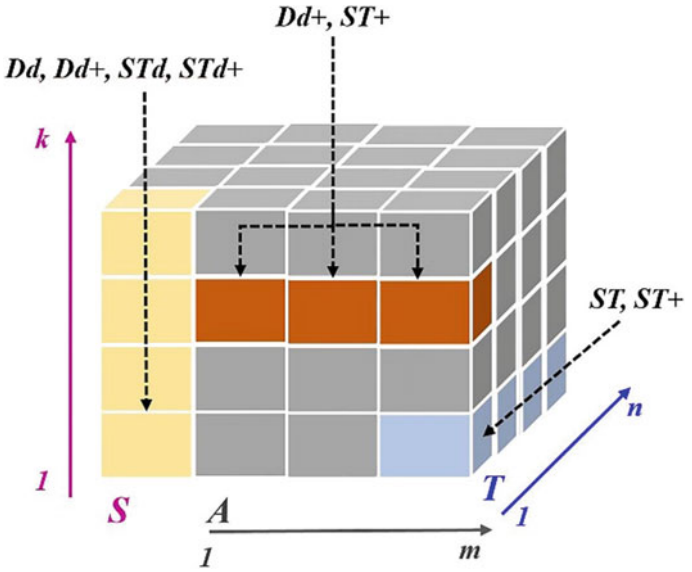

Fig. 25.1 Four spatial data types, Dd, Dd+, ST, ST+, and the types of attributes they contain (S, T, A)

25.3 A Typology for (Geo)Privacy

Privacy is always relational. It does not make sense for a lonely man on a desert island. At its simplest, privacy relates two parties—a human to a human, a human to a group of humans, a human to a private corporation or a human to a government institution. These relations can be arranged in a typology of (geo)privacy (Table 25.1). This grouping is a gross simplification of reality. For instance, LBS involve no less than thirteen human, machine and software parties—the mobile device, the hardware manufacturer, the operating system, the operating system manufacturer, the mobile application, the mobile application developer, the core application, the third-party software, the third-party software developer, the LBS, the LBS provider, the network operator and government (Herrmann 2016). Further, government institutions and private corporations often cooperate. The National Security Agency obtained direct access to the systems of Google, Facebook, Apple and other US Internet giants as part of the Prism program, which allows for officials to collect material including search history, the content of emails, file transfers and live chats (Greenwald and MacAskill 2013). Nevertheless, the four ideal types of relations help create a rough grid into which finer resolution grids may be inserted in future iterations.

At the heart of the typology is Alice. We may imagine her as a member of ACM who must comply with the ACM Code of Ethics or as a member of a (geo)computing department at a European university, which must comply with the GDPR. Alice values (geo)privacy as a positive right, a right that obliges action by individuals, groups,

Table 25.1 A typology of (geo)privacy relations

		Goal incongruity	
		Low(er)	*High(er)*
(Alice's) Ability to control human behavior, machine behavior, outputs	*Low(er)*	Cell (4) Alice—Government institution **Privacy strategy**: Compliance; lodge complaint to DPA in case of violation of the GDPR; anti-surveillance resistance	Cell (3) Alice—Private corporation **Privacy strategy**: Control behavior of corporation (via GDPR); lodge complaint to DPA in case of violation of the GDPR
	High(er)	Cell (1) Alice—Bob **Privacy strategy**: Right and duty of partial display	Cell (2) Alice–(Bob–Carol–Dan-etc.) **Privacy strategy**: Geoprivacy by design

or institutions to determine for themselves when, how, and to what extent information about them is communicated to others (Westin 1967). To transfer from Westin's times to the information age, Alice values privacy as the positive right of individuals, groups, or institutions *"to control the life cycle (especially the generation, access, recording, and usage) of their information and determine when, how, and to what extent their information is processed by others"* (Floridi 2014, p. 114). The relationality of privacy highlights the possibility that the privacy goals of two binary parties can be incongruous and results in the horizontal dimension of the typology in Table 25.1. Incongruity can be low or high. The vertical dimension refers to Alice's ability to control the transformation process of her volunteered, observed or inferred information or that of her research subjects. Her ability is high when she can control the entire transformation process—the behavior of humans (incl. herself) and machines and outputs. It is low when she can control some or none of these (Ouchi 1979; Ciborra 1985). Alice's ability (low or high) to control the transformation process results in the vertical dimension in Table 25.1.

In Cell (1), two humans (Alice and Bob) are interacting face-to-face in a private or public space. This is the archetypal human-to-human interaction. Both Alice and Bob are conscious of being observed by each other and other humans and have similar privacy goals—to uphold a tacit social code, the *'right and duty of partial display.'* The sociologist Erving Goffman (1957) described how all humans reveal personal information selectively to uphold this code while constructing their public personae. Hence, the low incongruity between Alice's and Bob's goals to protect their privacy—both strive to uphold this tacit social code, to protect (or curate) their public personae, and modulate it gradually over time, as the relation expands or shrinks. As Fried (1968) explains, Alice may not mind that Bob knows a general fact about her but may feel her privacy is invaded if he knows the details. For instance, Bob may comfortably know that Alice is sick, but it would violate her privacy if he

knew the nature of the illness. If Bob is a good friend, he may know what particular illness Alice is suffering from but it would violate her privacy if he were actually to witness her suffering. Both control their behavior and the knowledge they share (outputs) about each other and may choose to modulate them over time. Goffman's theory applies in settings where participants can see one another face-to-face and has implications for technology-mediated interactions, e.g., in email security (Agre and Rotenberg 1997). When emailing each other, Alice and Bob may choose from a continuum of strategies to safeguard their privacy depending on context. They may refrain from emailing, they may email each other but self-censor, they may delegate privacy protection to mail encryption and firewalls, or they can work socially and organizationally to ensure that members of their community understand and police norms about privacy (Bowker et al. 2010).

Cell (2) describes the interaction of a human, e.g., Alice, the research leader of a participatory sensing campaign, with a group of campaign participants (Bob, Carol, Dan, Eric, etc.).

