UrbanData 2 Decide
Integrated Data Visualisation and Decision Making

D2.4 SOCIAL, ETHICAL AND LEGAL ASPECTS OF BIG DATA AND URBAN DECISION MAKING

PROJECT

Acronym:	**UrbanData2Decide**
Title:	SOCIAL, ETHICAL AND LEGAL ASPECTS OF BIG DATA AND URBAN DECISION MAKING
Coordinator:	SYNYO GmbH
Reference:	847511
Type:	Joint Programme Initiative
Programme:	Urban Europe
Start:	September 2014
Duration:	26 months
Website:	http://www.urbandata2decide.eu
E-Mail:	office@urbandata2decide.eu

Consortium:
SYNYO GmbH, Research & Development Department, Austria (SYNYO)
University of Oxford, Oxford Internet Institute, UK (OXFORD)
Malmö University, Department of Urban Studies, Sweden (MU)
Open Data Institute, Research Department, UK (ODI)
IT University of Copenhagen, Software Development Group, Denmark (ITU)
ZSI Centre for Social Innovation, Department of Knowledge and Technology, Austria (ZSI)

SYNYO UNIVERSITY OF OXFORD MALMÖ UNIVERSITY ODI open data institute IT University of Copenhagen ZSI

DELIVERABLE

Number:	**D2.4**
Title:	**SOCIAL, ETHICAL AND LEGAL ASPECTS OF BIG DATA AND URBAN DECISION MAKING**
Lead partner:	ZSI
Work package:	WP2: Basic Exploration, Stakeholder Studies and Requirement Analysis
Date:	May 2015
Authors:	Susanne Dobner, ZSI Christian Voigt, ZSI
Contributors:	Markus Rasmusson, MU Joshua Ddamba, ITU Ulrich Atz, ODI Richard Norris, ODI
Reviewers:	Markus Rasmusson, MU Per Olof Hallin, MU Nicklas Guldacker, MU

The UrbanData2Decide project is co-funded under the Joint Programming Initiative, 2nd call Urban Europe.

TABLE OF CONTENT

1 INTRODUCTION

The rise of information technologies, the amount of information and data gathered daily alongside unprecedented forms of online communication and participation (via social media channels or blogs) have been increasingly challenging our notions of privacy. Given the focus on data in the 'UrbanData2Decide' project, it is imperative to discuss the social, ethical and legal issues around big data in greater depth. 'UrbanData2Decide' is collecting data from different data sources, including open data, social media data and potentially data with restricted public access, e.g. local crime data, as well as interpreting combinations of these various data sources. Consequently many privacy issues are raised and must be addressed within the project team consecutively. This report complements earlier project reports on 'Data Sources and Visualisation Methods' (D2.1), 'Urban Decision Making and Expert Integration Report' (D2.2) as well as the report on 'Stakeholder roles, workflows and requirements' (D2.3).

This report focuses on the **social, ethical and legal aspects of big data**. Each of these three aspects includes different dimensions of privacy concerns.

- The first chapter of the report is dedicated to introducing current privacy debates. An important insight can be gained by asking 'what would we lose if we do not have privacy?' In addition prevailing concepts of privacy, namely descriptive and normative concepts, as well as the contextual integrity framework by Nissenbaum are described.

- The second chapter, 'Informational privacy as a social construct' introduces several definitions of privacy, while emphasizing on the difficulties of finding an all-encompassing definition of privacy. Moreover, we discuss three main types of privacy encroaching technologies with current practical examples: tracking and monitoring, aggregation and analysis, and large-scale dissemination. . Further, questions of widespread privacy concerns vs. actual privacy actions and gender-specific, diverse privacy actions are discussed.

- The third chapter 'Internet research as ethical practices' puts a special focus on privacy preservation or research participants (including the confidentiality of expert inputs), the liabilities of platform providers and the impact software design can have on the democratic character of big data interpretation and decision-making. In addition, it outlines the benefit of a participatory design approach and supporting privacy self-management skills to address the above-mentioned challenges.

- The fourth chapter 'Social, Ethical and Legal Guidelines' offers a set of guidelines in regards to social privacy concerns, ethical guidelines, with a specific emphasis on software engineering, and legal guidelines. The legal guidelines start out by discussing international privacy framework by OECD and European Union before briefly describing the three main EU Directives on Privacy and Data Protection as well as their national implementation in the case study countries: Austria, UK, Sweden and Denmark.

- The fifth chapter summarizes the main findings from previous chapters and draws conclusions in regards to privacy concerns within the 'UrbanData2Decide' Project.

1.1 What is at stake?

Privacy is not a new concern, but has over the past centuries been shaped by and co-evolved with new technologies. It has thus continuously developed with the rise of information technology. In 1890 Samuela Warren and Louis Brandeis warned in their article published in the Harvard review: "*Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of the private and domestic life* […]" which triggered a first public discussion about the legal right to privacy. Since the beginning debates about privacy have been bound to technologies (DeCew 1997). In the case above privacy concerns are tied to the collection, publication and reproduction of photographs. Fast forwarding two centuries, through the rise and proliferation of the Internet, privacy discussions and attempts to conceptualize it have gained unprecedented momentum. However, the growing momentum does not prevent privacy to quickly turn into an empty term, especially if the respective context is missing. For instance it makes a difference whether revealing personal health details in case of an accident (e.g. if a special operation is required) or to pass health details on to your insurance (e.g. which may affect your insurance claims). Debates around 'what is privacy and where does it apply to?' often ask whether privacy can or should be considered as a claim, a right, an interest, a (personal and/or societal) value, a preference, currently merely a state of existence? Additionally, Nissenbaum (2010) asks whether privacy applies equally to informal expressions, formal statements and actions.

In order to better understand what protecting personal privacy can mean, Van den Hoven (2001) turns prevailing debates about privacy, (i.e. what is it actually that we feel is threatened?') around by asking '**What would we lose if we did not have privacy?**'

Van den Hoven names four moral reasons for protecting personal data:

- ***Prevention of harm*** (unrestricted access by others to passwords or personal characteristics)
- ***Informational inequality*** (personal data have become commodities, personal data transmission and use by third parties)
- ***Informational injustice and discrimination*** (personal information may change depending on the context – sharing personal health information with a doctor is something else than with an insurance or work place)
- ***Encroachment of moral autonomy*** (lack of privacy may affect individuals choices, similar discussions are found in debates about surveillance/CCTV where people tend to change their behaviour when knowing that there is surveillance)

Similarly, Wolfie (2014) discusses four societal implications caused by current digital surveillance and the loss of privacy.

- ***Losing control***
  the term 'digital fingerprint' expresses the difficulty (and often impossibility) to delete or change information that is captured digitally. Sometimes this is also referred to as leaving a 'digital trace'.

- ***Lack of transparency***
  often it is not clear who captures what kind of information, and for how long. Information about online searches or bought goods in a store might be passed on to third parties.

- ***Decontextualization***
  occurs when information is being passed on to third parties (e.g. companies) and decontextualized, for instance Facebook information being passed on to marketing companies.

- ***Misleading and wrong prediction***
  wrong prognosis models can harm individuals (e.g. when your medical history affects a job application).

All of the implications mentioned above are somewhat combined in what Boyd & Crawford (2012) include in their broader notion of a "*digital divide*", which refers to the *power imbalance* between the user (whose actions become more and more transparent) and companies (who benefit from user data in increasingly opaque). Discussing the commonly known risks and fears of losing privacy is as important as asking what kind of means and ethical, social and legal concepts it requires to safeguard privacy.

1.2 The privacy debate: Descriptive or normative?

Debates around definitions and ethics of privacy often bring various conceptualizations of privacy into focus. Attempts to conceptualize privacy have a long history and are to be found in philosophical, sociological, psychological as well as legal realms.

The two most commonly known concepts of privacy are the concepts of descriptive and normative definitions, both of which challenge and contrast different aspects of privacy.

"Still, privacy is – virtually – always used normatively."
(Robert Post, 1989 in ,The Social Foundations of Privacy')

Normative concepts question why privacy is important to society and individuals, and emphasize reasoning around protecting privacy. Also normative definitions aim to link privacy to 'higher-order

values', e.g. contrasting privacy interests of individuals and society. It thus relates privacy to other (societal) values like health or security. By discussing individual liberty versus national security, i.e. accessing and storing personal data, Etzioni (1999) states that "*privacy is not an absolute value and does not trump all other rights or concerns for the common good*" (Etzioni 1999, 38).
Two other examples of clashing 'higher-order values' are:

- personal autonomy versus freedoms of business institutions (cf. free-market economy)
- moral autonomy versus social order (cf. free speech)

The examples above demonstrate that privacy considered (only) as an individual right does not always outrank greater societal interests. Thus sometimes the common good may demand a balanced response, weighing privacy rights against other values, e.g. EU-wide data retention by mobile network providers is a much discussed issue and more specifically considered in section 4.4.6.