The goal incongruity between Alice and the group may be high if the group members are not aware of possible breaches to their privacy and their implications. As campaign leader, Alice has a high ability to control outputs and the behaviors of group members and machines and takes a series of privacy-preserving measures for the entire group before, during and after the campaign, a strategy Kounadi and Resch (2018) call *'geoprivacy by design.'* Kounadi and Resch (2018) propose detailed privacy-preserving measures in four categories, namely, 6 measures prior to the start of a research survey, 4 measures for ensuring secure and safe settings, 9 measures for processing and analysis of collected data, and 24 measures for safe disclosure of datasets and research deliverables. Table 25.2 provides illustrative examples in each category. Interestingly, measures to control human behavior include two subtypes:

Table 25.2 Examples of measures that control the transformation process

	Measures controlling human/machine behavior and outputs
Prior to start of campaign	human behavior (participation agreement, informed consent, institutional approval); outputs (defined criteria of access to restricted data)
Security and safe settings	human behavior (assigned privacy manager, trained data collectors); machine behavior (ensuring secure sensing devices, ensuring a secure IT system)
Processing and analysis	outputs (deletion of data from sensing devices, removal of identifiers from data set)
Safe disclosure	outputs (reduction of spatial and temporal precision, consideration of alternatives to point maps) human behavior (providing contact information, using disclaimers, avoiding the release of multiple versions of anonymized data, avoiding the disclosure of anonymization metadata, planning a mandatory licensing agreement, authenticating data requestors)

outreach measures, e.g., participation agreements, and measures of self-restraint, e.g., the use of disclaimers, avoiding release.

Cell (3) describes the interaction of Alice with a private corporation, as a user of a location-based service, of which Google Maps is the most popular and commonly used. Alice *volunteers* her location to the LBS to get directions to a desired destination (Herrmann 2016). In this case, the goal incongruity between Google and Alice is high, as evident from comparing Alice's commitment to (geo)privacy with that of Google's former executive chair Eric Schmidt. *"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place"* (Newman 2009). Alice's ability to control how her location information is used by LBS to infer other information about her is low. As an EU citizen, she can rely on the GDPR to (partly) control the behavior of the LBS provider. Another strategy is lodging a complaint to her national Data Protection Authority (DPA). DPAs are independent public authorities in each EU state that supervise application of the GDPR and handle complaints lodged concerning violations of GDPR. If the private corporation where Alice works systematically monitors its employees, including their workstations and Internet activity, a Data Protection Impact Assessment (DPIA) may be required.

Cell (4) describes the interaction of Alice with government institutions. Alice trusts that her government will respect her right to information privacy (thus the goal incongruity is low) but may be in the dark regarding the transformation process unless a whistleblower leaks a secret surveillance program (e.g., Greenwald and MacAskill 2013) or the abuse of private data (The Guardian 2018). Further, if the public organization where Alice works engages in processing that is likely to result in a high risk to the rights and freedoms of individuals, Alice may lodge a complaint to the DPA and request a DPIA. Such processing may include the systematic and extensive evaluation of personal aspects of an individual, including profiling, the processing of sensitive data on a large scale, or the systematic monitoring of public areas on a large scale.

Another strategy for Alice is collective, e.g., participating in popular resistance to unpopular government action. When the government of the Federal Republic of Germany announced a national census on 27th April 1983, German citizens protested so strongly that a dismayed German government had to comply with the Federal Constitutional Court's order to stop the process and take into account several restrictions imposed by the Court in future censuses. Asking the public for personal information in 1983, the fiftieth anniversary of the National Socialists' ascent to power, was apparently bad timing, to say the least (Der Spiegel 1983). When the census was finally conducted in 1987, thousands of citizens boycotted (overt resistance) or sabotaged (covert resistance) what they perceived as Orwellian state surveillance (Der Spiegel 1987).

Notably, these remarkable events took place in an era where the government was the only legitimate collector of data at such a massive, nationwide scale and at a great cost (approx. one billion German marks). Currently, state and corporate surveillance are deeply entangled. In response, technologically savvy digital rights activists have been influential in several venues, including the Internet Engineering Task Force (IETF) and the Internet Corporation for Assigned Names and Numbers

(ICANN), through the Noncommercial User Constituency (NCUC) caucus. However, their efforts have largely remained within a community of technical experts ('tech justice') with little integration with 'social justice' activists (Dencik et al. 2016).

25.4 Measures to Preserve Geoprivacy

In Sect. 25.2, we characterized various data types that deserve specific scrutiny when privacy is concerned. This characterization was motivated by the perspective of a variety of attackers' strategies (either theoretically possible or practically realized) to identify private information on a subject. Below, we describe geoprivacy-preserving measures to counter such attacks. Section 25.3 highlighted the relationality of privacy and described the four fundamental relations that are critical to understanding privacy as a societal phenomenon. In real life, the social graph is not bipartite and humans cannot be bluntly labeled as either 'attacked' or 'attacker'. Relations are often transitive and privacy-relevant information may travel along longer paths, which implies that intermediate agents may have dual, possibly frictional, roles. One rather regulated, yet much-discussed case, is that of patients whose hospital visits are covered by health insurance companies. Geoprivacy may be related to the living or working conditions of the patient. The patient's typical direct relation with the insurance company does not make this case less trivial. A second example that played out recently in the Netherlands was that of a citizen with a tax dispute, and the national museum foundation that had issued an annual pass to that person (van Lieshout 2018). The tax office accessed the person's museum visit details to prove that he actually lived in the Netherlands, and not abroad, as he claimed.