By contrasting privacy interest of individuals to those of society (i.e. ensuring the wellbeing or the safety of society) it can be argued that the interests of the society will in most cases outrank those of individuals. Following Etzioni's approach (1999) societies interests are always prone to undermine individual rights to privacy. However, Solove points to the fallacy of solely contrasting individual and societal interests (in which case the latter will almost always prevail):

- Firstly, it can be considered problematic to always discuss the rights of individuals to privacy as being in conflict with society's interests. Instead ensuring privacy should *"involve balancing societal interests on both sides of the scale*." (Solove 2007, 15)
- Secondly, framing privacy as opposing individual and society's interest will hardly ever favor individual rights with society mostly winning a battle that is neither fair nor always necessary.

Descriptive concepts of privacy stress the possibility of conflicts between privacy and other values. Hence in certain circumstances less privacy might be better than more and reductions in privacy need not constitute violations of individual rights. In general, privacy is not "*a binary property*" (Waldo et al. 2007, 58) i.e. either an individual has or does not have privacy, rather should we understand privacy in terms of degrees of freedom, which under some circumstances are restricted or lost. Moreover, such a neutral conception of privacy allows talking about states of increased and decreased privacy (e.g. stakeholder dependent privacy accounts), which is more practical in addressing concrete privacy issues pragmatically. The impact of changes in privacy protection can then be compared in a before and after fashion.

According to Waldo et al. (2007) two conceptions reflect the most commonly used views (especially in philosophical privacy debates), which are related to each other while at the same time offering different viewpoints:

> (1) Privacy as **restrictions on the access** of other people to an individual's personal information,

(2) Privacy as an **individual's control over** personal information, e.g. health status which often has to be shared.

Whereas the '*restrictions on access'* theory emphasizes the importance of restricted access to personal information in order to safeguard privacy, the '*control over information'* theory gives priority to the individual control over personal information. The *'restrictions on access'* theory takes into account different zones and contexts (for instance public-private) where individuals have not always the same opportunities to restrict access. Tavani (2008) criticises that the importance of control as in choosing to grant or restrict access to personal information has not been considered sufficiently. In regard to the *'control over information'* theory, a crucial aspect is how much a person is in control of his or her personal information (e.g. people may be obliged to share information like medical histories with doctors and have that information digitalized on a general health insurance card). Analyses concerning the control of individuals over their personal information aim to develop guidelines about "*the extent of what must be controlled for privacy to be maintained*" (Waldo, Lin & Millett 2007, 61).

1.3 Nissenbaum Contextual Integrity Framework

Nissenbaum's contextual integrity framework may be considered as both, a combination of the discussion around normative and descriptive concepts of privacy, as well as an attempt to go beyond the binaries of descriptive versus normative, or public versus private. With her framework Nissenbaum (2004) "*puts the context back into the equation*" and discusses differences in information sharing, e.g. 'I tell my doctor something else than the bank'. The main argument behind the contextual integrity framework is that context (e.g. the type of information and required in a given setting) sets the appropriate benchmark of privacy" (Nissenbaum 2004, 102). The "*key objective of Nissenbaum's "decision heuristics" is to provide an approach that enables us both to: (a) understand the "source or sources of trouble in new and emerging technologies*," and (b) evaluate the "system or practice in question" (Nissenbaum 2010, 181).

"*A right to privacy is neither a right to secrecy nor a right to control but a right to appropriate flow of information...but what this amounts to is a right to contextual integrity and what this amounts to varies from context to context*." (Nissenbaum 2010, 127)

According to Nissenbaum, privacy violations do not occur when 'too much' data accumulates but when principles of transmission change or are not specified. She argues that the actual problem with privacy "*is the inappropriateness of the flow of information due to the mediation of technology*" – unlike losing control over data, we should worry about a loss of control about data flows (Madrigal, 2012). Madrigal's theory is based on two principles:

(1) the activities people engage in take place in a "plurality of realms" (i.e. spheres or contexts);

(2) each realm (bank, hospital, etc.) has a distinct set of norms that govern data usage.

The contextual-integrity model proceeds on the assumption that there are "*no areas of life which are not governed by norms of information flow*" (Nissenbaum 2004, 137). There are two types of informational norms in Nissenbaum's privacy scheme:

(a) **Norms of appropriateness** relates to the given context or situation where we share or access personal information, for instance sharing information about our physical condition seems appropriate and useful in a hospital. Whereas revealing this type of information in an insurance office, with an employer or at the bank seems inappropriate. (ibid. 120f.)

(b) **Norms of distribution** concern the transfer of information among people or parties, which the suggested norms of distribution try to regulate. The informal information shared among friends (usually bidirectional) may differ from the information exchange in a healthcare context, where the confidentiality of information has to follow a multitude of norms and regulations (ibid. 122f.)

The contextual integrity framework by Nissenbaum is referenced again later in the 'Social Guidelines: 'Contextual Integrity Models' in chapter 4.1.

2 INFORMATIONAL PRIVACY AS A SOCIAL CONSTRUCT

The following chapters describe different aspects of informational privacy, including the different ways to define privacy, technologies which may encroach on personal privacy, as well as incongruences between expressed privacy concerns and the corresponding privacy actions taken by the same individual.

Firstly, different ways to define privacy are discussed, including a taxonomy of privacy by Solove identifying different privacy violations in the stages of data collection to data processing. Secondly, activities and mechanism enabling an encroachment of (personal) privacy are reflected upon. Specifically, the everyday occurrence of some privacy encroaching mechanisms (e.g. paying with credit card or using store loyalty card) as well as contradictions of privacy and convenience are discussed. Thirdly, the 'I have got nothing to hide' argument is discussed. Fourthly, gender as one amongst many factors impacting upon awareness and privacy actions is described.

2.1 Ways to define privacy

"*Currently, privacy is a sweeping concept, encompassing (among other things) freedom of thought, control over one's body, solitude in one's home, control over information about oneself, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations. Time and again philosophers, legal theorists, and jurists have lamented the great difficulty in reaching a satisfying conception of privacy*.'" (Solove 2002, 1088)

A common thread through privacy literature is the difficulty of capturing a satisfying definition of the term itself. Many authors have been discussing possible concepts of privacy, yet these are often very broad and indistinct. Some definitions go as far as defining privacy as "*the right to be alone*" (Warren & Brandeis 1890, 193) which encompasses too many things to be useful in the context of current privacy concerns (e.g. online privacy, big data and social media). Commonalities with other terms, such as intimacy lead to similar arguments. It is problematic, so Solove (2002) as not all information is considered intimate but may be regarded private (e.g. social security number). This is given the assumption that the term intimacy itself is thoroughly defined. Whereas a clear cut definition of privacy is currently not to be found, some privacy theorists, including Solove (2002) and Nissenbaum (2004) suggest putting the context to the fore, i.e. the circumstances under which privacy may be threatened or encroached instead of attempting to find an all-encompassing definition. Further, Solove argues that "privacy is not reducible to a singular essence; it is a plurality of different things that do not share one element in common but that nevertheless bear a resemblance to each other." Within their definitions of privacy, legal scholars, Anita Allen and Jerry Kang emphasize the control of and access to information as the most essential dimensions.

Privacy is "*an individual's control over the processing— i.e., the acquisition, disclosure, and use— of personal information*" (Kang 1998, 1203).

Following Allen, privacy involves three dimensions: physical privacy, informational privacy (similar to Kang) defined as "*confidentiality, secrecy, data protection and control over personal information*" and proprietary privacy, as "*control over names, likenesses and repositories of personal information*." (Allen-Castellitto 1999, 723 cit. after Nissenbaum 2010, 71) Aspects of privacy such as constraint (the degree of access others have to -personal- information) and control over personal information appear most prevalent in current privacy definitions.

Towards 'A Taxonomy of Privacy'

"*The term 'privacy' is best used as a shorthand umbrella term for a related web of things. Beyond this kind of a use, the term "privacy" has little purpose. In fact, it can obfuscate more than clarify*." (Solove 2007, 12)

In order to describe the "*related web of things*" Solove developed **a taxonomy of privacy** which includes important privacy dimensions. The taxonomy sets an attempt to map out diverse problems that constitute privacy violations, as they can occur in the stage of collecting information about a person to information processing. According to Solove, the taxonomy aims to contribute a "*set of necessary or sufficient conditions to define privacy*" (Solove 2007, 12) to current debates. The taxonomy is divided into the four dimensions listed above: Information Collection, Information Processing, Information Dissemination and Invasion.

Information Collection describes two problematic ways data can be collected

- *Surveillance*
- *Interrogation*

Information Processing

- *Aggregation*
- *Identification*
- *Insecurity*
- *Secondary Use*
- *Exclusion* (e.g. the inability to access and decide how personal data is being used)

Information Dissemination (i.e. the way information is transferred)

- *Breach of Confidentiality*
- *Appropriation*
- *Distortion*

Invasion can occur through

- *Intrusion or*
- *Decisional Interference*

2.2 Privacy encroaching technologies

Privacy encroachment has been highly influenced by and co-evolved with technological developments, which provide unprecedented possibilities regarding the amount and speed to collect (personal) data. Many activities and mechanism, described below, are engrained in our everyday lives when for instance information about our daily grocery shopping is collected.

The following paragraphs discuss **privacy encroaching technologies** by the example of **three types of activities**:

- tracking and monitoring,
- aggregation and analysis and
- large scale dissemination.

The taxonomy (see 2.1.1) discusses processes, such as information collection or processing in which privacy violations can occur. Yet, keeping Nissenbaum's contextual framework in mind, not all activities listed have negative consequences for individuals (or companies) per se. Personalisation of shopping trails for instance may have "*positive and negative externalities for consumers*" (Donovan et al. 2014, 71). There seems to be a balancing act between appreciating advice on future purchases and feeling tracked. Thus, it is highly dependent on the contextual use and processing of information as well as additional risks of vast amounts of data being in the hand of few people or companies can trigger severe privacy encroachments.

Tracking and Monitoring

Through the widespread use and rise of certain technologies processes of tracking information and monitoring (for instance 'shopping trails') have become alleviated. Certainly negative implications of tracking personal information are discrimination, exploitation and manipulation. An incident[1] made public in 2012 in the US (see grey text box below), has contributed to a public discussion and outcry regarding the tracking and monitoring of personal information.

***Target Store Using 'Predictive Analytics'* to identify customers' life changes**

The Guest Marketing Analytics and Statistics department of Target combine data to predict major life changes of their customers, in order to specifically target them with advertisements. The usage of predictive analysis became well-known entering public debates in 2012 when Target sent advertisement leaflets for maternity fashion and baby food to a presumably pregnant young woman. Her father found out about her pregnancy due to the ongoing leaflets in their mailbox.

Some examples of how information is gathered and can be used are listed here[2]:

- paying with **credit cards** provides evidence of a person's whereabouts
- **telephone bills** to extract payment provide information about a person's conversations
- **proxy cards** intended to provide secure access enable tracking of comings and goings
- **monitoring energy usage** patterns for utility companies (smart energy) indicate the presence, absence, and general activities of building
- **Event Data Recorder** (Unfalldatenspeicher) in cars for lower insurance rates
- **RFID-Tags (Radio Frequency Identification)** which are (also) used at marathons to track time

[1] 'How Companies Learn your Secrets?' published in the NY Times in February 2016 http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0

[2] Some more examples on how data is being gathered can be found in our report D2.2 'Urban Decision Making and Expert Integration Report'.

Aggregation and Analysis

The amount of information that is collected, while people are shopping, traveling, and use social networks allows to not only analyse current information but also possible future interests or moves of a person ('predictive analytics'). The **substantial increase of information stored and available online is also due to an increasing democratization of computerized information storage systems**. The democratization of computerized information storage systems has been enabled by decreasing costs of hard-, and software, alongside the amount of open software tools. This in turn allows an unprecedented number of organizations, including small NGOs or initiatives to collect, store and organize their information digitally. The open software DemocracyOS[3] for instance offers a tool for collaborative online decision-making which enables users to build their own proposals, debate and vote. The availability of tools such as DemocracyOS drives a proliferation of records of personal information and a growing number of social actors. This however brings about new challenges. Given the amount and detail of information, for instance people's political affiliations, or nutrition behaviour (if e.g. information of food coop participants is collected), new forms of detailed analysis of individuals are possible.

Large Scale Dissemination

> *"What you say on Twitter may be viewed all around the world instantly.* ***You are what you Tweet"***

Given the rise of new forms of communication triggered by social media (twitter, Facebook, online blogs), the speed and audience size has greatly increased in the past century. With new media the medium of storage and presentation makes a difference to what is revealed, even if the message remains unchanged. Additionally, the power of sharing tools in social media increases the speed and level at which reputations can be affected. For instance, the number of defamation actions brought over derogatory posts on social media has surged by more than 300 % (e.g. bad reviews on Trip Advisor)[4]. In reaction to increasing technologies encroaching privacy, so called privacy preserving technologies are being developed, including data anonymization (removing personally identifiable information – PII), encryption, anonymizing networks (TOR), or anonymous remailer.

2.3 Privacy concerns versus privacy actions

In general, knowledge about possible pitfalls and encroachment of personal privacy appears to be widespread, yet privacy concerns of individuals do not translate into consequent actions in most cases. On the one hand 56% of online users feel that they are protecting their personal information online, which in most cases they are not, or not fully (cf. TRUSTe[5] 2006). On the other hand people

[3] http://democracyos.org/

[4] http://www.independent.co.uk/news/uk/home-news/libel-cases-prompted-by-social-media-posts-rise-300-in-a-year-9805004.html

[5] TRUSTe is the leading privacy certification and seal program and market information group TNS; for more information, please see https://www.truste.com/

"*choose options that offer convenience, speedy passage, financial savings, connectivity, and safety rather than those that offer privacy*" (Nissenbaum 2010, 105).

For instance, when visiting a website, according to the 'E-Privacy Directive' in the European Union, users need to be informed about (1) which kind of data is collected and (2) how or if this data is stored and processed. The information about privacy policies of a website respective a company is commonly known as 'cookies'. However the number of users who actually read through and understand privacy statements on websites remains rather small as the study by TRUSTe points out. Only 20 percent of people claim to read privacy policies "*most of the time*" (TRUSTe and TNS 2006).

Additionally, companies who offer a 'frequent shoppers' card or store loyalty programs to their customers promise them convenience in terms of time saving (personalized advertisements) and special offers (customized to their taste). Shops and companies save latest purchases by individuals in order to optimize and personalize advertising as well as sell data to third parties (e.g. marketing companies). Services, such as a frequent shoppers card offers convenience to many people who receive personalized ads or recommendations which book or jacket to buy next. In addition to many services and activities (e.g. paying with credit card) offering convenience also non-transparency of how personal information is being processed and stored by companies, alongside not being aware of possible privacy encroachments are important to consider.

The term 'mass surveillance' is often raised in public discussions and knowledge of how personal data is being collected, stored and analyzed with the seemingly inherent risks to (partial) privacy loss, a prevalent arguments in privacy debates, as Solove discusses is best summarized as the ***'I have got nothing to hide'* argument**.

The 'I have got nothing to hide'

The 'I have got nothing to hide' argument might be an "*all-too-common refrain*" (Stone 2006 cit. in Solove 2007, 747); it is however not a trivial one considering that it "*reflects the sentiments of a wide percentage of the population*" (Solove 2007, 747). The argument of having something to hide delivers the notion that privacy is (only) about hiding something bad, i.e. that a person is "*engaged in illegal conduct*" (ibid. 751).

This is not unproblematic as this notion "*myopically views privacy as a form of concealment or secrecy*" (ibid. 746). Solove (2007) also discusses the prevalent argument of privacy (always) counterbalancing security, whereas the value of privacy is usually rather low as personal information is regarded as not sensitive when weighed against national security concerns. Solove concludes that the "*nothing to hide" argument speaks to some problems, but not to others. It represents a singular and narrow way of conceiving of privacy.*" (ibid. 772)

2.4Gender and Privacy Concerns

The ways people deal with and are aware of privacy matters can be influenced by a variety of factors, including gender. Generally, it is assumed that privacy concerns and ethics materialize in many shapes; depending on

(1) A person's e-literacy, age or gender, and/or
(2) different dealings with (online) privacy in personal and professional life, for instance using twitter for professional use only, i.e. not tweeting private photos (Taddicken 2013).

Privacy concerns are realized through many activities, including reading privacy notices on websites or using encryption while e-mailing. Research raises questions such as who tends to read privacy guidelines on websites (e.g. cookies) and who follows up on reading legislation, or privacy updates on social media sites and increasingly focuses on gender.

According to various researchers women tend to be more aware about online privacy and coherently read more often privacy statements. Thus, Sheehan (1999) found women to be more sensitive to online privacy than man "*and engage in noticeably different self-protective behaviours*" (Hoy & Milne 2010, 28). Milne and Culnan (2004) results suggest that women tend to read and trust privacy statements, such as cookies notices, on websites more often than men. Other research strands look at gender in computer and software ethics (see chapter 4.2 on ethical guidelines). However feminist ethics have gained rather little attention within the realms of computer ethics.