To identify core geoprivacy measures, we must define the landscape of variables and the values they take, and explore their interrelationships. Six *fundamental variables* are discussed below, along with their values, and are summarized in Table 25.3. The first variable is the 'attacked', who is any subject in a dataset that may be harmed from potential inferences. The attacked is an individual such as Alice—i.e., aware of privacy risks and subscribing to the ACM code of Ethics—or someone who is unaware of privacy risks and relevant legislation and regulations. The second variable is the *'attacker'*, who could use the data for a malevolent purpose. The attacker may be a government institution, corporation, researcher, or other individual. The third variable is the *'data type'*, any of the four types discussed in Sect. 25.2 (i.e., *Dd*, *Dd+*, *STd*, or *STd +*). The fourth variable is the *'purpose of attack'*, which may assume two values: (a) private attribute(s) of the attacked are identified (attribute(s) is unknown but the attacked is known) and (b) the attacked who has certain private attribute(s) is identified (attacked is unknown but the attribute(s) is known). In attacks of the first category, the attacker knows that the attacked's details are contained in a dataset and the attacker aims to draw inferences on the attacked. In those of the second category, the attacker knows the private information and aims to infer the identity of the attacked.

Table 25.3 Fundamental geoprivacy variables and their associated values

Variable	Values
Attacked	1. Any individual
Attacker	2. Government/Institution 3. Corporation 4. Researcher 5. Any individual
Spatial data types	1. Discrete location data (*Dd*) 2. Discrete location data with covariates (*Dd+*) 3. Space-time data (*STd*) 4. Space-time-attribute data (*STd+*)
Purpose of attack	1. Identify private attribute(s) of the attacked 2. Identify the attacked who has certain private attribute(s)
Attacker's strategy	1. Key-identifier exploitation 2. Combine to uniqueness 3. Re-engineering locations 4. Analyzing locations 5. Homogeneity attack 6. Background attack 7. Composition attack
Privacy-preserving measures	1. Pseudoanonymity 2. K-anonymity 3. Spatial *k*-anonymity 4. *l*-diversity 5. Differential privacy

We have used terminology from Sects. 25.2 and 25.3 to define four of the six fundamental variables. Two more variables in the geoprivacy landscape are discussed next. The fifth is the *'attacker's strategy'* (also referred to as "*inference attacks*") that can take seven forms: (a) key-identifier exploitation, (b) combine to uniqueness, (c) re-engineering locations, (d) analyzing locations, (e) homogeneity attack, (f) background attack, and (g) composition attack.

The simplest type of inference is *key-identifier exploitation*. It requires the presence of key identifiers in the dataset. The accuracy of such inferences range from low to high depending on the relationship type that the data represents (i.e., one-to-many or one-to-one). For example, a location representing a block of flats links it to many households (and even more people) whereas an address in a single-family residential area only links the location to a small number of family members. Other key identifiers represent a strict one-to-one relationship (e.g., a fingerprint or iris scan). Datasets collected by a governmental institution are more likely to contain such key identifiers, while subjects such as Alice have little control over the inferences that the institution can draw about them.

Individuals may be identified if the data comprise a combination of quasi-identifiers in the dataset that allows for the unique identification of subjects (i.e., *combine to uniqueness*). Unlike pseudonyms, quasi-identifiers are real attributes such as sex and age, which can be further processed or linked to external data to

disclose the subject's identity. Such disclosure may occur if hospitals share their medical records with governmental institutions such as a country's census bureau (Cell (4) relation). A hypothetical *Dd* + contains attributes such as the date of visit, age, gender, occupation, municipality of residence, and final diagnosis. A data analyst from the census bureau can identify a unique combination of quasi-identifiers in which there is a visitor diagnosed with a given disease who is male, lives in a known municipality, and has a known professional occupation. The combination of such facts in a certain municipality may lead to unique subject identification with a simple Internet search. However, only a fraction of the subjects may be identified in this way.

As explained in Sect. 25.2, in examples regarding *Dd* and *Dd* +, *re-engineering of locations* is performed using geoprocessing and spatial analysis techniques. When these locations represent private information, re-engineering of location implies identification of the attacked. For example, a researcher publishes a map of the distribution of pregnant teenagers in a study area as dots on a map (a Cell (2) relation). The map is georeferenced to a known coordinate system, and the dots are digitized as circles. Then, the centroid of each circle can be extracted as a single location. Geocoding can be used to reveal the addresses of the studied teenagers.

The analysis of locations of individuals may yield various inferences including the location of their home, which is a key identifier for a data subject. When key identifiers are inferred or re-engineered, the risk of identification is typically lower than when the key identifier is available in the dataset because of possible errors and inaccuracy in the inferencing processes. For example, an LBS stores the time and location of all user service requests (a Cell (3) relation). An attacker who has access to data on service requests may wish to infer the home locations of the users. First, the attacker excludes all service requests during working hours and weekends and splits the dataset by user. The remaining datasets represent sets of possible home locations for each user—requests sent at night and during weekdays, where people are more likely to be at home. The following analysis may be repeated for each user separately: (a) apply spatial clustering and identify the cluster with the highest density and (b) extract the point with the smallest accumulated distance to all other points (i.e., a point set centroid) within the highest density cluster. The extracted point is inferred as the home location of the user.

Anonymized data may disclose information if they yield homogeneous groups of subjects regarding their private attributes. This strategy is referred to as a *homogeneity attack* and requires that a dataset (either in its current form or after processing) includes a private attribute of categorical or ratio scale. For example, a researcher collects Twitter data during a three-month period (a Cell 2 relation). The home location of subjects is estimated using spatial analysis and the subjects' political preference (i.e., a categorical private attribute) is inferred using natural language processing and machine learning techniques. The researcher publishes the dataset in anonymized form, aggregating the home locations to zip code, including the political preference, and excluding all Twitter-relevant information (e.g., account names). An attacker knows a subject who uses Twitter frequently and where this person lives. However, all records associated with the zip code of the subject display a single

political preference. Thus, that subject's political preference is disclosed due to a lack of diversity in the private attribute.

A *background attack* is possible when an attacker has knowledge (in the form of background information) on the distribution of a private attribute. For instance, mobile operators collect call data records that contain the location, time and a user identifier for each call (a random UID distinguishes users) (a Cell (3) relation). The operator can apply spatiotemporal analytics to infer the most visited venue during weekends for each subject. Anonymized copies of the data may be shared with another corporation for advertising purposes. The operator may have aggregated subject home locations by zip code (the home location is already known to the operator because of contract information), and may include visited venues during weekends in addition to other information. An attacker from the corporation knows that a subject is in the dataset and may know their home address. In the records of the zip code of the known person, it is possible that four different restaurants are revealed as frequently visited. The attacker knows that due to the subject's religion, three out of the four restaurants are unlikely. Thus, private information about the user is disclosed using background information.