Questions raised in computer and software ethics are, for instance: to what degree does gender influence the recognition of "*unethical conduct in the use and development of information technology*" (Adam 2008, 593) and subsequently their decision-making? A feminist ethical approach towards computer and software ethics, as stated by Tong already in 1999 yet still accurate, should strive "*to create a gender-equal ethics, a moral theory that generates non-sexist moral principles, policies and practices*" (Tong 1999 cit. in Adam 2008, 590). Further, feminist ethics (1) form a substantial critique of traditional ethical theories, which are masculine in its conception and (2) challenge mainstream ethics (ibid. 593).

Adam critically looks at current research focusing on differences of women and men's ethical decision making with respect to computer ethics problems and found there to be an 'under theorizing of gender and ethics'. This, as she argues further, often reproduces "*stereotypical judgement about an expectation of men's more "laddish" behaviour against a "well-behaved" female stereotype where women are seen as guardians of society's morals*" (ibid. 601). Thus human behaviour concerning online privacy cannot be portrayed reduced to these gender differences. There are various key aspects influencing human (online) behaviour, such as the circumstances when to use Internet or where to have access to the web as well as culture, education and other social aspects (Bellman, Johnson & Kobrin 2004; Yao, Rice & Wallis 2007).

3 INTERNET RESEARCH AS ETHICAL PRACTICES

In this chapter we distinguish between three main groups in Internet research and the ways ethics shape their practices. The three groups are as follows:

(1) **ICT specialists and programmers** producing the tools that constitute the Internet itself as well as the tools that help to analyze information flows on the Internet.
(2) **Providers and maintainers of infrastructures** are the second group. They often have the power to enforce communication policies on their platforms or restrict access to their data (cf. limited ability of Twitter API to reach back in time).
(3) Lastly, **Internet users and Internet researchers** (while the distinction here is increasingly blurring) are the third group. They are both, content providers - through the data they generate consciously (e.g. Tweets) or in unintended ways (e.g. page hits or searches) - and beneficiaries of content analysis on the Internet.

By its very nature, Internet research raises multiple ethical and political challenges. Crosscutting humanities, social sciences and physical sciences, online researchers - more than ever - are aware that interpreting data is not a neutral activity but culturally determined. Hence, the task at hand for online researchers is to critically question and expose their own perspectives and resulting predilections or unintended side-effects (Johns, Chen, & Hall, 2004). Additionally, Internet users and the data traces they leave are not mere passive objects of research, but also beneficiaries of the products coming out of Internet research (e.g. better filtering of information during search). The very terms 'information society' and 'knowledge-based economy' indicate how important information has become as a means to drive innovation and growth. Rather than oversimplifying the challenges of ethical Internet research, we need to acknowledge the active / participatory role Internet users play in producing Internet research (Floridi, 2008).

Hence, since the 1990s the principle challenge remains to develop a set of guidelines that can claim to be founded in universally recognized norms while accounting for disciplinary, national and demographic differences (Buchanan & Ess, 2008).

3.1 Software ethics: Ethical values as design input

Information technology is about processing information, so that it is only logical that the way we design IT has a direct impact on privacy and other issues when IT is put to use. Yet, acknowledging the fact that 'artefacts have politics' (Langdon Winner) and that we can incorporate our moral values into technological designs has only happened with the 'design turn in applied ethics' in the 1990s (Van den Hoven, 2008). Before that, privacy issues were taken as given and their path-dependent analysis ignored the fact that we can change technologies in ways that prevent those issues from coming into being. Frontloading ethical a concern into a technical design is easier said than done, since many issues only become visible during large-scale use. Jenkins and McCauley (2006) gave the

example of a data algorithm oppressing the visualisation of ponds, thereby leading land-use planners to possibly devastating decisions, destroying ecologically valuable wetlands. To address the complexities of considering ethics as parts of designs Friedman at al. (2013) argue that three types of investigations are needed:

- **Conceptual**: conceptualisations deliver context, which in turn explicate affected stakeholders with their needs and values. Moreover, concepts help to specify the practical meaning of values which are often defined in rather abstract terms (e.g. trust, ownership, autonomy etc.)

- **Empirical**: empirical data help to overcome the shortcomings of conceptualisations. In order to understand to what degree a possible value is important to a given population, we need to gather the relevant data or observe the use of a technology in context.

- **Technical**: technologies transform values into concrete features. Here we can expect interesting outcomes, since most features have multiple outcomes (e.g. in collaborative group work systems, the desire for privacy of one individual affects the desire for awareness of group activity of another individual) (ibid.).

Other examples of value sensitive designs are the control of web browser cookies (incorporating the value of informed consent) and public deliberation systems for city-planners (incorporating the value of democratic decision making) (Friedman, Kahn Jr, Borning, & Huldtgren, 2013).

Underlying value sensitive designs is the question '**What is a value?**'. Coming back to the national and cultural differences in questions in ethics, Buchanan and Ess (2008) distinguish - very broadly - between the Anglo-American, utilitarian approach to ethics and the more deontological European approach. The former compares the potential benefits and costs of a given ethical choice, searching for the 'greatest good for the greatest number'. The latter presumes a more categorical right, which cannot be debated under no circumstances whatsoever. The European position would imply that research designs have to be tweaked until all relevant rights are respected and protected - even if this would mean to forego a research project if protecting participants' rights could not be guaranteed (ibid).

The conclusion in this debate is that ethical designs depend on cultural expectations, which have a normative importance and need to be accounted for - even or especially when Internet research affects people from potentially very diverse cultural backgrounds and consequently very diverse expectations around their privacy and transparency about the usage of their data. Buchanan and Ess (2008) compare privacy implementations in Germany and China. In Germany, data privacy protection is seen as an intrinsic value (i.e. a necessary condition for independent self-development and the execution of basic democratic rights), whereas in China data privacy is primarily an instrumental value (i.e. a precondition for a much desired growth in e-commerce developments).

More general guidelines in the field of computing are listed in section 4.2.1, including the ACM Code of Ethics, AITP Code of Ethics, and Software Engineer's Code of Ethics.

3.2 Platform ethics: Social media and big data forcing

> "*What I call big data's 'forcing function' is the result of the volume, velocity, and variety of big data's growth. Touching nearly all aspects of our lives, we are just now beginning to understand how it will influence our values and the meaning of words like identity, privacy, ownership, and reputation*." (Davis & Patterson 2012, viii)

When talking about Internet Service Providers (ISPs), or the providers of large or fast growing content platforms, ethical issues are mostly related to the appearance of illegal and immoral content. However, the liability of service providers has been regulated in the European Directive 2000/31/EC (Directive on Electronic Commerce). There it is established that as long as the provider does not have knowledge of the illegal information in question, it cannot be held accountable. However, providers are required to take actions as soon as issues are brought to their attention.

The associated duties and responsibilities are not trivial, if we consider that about 2.5 quintillion bytes of data are created daily - the equivalent of 57 billion 32 GB iPads (Desouza & Smith, 2014) or that *"some 90% of the world's data has been created in the last two years"*[6]. On the one side, the data flood can overwhelm administrators of comparatively smaller platforms if confronted with multiple 'takedown requests'. On the other side, organizations want to exploit this vast amount of data available today. 'Exploiting' means to base decisions on data rather than intuition. Data-driven decision making has been demonstrated to significantly impact productivity (Brynjolfsson, Hitt, & Kim, 2011). Big data is not only to make organisations more efficient, it is also used to better understand customers' needs or anticipate unintended consequences of market interventions. Big data bears both risks as well as opportunities; it certainly raises new challenges regarding security, legal compliance and privacy and has potential for innovative use for common good (e.g. movement patterns of people analysed to improve traffic and urban planning[7]) as well as deeper insights into health issues, e.g. geographic distribution of outbreak of disease.

Let's look at the value of information provided through platforms such as Twitter to illustrate the tradeoff between allowing for maximum diversity of expression on one side and monitoring tweets for potentially harmful content on the other side. The average number of tweets sent per *day is* 58 million and with one statement, including a photo or a link a lot more information is sent. David and Patterson (2012) dismantled the **anatomy of a tweet**. According, to their anatomy a tweet consists of:

[6] http://www.economist.com/node/21537967
[7] AT&T Labs in a city in New Jersey, USA, used big data, specifically the flow of mobile devices from cell tower to cell tower moving through cities (Davis & Patterson 2012, 36).

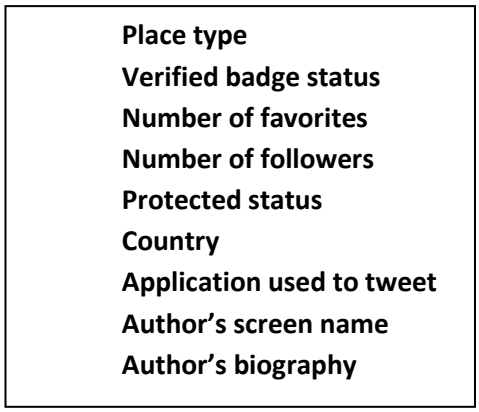

Figure 1: Anatomy of a tweet (Patterson 2012)

Davis and Patterson (2012) further raise concerns of **data ownership** by asking: "S*ince people buy their device does the data generated by the use of those devises belong to the individual device owners- or the company who owns and maintains the technological infrastructure that makes that usage possible*?"

Similarly, a critical view on the use of social media data by companies or for research purposes is raised by Boyd and Crawford (2012). They ask if data like tweets or likes on Facebook should be unrestrictedly and publicly available. This especially concerns data where identification can be made easily.

Closely related to platforms being accountable, is the preservation of due processes related to the use of big data analytics. This is no trivial challenge, as **audit trails for a judicial reviews**, hearings of the affected people (to get a balanced view) or prior notice to the point of decision making is often deemed to be unfeasible in a big data context (Crawford & Schultz 2014).

3.3 Participants ethics: Do we need more privacy self-management?

Current state of privacy provisions concern "*primarily of rights to notice, access, and consent regarding the collection, use, and disclosure of personal data. The goal of this bundle of rights is to provide people with control over their personal data, and through this control people can decide for themselves how to weigh the costs and benefits of the collection, use, or disclosure of their information*" (Solove 2012). Solove labels this approach "**privacy self-management**".

Underlying the call for more privacy self-management is the enormous growth of social media use through smartphones, which has created new waves of privacy concerns. Especially when considering the amount and detail of data gathered by smartphones, GPS tracking locations, use of social media and the increasing merge of private and professional use on mobile devices that are always 'with us'.

The development of mobile phones with Internet access reads back to 1996 with the Nokia 9000 Communicator, the first mobile phone with Internet connectivity was the Blackberry in the year 2001, also being the first email-enabled mobile phone. Since 2008 there are more mobile phones with Internet access than PCs in the world. In order to provide some figures, stressing the vast amount of use and data found within the realms of social media today, some figures[8] showing the rapid development of social media companies within the last 15 years are listed below:

- Pieces of content shared on **Facebook** each *month* **70 billion**
- Articles hosted by **Wikipedia 17 million**
- Pictures hosted by **Flickr 5 billion**
- **100 hours** of video are uploaded to **YouTube** every minute (i.e. 2000 y /year)

However, current big data[9] analytics takes place after aggregating many seemingly innocuous data, the question whether revelations in the future will be beneficial or harmful is far too complex and abstract for most people to make an informed consent and to self-manage their privacy (Solove, 2007). For example, in 2013 a privacy lab was able to identify 42% of anonymous contributors to a high profile DNA research study, publishing its data online - as participants included birth date, zip code and gender[10] (Sweeney, Abu, & Winn 2013).

Hence, Solove (2012) suggests two types of adjustments to privacy regulations:

- Rather than requiring people to micro-manage their privacy under high uncertainty or applying restrictive national laws**, an agency might decide** on new uses of aggregated, big data in the future;
- Regulations should take **a more substantial stance on privacy protection**, referencing basic privacy norms without falling into an overly paternalistic regime.

In a similar vein, Boyd and Crawford (2012) suggest to enhance the principle of privacy (protecting data owners) with the principle of accountability (holding accountable the users of big data). Accountability would have the benefit of encouraging users or researchers of big data to think about possible ramifications of their efforts in rigorous ways (ibid).

Another ethical dimension of Internet usage concerns the quality of information in general. Not only does it make a difference who can access information, but also for what purposes this information is accessed. Vedder (2008) raises the question of who is responsible for possible negative conse-quences if information on the Internet is misconstrued or misleading. This is particularly relevant for

[8] http://www.statisticbrain.com/social-networking-statistics/

[9] Big data is different to past data-analytics not only because of its size but also because of its relationality to other data (Boyd & Crawford, 2012). Another typical reference for big data are the three Vs: volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources).

[10] http://www.forbes.com/sites/adamtanner/2013/04/25/harvard-professor-re-identifies-anonymous-volunteers-in-dna-study/

information in the legal and health domain, where context is particularly important for the validity and applicability of information. Vedder suggests that information quality can be judged (a) by content criteria which requires certain upfront familiarity with a topic, and (b) by pedigree criteria which requires a widely acknowledged, credibility-conferring system.

Apart from increasing the recognisability of credibility-conferring systems (e.g. through certifications), Vedder (ibid.) argues that increasing the background knowledge and expertise of Internet users is paramount, so that self-management is not limited to the regulation of data privacy but also data quality.

4 SOCIAL, ETHICAL AND LEGAL GUIDELINES

The previous chapters (1 – 3) have discussed different privacy concepts, concerns and differences in privacy actions among users as well as new waves of privacy concerns due to the unprecedented amount of data, including big data, social media data and open data found today.

The following chapter will discuss the most important privacy guidelines on three levels: on social, ethical and legal levels. Firstly, Nissenbaum's (2004) contextual framework is taken as basis for a social guideline, ensuring valid data interpretations in research and policy decisions. Secondly, common ethical guidelines are briefly described as well as the importance of user-centred design (e.g. when an interface is planned) are stressed. Thirdly, legal guidelines are described, starting from the international framework developed by the OECD to specific EU Directives on (1) Data Protection, (2) E-Privacy and (3) Data Retention as well as their national implementations in the relevant case study countries: Austria, UK, Sweden, and Denmark. In order to better grasp the local landscapes of relevant organizations, including NGOs, dealing with privacy issues, brief lists of organizations in each of the four case study countries are provided.

4.1 Social Guidelines: Contextual integrity models

The difficult question of how privacy can or should be treated in the context of rapid growth of Internet technologies, social media and mobile phones leading to an unprecedented amount of (online) data, is not an easy endeavour.

"*In the past few decades, we have seen a radical intensification in the social practices of gathering, storing, manipulating, and sharing information about people (henceforth "personal information")*" (Barth 2008, 1).

There are hardly any one-size fits all answers or solution available, considering the difficulty to conceptualize notions of privacy and the interweaving into our everyday lives. The contextual integrity framework does not aspire to provide a one-size fits all concept but it provides a framework for better understanding current privacy expectations and privacy practices (ibid. 2).

Many approaches treat information as a binary concept where information is either private or not, or describe different levels of privacy (Krupa & Vercouter 2010). The contextual integrity model "*defines privacy in a socially relevant way.* [on an abstract level] ... *all information is regarded as evenly sensitive/insensitive*" (ibid. 151f.). Then **the situation and context determines whether privacy is being violated or not**. In addition the purpose of collecting and analyzing data is essential; data can be collected in various contexts and for different purposes. Some examples for four different areas and reasons behind data collection are listed below:

- *Health* (blood-pressure monitors; EEGs; or monitoring outbreaks of infectious diseases)
- *Marketing* (analyzing consumption patterns through e.g. monitoring online transactions)
- *Research* (analysis of mobility patterns by collecting personal data and GPS tracking)
- *Local Police* (CCCTV video surveillance of public parks)

The above discussion about social guidelines stresses the importance to consider the contexts of (1) data collection and (2) data usage in research. In addition also the type of data and its source, for instance whether it is retrieved from social media sites (tweets), open data portals provided by governments, or non-public data from local authorities (e.g. police), needs to be reflected upon. In order to classify research projects according to their conflict potential, a matrix has been suggested including two dimensions: (a) the use of public vs. private data and (b) the use of sensitive vs. non-sensitive data. Buchanan und Ess (2008) suggest that data from the private and sensitive quadrant should be considered off-limits and all other data could be used with the appropriate guidelines and policies.

Another implication in visualizing or providing data in their raw format is the danger that out of context data are more prone to endorsing a person's image or a stereotype of a neighborhood[11]. Thus, awareness and caution of possible misuses of data needs to be considered at an early stage in the research process.

4.2 Ethical Guidelines: Codes of conduct

"*Big-data technology has no value framework. Individuals and corporations, however, do have value systems.*" (Davis & Patterson 2012, 8)

The rapid growth of Big Data raises new questions and challenges regarding privacy and ethical standards. As Davis & Patterson (2012) point out big data not only concern questions and concerns about personal privacy but "*it generates new questions about personal identity, notably who owns our personal data and how does the increased presence and availability of more data influence our reputation*". Thus ethics of big data not only entail questions relating to the collection, storing and processing of data (i.e. legal regulations and codes) in a company or organization but also more

[11] See for instance Goss, J. (1995) "We Know Who You Are And We Know Where You Live": The Instrumental Rationality of Geodemographic Systems. Economic Geography, 71(2).

general (and widely shared) concerns about personal identity and reputation. Speaking in the words of Brad Peters[12] "*Big Data changes the social contract*".