The term *composition attack* refers to a privacy breach that occurs when exploiting independent anonymized datasets from different sources that involve overlapping subject populations (Ganta et al. 2008). A composition attack may build on the attacker's knowledge about a subject or the distribution of the private attribute and relies on the existence of further sources of auxiliary information. For example, in the mobile operator case, a subject may visit only two restaurants due to their eating habits. The data may have been anonymized to include the zip code and the most visited venues during weekends. Because the attacker also possesses Foursquare check-in data and knows that the subject is a frequent Foursquare user, they can search the venue results within the subject's zip code. There may be six distinct venues in the second dataset but only one appears in both datasets for the same zip code, and so the most visited venue by the attacked during weekends is disclosed.

The sixth variable is the *'privacy-preserving measures'* that mitigate an attack strategy by controlling the final digital outputs (Table 25.3). Data holders with full control of the transformation process may apply various privacy-preserving measures. Alice, as a sophisticated attacked, should consider the attacker's strategies and the privacy-preserving measures and intervene in her outputs by controlling, blurring, or censoring her digital behavior. The degree to which this is possible depends on her ability to control the transformation process (see Table 25.1). Next, we discuss five measures at her disposal, namely, (a) pseudonymity, (b) k-anonymity, (c) spatial k-anonymity, (d) l-diversity, and (e) differential privacy.

Pseudonymity is the use of pseudonyms as identifiers (or as key identifiers) (Pfitzmann and Köhntopp 2001). Unlinked pseudonyms are fake identities associated with data subjects. A pseudonym can be used to permit a subject's distinguishability, such as a UID as a random number. If distinguishability is not needed, given the use forms of the data, all key identifiers should be removed. However, if we consider that the attacker can apply strategies beyond *key identifier exploitation*, such as *combine to uniqueness,* pseudonymity mitigates but does not eliminate disclosure risk. *Combine*

to uniqueness can be prevented with *k-anonymity*, which ensures that any subject is a member of a group of size *k* with the same values of the quasi-identifiers (Samarati and Sweeney 1998). Thus, a *key-identifier exploitation* attack is mitigated by a *k* level of anonymity. The larger the *k*, the more difficult it is to identify a subject.

A similar measure to *k*-anonymity is *spatial k-anonymity*, in which a location cannot be distinguished among *k-1* other locations. This can mitigate the risk from analyzing locations, and its application varies depending on the data type. For example, to prevent re-engineering from a *Dd*, every location should be an approximation of *k* locations (such as residential addresses) within an area. In this case, randomly displacing residential addresses based on some uniform distribution is preferable over a normal distribution because the latter may provide hints to potential attackers (see Sect. 25.2). To prevent the inference of home locations from an *STd*, each subject's location should ambiguously map information to at least *k* other subjects for every moment in time. This approach can be done by decreasing the spatial resolution.

Machanavajjhala et al. (2006) showed that *k*-anonymity mitigates but does not prevent identification due to homogeneity and background attacks. The authors proposed the *l-diversity* privacy measure, which requires a *k*-anonymous dataset to have at least *l* 'well-represented' values for the *private* attributes in each equivalence class. The characteristic *l* is the minimum number of times a value of a private attribute appears in a dataset. The last measure is *differential privacy*, which guarantees that any disclosure from the data does not change significantly due to the absence or presence of a subject in the database (Dwork 2006). *Differential privacy* returns answers to aggregate queries and, according to Ganta et al. (2008), certain variations of the measure may satisfy conditions to prevent *composition attacks*.

25.5 Toward a Sociocultural Understanding of Privacy

In the previous sections, we explored the *Ethics of Where* from the point of view of Alice, an individual complying with the ACM Code of Ethics and/or the rules of a GDPR-compliant European university. Alice's technological sophistication enables her to control (part of) the transformation process (from volunteered/observed to inferred information) and preserve her privacy from attackers (Table 25.3), as well as the privacy of her research subjects (Table 25.2). Her knowledge of GDPR legislation reassures her that the behavior of corporations and government institutions is controlled by law and enforced by sanctions. GDPR instruments (e.g., DPIA) enable her to lodge complaints to preserve her privacy as a private or public sector employee. She may tackle perceived privacy breaches of the data protection legislation by alerting her representative in the legislature, by joining a collective movement of peaceful protest or by bringing a case of privacy violation to a court of law, as in *Kyllo v. United States*.

In the future, we should tackle privacy at the sociocultural level, starting from a basic premise in social theory, as Alice's (privacy) preferences and commitments are shaped by and shape the culture of her community and society (Georgiadou et al.

2019). Her individual preferences and the culture—i.e., the shared beliefs, attitudes, way of life, or world view—of the community or society in which she is socialized are deeply enmeshed and mutually reinforcing, and there is no way to determine the dependent and independent variables. This means that we should consider privacy a social construction to account for the substantial differences in social organization in countries around the world, each with different preferred ways of social organizing and different attitudes to privacy. We may distinguish four ideal types of social organizing—individualist, hierarchist, egalitarian, or fatalistic (Douglas and Wildavsky 1983). Each type is supported by (and supports) a 'cultural bias': a compatible pattern of perceiving, justifying, and reasoning about nature, human nature, justice, risk, blame, and privacy. These ideal types do not exist in unadulterated form, but can help us identify which hybrids may be most effective in which institutional settings, and how these hybrids change over time.

Individualists tend to frame information privacy as a product that can be exchanged in the marketplace for a fair price. An excellent recent example of this approach is the advocacy of the GenerationLibre think tank (Laurent 2018) to extend the private property paradigm to personal data. GenerationLibre aspires to change the way the digital ecosystem works by giving user-producers: *"(1) The possibility for e-citizens to negotiate and conclude contracts with the platforms (possibly* via *intermediaries) regarding the use of their personal data, so that they can decide for themselves which use they wish to make of them; (2) The ability to monetise these data (or not) according to the terms of the contract (which could include licensing, leasing,* etc.*); (3)The ability, conversely, to pay the price of the service provided by the platforms without giving away our data (the price of privacy?)"* (p. 7).