4.2.1 Ethical guidelines for individuals and organizations

In an attempt to disentangle single aspects of big data ethics for both individuals and organizations, Davis & Patterson (2012) stress four main dimensions:

- **Identity**
  raises concerns, such as "*Big Data provides others the ability to quite easily summarize, aggregate, or correlate various aspects of our identity – without our participation or agreement*." (ibid. 16)
- **Privacy**
  "*Who should control access to data about you*?"
  The above question emphasizes concerns about the degree of control individuals have over their personal information? Many people can access huge amounts of information about individuals given that the Internet, anonymity when operating on the Internet has changed drastically over the past years. (ibid. 18)
- **Ownership**
  "*What does it mean to own data about ourselves*?"
  Do we own data about ourselves differently in the offline and online world? Given that data markets grow, questions of who actually owns which data (and at which point in the data trail) are crucial.
- **Reputation**
  "*How can we determine what is trust-worthy*?"
  Due to the growth in the use of social media, the audiences who can (potentially) form an opinion about a tweet and spread it via online channels have grown enormously.

Generally one can argue that ethics of big data is about discussing how our values influence our actions. Ethical practice as argued is an outcome of ethical inquiry; ethical inquiry is understood as an exploration of values. With their framework, David & Patterson (2012) seek a set of common values of ethics of big data to reduce value conflicts and ease opportunities for collaborative innovation. These 'opportunities' include examples such as adding a new product feature, designing new products or services, as well as a new combination of data.

As Davis and Patterson (2012) argue agreements how data is used should be made explicit and easily understood and accessible (ibid. 37). Additionally it is not only a question of providing information on how data is being collected, used, or processed, but also how the information is provided in terms of technological design and readability. Of course it is not only a question of design whether users actually pay attention and read the privacy information (e.g. notice about using cookies on websites)

[12] Peters, Brad, "The Age of Big Data", *Forbes online,* December 2012.
http://www.forbes.com/sites/bradpeters/2012/07/12/the-age-of-big-data/

or not, but how information is designed should not be underestimated. Ethical decision points can ease and facilitate ethical inquiries and dialogues in organization.

4.2.2 Ethical guidelines in Software Engineering

Given that software engineers design and program tools, including the mechanisms of writing a code to collect data and personal information of its users, ethics play an important role within the profession of software engineering. Within software engineering three codes of ethics are most prevalent: ACM Code of Ethics, AITP Code of Ethics, and Software Engineer's Code of Ethics. Each code of ethics is briefly described below.

ACM Code of Ethics

The ACM Code dates back to 1992, with last revisions made in 2004. Generally, the ACM Code[13] is divided into four sections: General Ethical Considerations, Specific IT Professional Responsibilities, Leadership Responsibilities, and principles for complying with the code.

Some of the ACM Code's "General Moral Imperatives" include:

- Contribute to society and human well-being;
- Avoid harm to others (includes substantive duty of assessing social consequences of systems);
- Honor property rights including copyright and patent;
- Give proper credit for intellectual property;
- Respect the privacy of others (includes substantive responsibility for data integrity);
- Honor confidentiality.

More specifically, the ACM Code of Ethics has eight guiding principles in regards to behaviors and ethical decision-making by software engineers, including practitioners, managers, supervisors, policy makers, as well as students. The eight principles are as follow:

1. PUBLIC - Software engineers shall act consistently with the public interest.
2. CLIENT AND EMPLOYER - Software engineers shall act in a manner that is in the best interests of their client and employer consistent with the public interest.
3. PRODUCT - Software engineers shall ensure that their products and related modifications meet the highest professional standards possible.
4. JUDGMENT - Software engineers shall maintain integrity and independence in their professional judgment.
5. MANAGEMENT - Software engineering managers and leaders shall subscribe to and promote an ethical approach to the management of software development and maintenance.
6. PROFESSION - Software engineers shall advance the integrity and reputation of the profession consistent with the public interest.
7. COLLEAGUES - Software engineers shall be fair to and supportive of their colleagues.

[13] For more information, including the entire ACM code of ethics, please see: https://www.acm.org/about/se-code

8. SELF - Software engineers shall participate in lifelong learning regarding the practice of their profession and shall promote an ethical approach to the practice of the profession.

AITP Code of Ethics

The Association of Information Technology Professionals (AITP) Code of Ethics[14] is addressed to IT professionals generally. AITP is the leading worldwide society of information technology business professionals and the community of knowledge for the current and next generation of leaders.
The code of ethics is formulated for four stakeholders: management, fellow IT professionals, society, and employers. The AITP guiding principles can be summarized as four topics:

- Integrity
  "We value professionalism and uphold the AITP Code of Ethics and Code of Conduct."
- Respect
  "We build an inclusive environment through mentoring, delivering on commitments, working together with trust, and enjoying the camaraderie of each other."
- Innovation
  "We learn, share insights, and encourage our members to make a difference today and for the future."
- Service
  "We keep current in technology, business, and academia. We contribute to the Association, IT profession, and society utilizing leadership, appropriate solutions, and sound processes."

Software Engineer's Code of Ethics

The Software Engineering Code of Ethics and Professional Practice (SEC) dates back to 1998, and consists of eight principles that express ethically responsible relationships related to software development.

- Public
- Act consistently with the public interest.
- Client and employer
- Act in the best interests of their client and employer, consistent with the public interest.
  Product
- Ensure products meet highest professional standards.
- Judgment
- Maintain integrity and independence of professional judgment.
- Management
- SE managers and leaders shall manage software development and maintenance ethically.
- Profession
- Advance integrity and reputation of profession consistent with 1.

[14] http://www.aitp.org/group/3203

- Colleagues
- Support colleagues.
- Self
- Participate in lifelong learning and promote an ethical approach to profession

It contains very little specific to software development. Most of the requirements are formulated at the level of an IT development project, which would apply across the board to almost all IT professionals. This Code also includes a number of prioritizations to help settle conflicts. For example, Number 2, the best interests of client and employer, as well as Number 6, the best interests of the profession, are explicitly superseded by Number 1, the public interest.

4.3 Legal Guidelines: Principles

This chapter discussed legal guidelines in regards to privacy on both, international (OECD and European Union) and national levels (e.g. the four case study countries). General privacy principles provided by the OECD and the European Union are discussed. These principles set a general framework under which specific EU Directives on data protection, processing or storing are framed.

Legal guidelines determining control and access over (personal) data can be found in almost all countries, despite variation in implementation[15], some general underlying principles in terms of processing of personal data for instance refer to the specification of the purpose, limited use of personal data, individuals to be notified and allowed to correct inaccuracies. However, up until recently, the number of countries who have privacy protection laws has merely been an estimate. In 2012 Greenleaf analysed how many countries worldwide actually have data privacy laws and how data laws developed over time. His results (from 2012) show that at least 89 countries have implemented data privacy laws; more than half of these countries are located within Europe. Further it was shown that growth of data protection laws across the global is accelerating and not expanding linearly. Looking at the implementation of laws by decade the 2000s are the strongest decade with 35 countries, followed by 2010s with 12 new countries implementing laws (Greenleaf 2010). Given the continuous growth of countries, Greenleaf predicts another 50 new laws in the current decade, which would result in 127 countries on the 'data protection implementation list'. Most of these new laws will come from outside of Europe.

International agreements and guidelines have impacted upon national and sub-national law for more than thirty years since the first privacy draft by OECD in 1980 (which was revised in 2013). All European Member states are required to have data privacy laws, which implement the EU Directives. Also within the European Economic Area (EEA) all member countries have data privacy laws. Countries or jurisdictions outside of the EEA can request a decision by EU that their laws showcase an

[15] DLA Piper 'Data Protection Laws of the World' provides a handbook listing data protection laws of 77 countries, alongside a virtual interactive heat map indicating the extent of law enforcement (limited to heavy). It further offers direct comparison of countries in regards to their laws on e.g. concrete laws, online privacy. http://dlapiperdataprotection.com/#handbook/world-map-section

"*adequate level of protection of privacy, to enable free flow of personal data from EU members states to organisations in those countries*" – so far the EU has made this decisions for nine jurisdictions (Andorra, Argentina, Canada, Switzerland, Faroe Islands, Guernsey, Israel, Isle of Man, Jersey).

4.3.1 Fair Information Practice Principles (1973)

The Fair Information Practice Principles developed in 1973 (by US Health, Education, Welfare Advisory Committee on Automated Data Systems) was one of the first privacy framework and "*became the dominant U.S. approach to information privacy protection for the next three decades.*" (Westin 2003, 436). The five guiding principles are:

- Notice: notifying about data collection
- Choice/Consent: usually not explicit (e.g. access of third parties allowed)
- Access/Participation: view, verify or correct your data
- Security: encrypt data, limit access to data within the organisation
- Enforcement: (a) self-regulation by collectors (b) suing perpetrators and (c) government enforcement

4.3.2 OECD Privacy Guidelines

The OECD Privacy Guidelines on the Protection of Privacy and Transborder Flows of Personal Data are the most widely used privacy framework internationally. These guidelines are closely related to data protection legislations by EU member states (see for instance EU Directive 95/46/EC Data Protection Directive). Privacy frameworks may be used as tools to help us think about and frame discussions about privacy, and understand privacy requirements.

"*OECD Member countries considered it necessary to develop Guidelines which would help to harmonise national privacy legislation and, while upholding such human rights, would at the same time prevent interruptions in international flows of data. They represent a consensus on basic principles which can be built into existing national legislation, or serve as a basis for legislation in those countries which do not yet have it*.[16]

The first privacy guidelines by the OECD were developed in 1980, with first revision in 2013. Given the growth in amount of personal data that is collected, used and stored (digitally and non-digitally) today, and variety of actors (e.g. number of social media users) the OECD privacy guidelines from 2013 refer to 8 guiding principles:

- *Collection Limitation* Principle: Collection should be limited and consented
- *Data Quality Principle*: Personal data should be relevant to the purposes, accurate, complete and kept up-to-date

16

http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

- *Purpose Specification* Principle: Purposes should be specified not later than at the time of data collection
- *Use Limitation Principle*: Personal data should not be disclosed, made available or otherwise used for purposes other than those specified except:
  a) with consent or b) by the authority of law.
- *Security Safeguards Principle*: Personal data should be protected
- *Openness Principle*: A general policy of openness about developments, practices and policies with respect to personal data
- *Individual Participation* Principle: An individual should know whether there are data relating to him/her, know which data there are, understand why access might be denied, and be able to correct or delete these data
- *Accountability Principle*: Data controller should be accountable for effectively implementing these principles

4.4 Legal Guidelines: Directives at a European level

Three EU Directives are explained in greater detail in order to look at their specific implementation in the four case study countries (Austria, United Kingdom, Sweden, and Denmark).

4.4.1 Article 8 of the European Convention on Human Rights

The European Convention on Human Rights entered into force in 1953 (it was adopted in 1950), the ratification of the convention is a prerequisite to enter the European Union. In 2010 last amendments to convention have been made. The European Court of Human Rights oversees the implementation of the European Convention by member states. Article 8, the *'Right to respect for private and family life'* states:

1) Everyone has the right to respect for his private and family life, his home and his correspondence.

2) There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others."[17]

[17] http://www.echr.coe.int/Documents/Convention_ENG.pdf

4.4.2 Directive 95/46/EC (Data Protection Directive)

The Data Protection Directive[18] was adopted in 1995 and refers to the protection of individuals with regard to the processing of personal data and on the free movement of such data. In accordance with this Directive, Member States shall protect the *fundamental rights and freedoms* of natural persons, and in particular their right to privacy with respect to the processing of personal data. European member states had to transpose the Directive into internal law by the end of 1998, however the Directive is not legally binding. Hence each state has enacted their own data protection legislation. However the Directive is planned to be fully adopted by member states in 2015, which would clear off all national legislation and national differences in implementation.

'Personal Data' defined by the Directive refers to

"any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."

Examples of personal data include address, credit card number, bank statements or criminal records.

'Processing of Personal Data' defined as

"any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as *collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination* or otherwise making available, alignment or combination, blocking, erasure or destruction."

However, in some cases data processing is legitimated given certain criteria (listed under Article 7 'Criteria for making data processing legitimate'), i.e. personal data shouldn't be processed, except when certain conditions are met which fall under three conditions: *transparency, legitimate purpose, proportionality.* Member States shall provide that personal data may be processed only if:

- the data subject has unambiguously given his consent; or
- processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract; or
- processing is necessary for compliance with a legal obligation to which the controller is subject; or
- processing is necessary for the purposes of the legitimate interests pursued by the

[18] Full title: Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 Article 1 - Object of the Directive

- controller or by the third party or parties to whom the data are disclosed, **except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject** which require protection under Article 1.

Case Example
'ELGA Gesundheitsakte' (Austria)

'ELGA Healthrecord' connects personal health data (e.g. doctors' visits, medication), which can be accessed by doctors to provide information about last doctor visits, current medication, dismissals etc. The Austrian medical chamber is highly critical of ELGA and has been raising concerns regarding the patients' loss of privacy. Currently ELGA will be implemented in 3 (out of 9) Austrian Federal States (including Vienna). This will include all public hospitals, pharmacies, care facilities and doctor's offices that share and have access to this information starting in the end of 2015. In contrast to Denmark (who has successfully implemented a similar record in 2003) Austria chose an 'opt-out' approach where people have to sign a form in order to not be part of ELGA. Additionally the transparency of the process right from the beginning to its implementation may have contributed to less controversial discussions in Denmark.

4.4.3 National Implementation of Data Protection Directive

Austria

The Data Protection Directive was implemented in the Austrian Federal Act concerning the Protection of Personal Data 'Datenschutzgesetz'[19] (DSG) in 2000. All Austrian federal states (nine states) have adopted data protection laws to implement the Directive. In Austria, recent discussion concerning the 'ELGA Healthrecord'[20] (see case example above) has raised concerns by many citizens regarding personal data protection.

Alterations to the current federal act due to member state wide binding implementation of directive would in the case of Austria for instance be the following[21]:

- Companies in member states with more than 250 employees are obliged to have a *data protection officer* (also obligatory for public authorities despite its size)
- In case new IT systems are installed protocols for *Data protection impact assessment*, e.g. what would happen if the IT system is leaked? need to be followed

The European Union has been pushing for (more) national certifications and seals of quality (Gütesiegel) in member states to raise the level of data protection, e.g. IT product will be more

[19] https://www.dsb.gv.at/DocView.axd?CobId=41936 Wiener Datenschutzgesetz (Wr. DSG), LGBl. Nr. 125/2001
[20] See http://www.elga.gv.at/index.php?id=faq and press articles
http://diepresse.com/home/wirtschaft/recht/1588903/ELGA_Bei-Datenleck-haften-Aerzte
[21]http://futurezone.at/archiv/neue-eu-datenschutz-verordnung-inwieweit-hilft-iso-27001-zertifizierung/24.595.175

transparent and can be assessed faster when seal of quality. A European Privacy Seal (EuroPriSe) started by Unabhängigen Landeszentrum für Datenschutz in Schleswig-Holstein (funded under eTEN-Program) has been in place since 2007. Receiving the European Privacy Seal[22] requires a two-step procedure to certify IT products and IT based services.

UK

The Data Protection Directive was transposed into UK law via the Data Protection Act 1998[23] (secondary legislation passed in 2000). The Data Protection Act regulates how personal information is used by organisations, businesses, researchers and government departments.

Sweden

The Personal Data Act was added to the Swedish Code of Statues in 1998 and is based on the EC Directive 95/46/EC on data protection. The main purpose of the Personal Data Act is to protect Swedish citizen's personal integrity when their personal data is being processed. The law is applicable in the public and private sector, and companies, government agencies, and other associations control the implementation of the law within their own organization (Datainspektionen, 2015, 1).

In 2007 the Personal Data Act was revised at the request of the Swedish government. The purpose of the revision was to investigate whether it was possible to modify the Personal Data Act to enact laws on the misuse of personal data, instead of handling of personal data, despite the current EU Directive. The revision concluded in a change of the Personal Data Act to make the handling of personal data easier, by making a clear difference between personal data handled in a structured form and data handled in an unstructured form (Datainspektionen, 2015, 1). Personal data handled in an unstructured form, such as in word processing systems, continuous text, or occasional audio and video recording, is therefore exempted from most of the laws in the Personal Data Act (as long as the personal integrity is not violated). Personal data handled in a structured form in traditional data files, databases and document handling systems is, however, still subject to all of the laws in the Personal Data Act (Regeringen, 2014).

Denmark

The Danish implementation of the data protection directive is known as "persondataloven", and is called The Act on Processing of Personal Data. The official version translated for the Danish Data Protection Agency is published in "Lovtidende" (Official Journal) on 2 June 2000 and only the Danish version of the text has legal validity. The current version has been amended until December 2012.

The act, in relation to the electronic data, stipulates the rules for the automatic and non-automatic systematic processing and transmission of personal data. The act does not cover social media data specifically but describes rules for handling media data. With respect to the project, the act describes that text, images, sound and video that has been already published do not apply to the act.

[22] https://www.european-privacy-seal.eu/EPS-en/Home
[23] http://www.legislation.gov.uk/ukpga/1998/29/contents

Concerning the project there should be no relationship in the database that will lead to the establishment of personal profile their by obtaining personal data from stored data.