Hierarchists may be willing to surrender some of their privacy to a legal/rational authority (e.g., government) they trust in exchange for another public good they value, e.g., security or economic growth. The Chairperson of the German Social Democratic Party (SPD), Andrea Nahles (2018), framed the problem: *"Empires like Google and Amazon cannot be beaten from below. No start-up can compete with their data power and cash. If you are lucky, one of the big Internet whales will swallow your company. If you are unlucky, your ideas will be copied."* Her solution is a Data-for-all law: *"The dividends of the digital economy must benefit the whole society. An important step in this direction: we [the state] must set limits to the internet giants if they violate the principles of our social market economy. […] A new data-for-all law could offer decisive leverage: As soon as an Internet Company achieves a market share above a fixed threshold for a certain time period, it will be required to share a representative, anonymized part of their data sets with the public. With this data other companies or start-ups can develop their own ideas and bring their own products to the market place. In this setting the data are not "owned" exclusively by e.g. Google, but belong to the general public."* However, as Morozov (2018) argues, Nahles' agenda *"needs to overcome a great obstacle: citizens' failing trust in the state as a vehicle of advancing their interests,"* especially in a country such as Germany with a long history of data privacy activism.

Morozov (2018) argues for an egalitarian approach to privacy as constitutive of who we are and as radical citizen empowerment. *"We should not balk at proposing*

ambitious political reforms to go along with their new data ownership regime. These must openly acknowledge that the most meaningful scale at which a radical change in democratic political culture can occur today is not the nation state, as some on the left and the right are prone to believe, but, rather the city. The city is a symbol of outward-looking cosmopolitanism–a potent answer to the homogeneity and insularity of the nation state. Today it is the only place where the idea of exerting meaningful democratic control over one's life, however trivial the problem, is still viable." Similarly, the Oxford-based *Digital Rights to the City* group proposes a deeper meaning to the right to information that amounts to the declaration that *"we will no longer let our information be produced and managed for us [presumably by the state or corporations], we will produce and manage our information ourselves"* (Shaw and Graham 2017). Fatalists are those persuaded by the abovementioned slogans *"you have zero privacy…get over it"* or *"if you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."* However, as Snowden said, *"arguing that you don't care about the right to privacy because you have nothing to hide is no different than saying you don't care about free speech because you have nothing to say"* (Reddit 2015).

25.6 Toward Digital Earth Ethics: The Ethics of Where

In the previous sections, we mentioned privacy arrangements in the legal systems of two polities—the United States and the European Union—a serious limitation in a chapter on Digital Earth Ethics that encompasses the entire planet. However, it is possible to see how privacy is dealt with differently in these two cases. The word privacy is not mentioned in the US Constitution except indirectly in the Fourth Amendment, which protects the right of people to be secure in their persons, houses, papers, and effects against unreasonable searches and seizures. In contrast, in the European Union, privacy is a human right according to Article 8 of the European Convention on Human Rights: a *"right to respect for private and family life, home and correspondence."* This is largely due to the events around World War II, where personal information was often used to target individuals and groups and facilitate genocide. In 2018, we witnessed a serious shake-up of the treatment of privacy, data protection, and cybersecurity by legal systems around the world. The EU's GDPR, put in place in 2018, is a landmark development for privacy and how we perceive it. In this transitional period, a number of countries seem to follow a similar pathway as the GDPR: for instance, Canada, Japan, and India are looking at comparable extraterritorial privacy regimes. A common denominator between them is privacy as a constitutional right. Similar legislative developments are manifesting in China and the African continent.

In China, the Cybersecurity Law, the most important Internet legislation to be passed in the country thus far, came into effect on June 1, 2017. The law is intended to align China with global best practices for cybersecurity. Network operators must store select data within China and Chinese authorities may conduct spot-checks on

a company's network operations. *"The Cybersecurity Law provides citizens with an unprecedented amount of protection to ensure their data privacy. The law defines "personal information" as information that can be used on its own or in conjunction with other information to determine the identity of a natural person, including but not limited to a person's name, birthday, identity card number, biological identification information, address, and telephone number. In other words, once such information is de-identified, it will no longer be subject to the requirement for personal information in the Cybersecurity Law"* (Lee 2018, p. 87). Other countries such as Korea and Singapore are less decided and may be consciously delaying their legislative moves until the scene becomes clearer.

In the African continent, approximately 40% of the countries have enacted data protection legislation that abides the OECD standards (1st generation), the EU DPD 1995 standards (2nd generation), or even features GDPR elements (3rd generation). The latter refers to Mauritius, one of Africa's dynamic but small economies, which updated its 2004 law in 2017 with a new Data Protection Act 2017 featuring elements of the GDPR. In June 2014, the African Union (AU) adopted the Convention on Cyber Security and Personal Data Protection, known as the Malabo Convention, the first treaty outside the EU to regulate the protection of personal data at a continental level. The Convention aims to establish regional and national legal frameworks for cybersecurity, electronic transactions and personal data protection, but its actual impact will depend on ratifications, which had not occurred by early 2016. In 2018, the AU created data protection guidelines that are broadly aligned with the GDPR for its Member States, with contributions from regional and global privacy experts including industry privacy specialists, academics and civil society groups (Georgiadou et al. 2019). On a global scale, there is a substantial imbalance in sensitive data flows, with mostly American Internet tech companies sourcing data globally. This imbalance is the substrate for a continuation of developments in technology, the legal scenery and contractual arrangements that we do not expect to be settled soon. Unfortunately, privacy and data protection as global goods intersect with cybersecurity and counterterrorism, which gives little hope for transparency and focus on solutions. Nevertheless, we should follow these developments closely (Raul 2018).

In addition to legislative efforts, global and regional institutions are busy developing ethical principles and guidelines. The *UNESCO Declaration of Ethical Principles in relation to Climate Change* addresses the responsibility to overcome the challenges and reinforces ethics at the center of the discussion on climate change. Member states have mandated UNESCO with promoting ethical science: science that shares the benefits of progress for all, protects the planet from ecological collapse and creates a solid basis for peaceful cooperation. The *Global Ethics Observatory (GEObs)*, a system of databases with worldwide coverage in bioethics and environmental ethics, science ethics, and technology ethics, helps researchers identify Who's Who in Ethics, Ethics Institutions, Ethics Teaching Programs, Ethics-Related Legislation and Guidelines, Codes of Conduct and Resources in Ethics. Other global actors in the responsible data movement, e.g., UN Global Pulse (2017), Red Cross/Red Crescent 510 (2018) and UNOCHA (2019), also develop data ethics guidelines as a cornerstone of their groundwork.

At the European Union level, the *High-Level Expert Group on Artificial Intelligence (AI HLEG)* proposed the first draft AI ethics guidelines to the European Commission in December 2018. These cover issues such as fairness, safety, transparency, the future of work, democracy and the impacts of application of the Charter of Fundamental Rights, including privacy and personal data protection, dignity, consumer protection and nondiscrimination. The *European Group on Ethics in Science and New Technologies (EGE)* provides the Commission with high-quality, independent advice on ethical aspects of science and new technologies in relation to EU legislation or policies. The EGE is an independent advisory body founded in 1991 and is tasked with integrating ethics at the international level, at the interinstitutional level with the European Parliament and the Council, and within the Commission itself. The *European Union Agency for Fundamental Rights (FRA)* is the EU's center of fundamental rights expertise. It helps ensure that the fundamental rights of people living in the EU are protected. Fundamental rights set minimum standards to ensure that a person is treated with dignity. The Agency seeks to instill a fundamental rights culture across the EU by collecting pertinent and timely data and information, by sharing evidence-based insights and advice with policy- and decision-makers, by raising rights awareness and promoting fundamental rights through cutting-edge communications, and by engaging with a wide array of diverse stakeholders from local and international levels.

However, requiring global and regional guidelines and principles to conceptualize ethics and privacy across diverse cultural and political contexts and to substitute for lacking or existing weakly enforced legislation (Taylor 2017) may further deplete local institutional capacity and harm citizens. The question of how to improve digital data flows between local and global institutions while maximizing the use of innovative geospatial technologies and protecting citizens' rights as well as local institutions is particularly relevant in view of the increasing use of artificial geospatial intelligence, mobile technology and social media to extract information about individuals and communities.

Finally, legal and social strategies and privacy-preserving technological measures should form the backbone of university curricula for students and working professionals. The advent of the GDPR in 2018 created a sudden educational demand for university courses, of which the massive open online course (MOOC) *'Privacy by Design and GDPR'*, designed by computer scientists at Karlstad University in Sweden, is a recent EU-focused example (Fischer-Hübner et al. 2018). The educational challenge is to gradually expand the content of such courses for a global audience of students from countries with emergent privacy and data protection legal frameworks, different understandings of privacy and different social organization as well as different levels of technological sophistication in countering privacy attacks, because *The Ethics of Where* should eventually be everywhere.

Acknowledgements We thank the anonymous reviewers for their insightful comments and suggestions to probe deep, and we are grateful to Deirdre Mulligan for allowing us the use of the extensive quote on the Kyllo v. United States case in Sect. 25.1. Our colleague Marga Koelen continues to be a marvelous traveling companion into the land of ethics and location privacy.

References

ACM Code 2018 Task Force (2018) ACM code of ethics and professional conduct. https://www.acm.org/code-of-ethics. Accessed 3 Apr 2019

Agre EP, Rotenberg M (1997) Technology and privacy: the new landscape. MIT Press, Cambridge, MA

Alrayes F, Abdelmoty A (2014) No place to hide: a study of privacy concerns due to location sharing on geo-social networks. Int J Adv Secur 7(3/4):62–75

Baker M (2016) Is there a reproducibility crisis? Nature 533:452–454

Beck J (2018) People are changing the way they use social media. The Atlantic. https://www.theatlantic.com/technology/archive/2018/06/did-cambridge-analytica-actually-change-facebook-users-behavior/562154/. Accessed 3 Apr 2019

Bowker GC, Baker K, Millerand F et al (2010) Toward information infrastructure studies: ways of knowing in a networked environment. In: Hunsinger J, Klastrup L, Allen M (eds) International handbook of internet research. Springer, Dordrecht, pp 97–117

Brownstein JS, Cassa CA, Kohane IS et al (2006) An unsupervised classification method for inferring original case locations from low-resolution disease maps. Int J Health Geogr 5(1):56

Cassa CA, Wieland SC, Mandl KD (2008) Re-identification of home addresses from spatial locations anonymized by Gaussian skew. Int J Health Geogr 7(1):45

Ciborra CU (1985) Reframing the role of computers in organizations: the transactions costs approach. In: Gallegos L, Welke R, Wetherbe J (eds) Proceedings of the 6th international conference on information systems. Indianapolis, IN, pp 57–69

Confessore N, Dance GJX, Harris R et al (2018) The follower factory. New York Times. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html. Accessed 3 Apr 2019

Cukier K, Mayer-Schoenberger V (2013) The rise of big data: how it's changing the way we think about the world. Foreign Aff 92(3):28–40

De Montjoye YA, Hidalgo CA, Verleysen M et al (2013) Unique in the crowd: the privacy bounds of human mobility. Sci Rep 3:1376

Dencik L, Hintz A, Cable J (2016) Towards data justice? The ambiguity of anti-surveillance resistance in political activism. Big Data Soc 3(2):1–12

Der Spiegel (1983) Volkszählung: laßt 1000 fragebogen glühen. http://www.spiegel.de/spiegel/print/d-14022649.html. Accessed 3 Apr 2019

Der Spiegel (1987) Datenschrott für eine milliarde? http://www.spiegel.de/spiegel/print/d-13522320.html. Accessed 3 Apr 2019

Douglas M, Wildavsky A (1983) Risk and culture: an essay on the selection of technical and environmental dangers. University of California Press, Berkeley, CA

Duckham M, Kulik L (2006) Location privacy and location-aware computing. In: Drummond J, Billen R, Joao E et al (eds) Dynamic & mobile GIS: investigating change in space and time. CRC Press, Boca Raton, FL, pp 35–51

Dwork C (2006) Automata, languages and programming. In: Bugliesi M, Preneel B, Sassone V et al (eds) 33rd international colloquium, ICALP 2006. Springer, Venice, Italy, p 112

Ehlers M, Woodgate P, Annoni A et al (2014) Advancing digital earth: beyond the next generation. Int J Digit Earth 7(1):3–16

Fischer-Hübner S, Martucci LA, Fritsch L et al (2018) A MOOC on privacy by design and the GDPR. In: Drevin L, Theocharidou M (eds) Information security education – towards a cybersecure society. Springer, Cham, pp 95–107

Floridi L (2014) The fourth revolution: how the infosphere is reshaping human reality. Oxford University Press, Oxford, UK

Fried C (1968) Privacy. Yale Law J 77(3):475–493

Gambs S, Killijian M-O, Cortez MNDP (2010) Show me how you move and I will tell you who you are. In: Proceedings of the 3rd ACM SIGSPATIAL international workshop on security and privacy in GIS and LBS. ACM, San Jose, California, pp 34–41

Ganta SR, Kasiviswanathan SP, Smith A (2008) Composition attacks and auxiliary information in data privacy. In: Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, Las Vegas, Nevada, pp 265–273

Georgiadou Y, De By RA, Kounadi O (2019) Location privacy in the wake of the GDPR. ISPRS Int J Geo-inf 8(3):157

Glanville J (2018) The journalistic exemption. London Rev Books 40(3):9–10

Goffman E (1957) The presentation of self in everyday life. Anchor Books, New York, NY

Goodchild MF, Guo H, Annoni A et al (2012) Next-generation digital earth. Proc Natl Acad Sci U S A 109(28):11088–11094

Greenwald G, MacAskill E (2013) NSA prism program taps into user data of Apple, Google and others. The Guardian, October 1, 2016

Halpern S (2013) Are we puppets in a wired world? https://www.nybooks.com/articles/2013/11/07/are-we-puppets-wired-world/. Accessed 3 Apr 2019

Halpern S (2018) The known known. https://www.nybooks.com/articles/2018/09/27/privacy-technology-known-known/. Accessed 3 Apr 2019

Harley JB (1991) Can there be a cartographic ethics? Cartogr Perspect 13(10):9–16

Hennepadvocaat Website (2019) Warmtebeeldcamera mag gebruikt worden voor opsporen hennepkwekerij. https://hennepadvocaat-hennepkwekerij.nl/warmtebeeldcamera-mag-gebruikt-worden-voor-opsporen-hennepkwekerij/. Accessed 6 Apr 2019

Herrmann M (2016) Privacy in location-based services. PhD Dissertation. KU Leuven–Faculty of Engineering Science

Lee J-A (2018) Hacking into China's cybersecurity law. https://ssrn.com/abstract=3174626 Accessed 3 Apr 2019

Johnson B (2010) Privacy no longer a social norm, says Facebook founder. https://www.theguardian.com/technology/2010/jan/11/facebook-privacy. Accessed 3 Apr 2019

Keßler C, McKenzie G (2018) A geoprivacy manifesto. Trans GIS 22(1):3–19

Kounadi O, Resch B (2018) A geoprivacy by design guideline for research campaigns that use participatory sensing data. J Empir Res Hum Res Ethics 13(3):203–222

Laurent D (2018) My data are mine: why we should have ownership rights on our personal data. https://www.generationlibre.eu/wp-content/uploads/2018/01/Rapport-Data-2018-EN-v2.pdf. Accessed 3 Apr 2019

Leitner M, Mills JW, Curtis A (2007) Can novices to geospatial technology compromise spatial confidentially. Kartogr Nachr 57(2):78–84

Li H, Zhu H, Du S et al (2018) Privacy leakage of location sharing in mobile social networks: attacks and defense. IEEE Trans Dependable Secure Comput 15(4):646–660

Liddell HG, Scott R (1940; revised version 2018) A Greek-english lexicon; machine readable text. Trustees of Tufts University, Oxford

Machanavajjhala A, Gehrke J, Kifer D et al (2006) L-diversity: privacy beyond k-anonymity. In: 22nd International Conference on Data Engineering (ICDE'06). IEEE, Atlanta, GA, p 24

Masser I, Wegener M (2016) Brave new GIS worlds revisited. Env Plan B 43(6):1155–1161

Morozov E (2018) There is a leftwing way to challenge big tech for our data. Here it is. https://www.theguardian.com/commentisfree/2018/aug/19/there-is-a-leftwing-way-to-challenge-big-data-here-it-is. Accessed 3 Apr 2019

Mulligan DK, Koopman C, Doty N (2016) Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy. Philos Trans A Math Phys Eng Sci 374(2083):20160118

Nahles A (2018) Die tech-riesen des silicon valleys gefährden den fairen Wettbewerb. https://www.handelsblatt.com/meinung/gastbeitraege/gastkommentar-die-tech-riesen-des-silicon-valleys-gefaehrden-den-fairen-wettbewerb/22900656.html?ticket=ST-775717-C1dsgQEyOCBKDWwzRAVA-ap4. Accessed 3 Apr 2019

Newman J (2009) Google's schmidt roasted for privacy comments. https://www.pcworld.com/article/184446/googles_schmidt_roasted_for_privacy_comments.html. Accessed 3 Apr 2019

Ouchi WG (1979) A conceptual framework for the design of organizational control mechanisms. Manag Sci 25(9):833–848
Pfitzmann A, Köhntopp M (2001) Anonymity, unobservability, and pseudonymity—a proposal for terminology. In: Federrath H (ed) International conference proceedings on designing privacy enhancing technologies. Springer, Berlin, Heidelberg, pp 1–9
Preoţiuc-Pietro D, Cohn T (2013) Mining user behaviours: a study of check-in patterns in location based social networks. In: Proceedings of the 5th annual ACM web science conference. ACM, Paris, France, pp 306–315
Raul AC (2018) The privacy, data protection and cybersecurity law review, edition 5. The Law Reviews website. https://thelawreviews.co.uk/edition/1001264/the-privacy-data-protection-and-cybersecurity-law-review-edition-5. Accessed 3 Apr 2019
Red Cross/Red Crescent 510 (2018) 510 Data Responsibility Policy. Version 2.0. Available from https://www.510.global/wp-content/uploads/2017/11/510_Data_Responsibility_Policy_V.2_PUBLIC-1.pdf. Accessed 12 Aug 2019
Reddit (2015) Ask me anything session on reddit, featuring Edward Snowden and Jameel Jaffer. https://www.reddit.com/r/IAmA/comments/36ru89/just_days_left_to_kill_mass_surveillance_under/. Accessed 3 Apr 2019
Romm T (2018) Apple's Tim Cook blasts Silicon Valley over privacy issues. https://www.washingtonpost.com/world/europe/apples-tim-cook-delivers-searing-critique-of-silicon-valley/2018/10/24/5adaa586-d6dd-11e8-8384-bcc5492fef49_story.html?utm_term=.9ec79fe25f7a. Accessed 3 Apr 2019
Samarati P, Sweeney L (1998) Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical Report SRI-CSL-98–04. In: SRI Computer Science Laboratory. Palo Alto, CA, pp 101–132
Shaw J, Graham M (2017) An informational right to the city? Code, content, control, and the urbanization of information. Antipode 49(4):907–927
Siegel E (2013) Predictive analytics: the power to predict who will click, buy, lie, or die. Wiley, Hoboken, NJ
Sprenger P (1999) You have zero privacy … get over it. https://www.wired.com/1999/01/sun-on-privacy-get-over-it/. Accessed 3 Apr 2019
Taylor L (2017) What is data justice? The case for connecting digital rights and freedoms globally. Big Data Soc 4(2):2053951717736335
The Guardian (2018) The Cambridge Analytica files. https://www.theguardian.com/news/series/cambridge-analytica-files. Accessed 3 Apr 2019
Tiku N (2018) Europe's new privacy law will change the web, and more. https://www.wired.com/story/europes-new-privacy-law-will-change-the-web-and-more/. Accessed 3 Apr 2019
UN Global Pulse (2017) Privacy and data protection principle. Available from https://www.unglobalpulse.org/privacy-and-data-protection-principles. Accessed 12 Aug 2019
UNOCHA (2019) Data Responsibility in Humanitarian Action: Building trust through dialogue Available from https://www.unocha.org/story/data-responsibility-humanitarian-action-building-trust-through-dialogue. Accessed 12 Aug 2019
U.S. Government Publishing Office. (2009). 45 CFR 164.514—Other requirements relating to uses and disclosures of protected health information. Available from https://www.gpo.gov/fdsys/pkg/CFR-2009-title45-vol1/xml/CFR-2009-title45-vol1-sec164-514.xml/. Accessed 12 Aug 2019
US Supreme Court (2001) Kyllo v. United States. 533 U.S. 27, 121 S. Ct. 2038, 150 L. Ed. 2d 94. https://supreme.justia.com/cases/federal/us/533/27/. Accessed 3 Apr 2019
Van der Sloot B (2014) Privacy as human flourishing: could a shift towards virtue ethics strengthen privacy protection in the age of Big Data. J. Intell. Prop. Info. Tech. & Elec. Com. L. 5:230
Van Lieshout M (2018) Belastingdienst eist inzage Museumkaart om belastingplichtige kaarthouder op te sporen. https://www.volkskrant.nl/nieuws-achtergrond/belastingdienst-eist-inzage-museumkaart-om-belastingplichtige-kaarthouder-op-te-sporen~bf250db4/. Accessed 12 Feb 2019

Vicente-Saez R, Martinez-Fuentes C (2018) Open science now: a systematic literature review for an integrated definition. J Bus Res 88:428–436
Warren SD, Brandeis LD (1890) The right to privacy. Harv Law Rev 4(5):193–220
Wegener M, Masser I (1996) Brave new GIS worlds. In: Masser I, Campbell H, Craglia M (eds) GIS diffusion: the adoption and use of geographical information systems in local governments in Europe. Taylor & Francis, London, UK, pp 9–22
Westin AF (1967) Privacy and freedom. Atheneum, New York, NY
World Economic Forum (2011) Personal data: the emergence of a new asset class. World Economic Forum in Collaboration with Bain & Company, Inc., Cologny, Switzerland
Zeile P, da Silva AR, Papastefanou G et al (2011) Smart sensoring as a planning support tool for barrier free planning. In: Proceedings of 12th international conference on computers in urban planning and urban management (CUPUM)
Zook M, Barocas S, Boyd D et al (2017) Ten simple rules for responsible big data research. PLoS Comput Biol 13(3):e1005399

Yola Georgiadou is Professor in Geo-information for Governance at the ITC Faculty, University of Twente, The Netherlands. Yola studies how social actors structure policy problems characterized by intense disagreement on values and uncertain spatial knowledge. Her methods are qualitative. Her normative orientation is "working with the grain" of local institutions.

Ourania Kounadi is an Assistant Professor in GIScience at the ITC Faculty, University of Twente. Her research interests include location privacy, spatial confidentiality, and spatial crime analysis. Regarding location privacy, she has conducted literature reviews, methodological and empirical studies. Currently, she is examining application tools to allow safe collection and dissemination of geographic information.

Rolf A. de By is an Associate Professor in Geospatial Information Processing at the ITC Faculty, University of Twente. His research interests are in the domain of applying advanced information systems that make use of earth observation and other geospatial data sources to important societal problems, especially those occurring in developing nations.