"Furthermore, this Act shall not apply to information databases which exclusively include already published texts, images and sound programs which are covered by paragraph 3 of section 1 of the Act on media responsibility, or parts hereof, provided that the data are stored in the database in the original version published. However, sections 41, 42 and 69 of the Act shall apply."

Where section 41 describes the rules of controller and the third parties working with the personal data and the securing of such data, 42 and 69 describes the data controllers compensation to any damage cause by violation of the provision of the act in relation to processing of data, unless such damage cannot be adverted through diligence and care in the processing. The act applies to processing of personal data such as CPR number, addresses, and transaction information and where the management, processing and securing of storing social media data in relation to identifying persons or individuals, unless published by the persons or individuals. The act was stated to come into operation on 1 January 2014.

4.4.4 Directive 2002/58/EC (The E-Privacy Directive)

Directive 2002/58/EC on Privacy and Electronic Communications concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) and free movement of data, communication equipment and services. The E-Privacy Directive complements the 'Data Protection Directive' as it specifically applies to legal persons (e.g. providers of websites, not only individuals). The E-Privacy Directive has been implemented in member states since 2009 directive. The Directive often referred to as 'Cookie Law' deals with the confidentiality of information, treatment of traffic data, spam and cookies.

The first general obligation in the Directive is to provide *security of services.* Service providers have to inform the subscribers in case a particular risk (e.g. virus) may harm the protection of their personal data. The second general obligation is for the *confidentiality of information to be maintained*. Referring to data retention, providers of services for instance are obliged to delete traffic data when it is no longer needed unless users have given consent or conditions under Article 15[24] are fulfilled.

The Directive also regulates the passing on of e-mail addresses to third parties (e.g. marketing companies), unsolicited e-mails (spam) unless recipients have agreed to receive such emails prior ('opt-in regime') (e.g. newsletter e-mail lists).

The Directive also requires websites to get consent from visitors to store or retrieve any information on a computer, smartphone or tablet with 'cookies'. Making users accept the cookies before further use of a website, the increased protection of online privacy by making consumers aware of how information about them is collected and used online, and give them a choice to allow it or not.

[24] see Article 15 'Application of certain provisions of Directive 95/46/EC' http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:en:HTML

Case Example
'Cookie Law'

"*The cookie regulations, intended to be uniform in Europe, ended up as an inconsistent mess*." (Schneider 2014)

Despite the initial idea to establish consistent cookie regulations in all EU member states, it has been implemented differently in many countries. Further, the cookie law has received major criticism from a UK company, 'Siltide'. Silktide established a 'no-cookie-law' website (www.nocookielaw.com) raising concern about the great number of pop-ups the new cookie law requires and the fact that still many website ignore it alongside the fact that many people usually don't read the cookie notifications anyway.

Reference: [25]

4.4.5 National Implementation of E-Privacy Directive

Austria

The E-Privacy Directive has been enforced in Austrian law through Austrian Telecommunications Act in 2011[26]. The following three guidelines have been implemented:

Information Requirement:

Inform the user on the types of data processed (including cookies), the legal basis for and the purpose of processing the data, and the duration of storage.

Opt-In:

Opt-in consent for processing the data (including cookies) required, except where for technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as strictly necessary in order for the provider of an information society service requested by the subscriber or user to provide the service.

Implied Consent:

Legislative notes to the amendments to the Telecommunications Act suggest that consent may also be inferred from browser or other application settings.

UK

The E-Privacy Directive was transposed into UK law via the Privacy and Electronic Communications Regulations[27] (2003,amended in 2011). The Regulations cover the transmission of automated recorded messages (phone, email or SMS) for direct marketing. The Regulations have been criticised by some website owners for being "unworkable" in practice.

[25] http://www.osborneclarke.com/connected-insights/publications/european-commission-germany-has-implemented-cookie-directive-really/
[26] http://www.fieldfisher.com/pdf/cookie-consent-tracking-table.pdf
[27] http://www.legislation.gov.uk/uksi/2003/2426/made

Sweden

The Electronic Communications Act replaced the Telecommunications Act and the Radio communications Act in 2003 and comprises all the laws on electronic communications networks and electronic communications services (PTS, 2015:1). In terms of privacy the Electronic Communications Act is subject to the Personal Data Act, meaning that personal data processed with electronic communications networks and electronic communications services must relate to the laws in the Personal Data Act unless stated otherwise (SFS, 2003).

Denmark

The Danish implementation of the E-privacy directive is done through Executive Order no 1148 of 9 December 2011 - commonly known as "cookie-bekendtgørelsen", and describes the guidelines and regulation for the use of cookies. The Danish implementation is described in the "Guidelines on Executive Order on Information and Consent Required in Case of Storing and Accessing Information in End-User Terminal Equipment ("Cookie Order")".

4.4.6 Directive 2006/24/EC (The Data Retention Directive)

The Data Retention Directive was put in place in 2006. Especially after two occurrences - the attacks in Madrid in 2004 and London in 2005 - the EU was eager to harmonize crime investigation and prosecution among the EU member countries. The Data Retention Directive requires countries to retain certain information over a period between six months to two years.

The Data Retention Directive thus

"*aims to harmonize Member States' provisions concerning […] the retention of certain data which are generated or processed by them, in order to ensure that the data are available for the purpose of the investigation, detection and prosecution of serious crime, as defined by each Member State in its national law*."

The Directive applies

"*to traffic and location data on both legal entities and natural persons and to the related data necessary to identify the subscriber or registered user. It shall not apply to the content of electronic communications* […]"

Initially all members states were required to retain for between 6 and 24 months all data necessary to trace and identify: the source, the destination, the date, time and duration and the type of communication, as well as the communication device and the location of mobile communication equipment.

From the very beginning the directive has been highly controversially discussed in many member states and rejected as unconstitutional by several member states. In 2008 it was challenged at the EU Court of Justice for the first time by the Digital Rights Ireland[28].

On 8 April 2014, the Court of Justice of the European Union declared the Data Retention Directive invalid, as it didn't "*meet the principle of proportionality and should have provided more safeguards to protect the fundamental rights to respect for private life and to the protection of personal data*."[29] However some EU member states have kept legal regulations on national level, e.g. UK "Data retention and investigatory power act".

Case Example

NGO Digital Rights Ireland

There have been numerous legal challenges against the EU's Data Retention Directive at both national and EU level. In 2006, the first legal challenge to the EU Court of Justice raised by NGO Digital Rights Ireland (supported by Slovakia) was on the grounds that the Data Retention Directive had the wrong legal basis. NGO Digital Rights Ireland claimed that the correct legal basis for data retention resided "*in the provisions of the EU Treaty concerning police and judicial cooperation in criminal matters*," rather than those on the internal market.

4.4.7 National Implementation of Data Retention Directive

Austria

After an on-going trial (starting in 2009) concerning 'Data retention' a law was implemented in April 2012. However discussions and criticism did not stop - the imitative 'AK Vorrat'[30] collected signatures to bring a claim to the National Constitutional Court (VfGH). In July 2013 the VfGH hold a first meeting regarding the claim with the final decision that data retention is unconstitutional on 27.06.2014 by European Court of Law. Despite the Austrian Government defending data retention in front of VfGH, the collection and storage of data is declared as unconstitutional by national court of constitutional law[31]. After recent terrorist attacks new voices in favour of data retention have been raised.

UK

The Data Retention Directive was transposed into UK law via the Data Retention Regulations 2009[32]. The 2010-2015 Government debated the adoption of a Communications Data Bill (Draft), which

[28] https://wiki.openrightsgroup.org/wiki/Data_Retention_Directive#Sweden

[29] http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/police-cooperation/data-retention/index_en.htm

[30] https://www.akvorrat.at/

[31] http://derstandard.at/2000002350932/Verfassungsgerichtshof-kippt-Vorratsdatenspeicherung

[32] http://www.legislation.gov.uk/uksi/2009/859/contents/made

would require ISPs and mobile phone companies to retain records of (but not content of) their users' Internet browsing activity (including social media), email correspondence, voice calls, Internet gaming, and mobile phone messaging services and store the records for 12 months. The 2015 Conservatives majority government has announced its intention to reintroduce this Bill.

Sweden

The EC Directive 2006/24/EC on Data Retention encountered heavy criticism in Sweden when it was implemented by the EU in 2006. Originally Sweden was supposed to implement the directive on the 15 of March in 2009 and an investigation on how to implement it in accordance to the Swedish Code of Statues was initiated at the request of the government in 2007 (SOU, 2007). The investigation concluded that the collection and retention of data should be done by the service provider and stored for 1 year, and after that destroyed. But the proposal faced heavy criticism on the violation of personal integrity and the political parties which constituted the government could not agree on how to design the legislation, which made it impossible to pass it (SVD, 2009, 1).

In March 2009 the government agreed on how to implement the directive and sought to impose the minimum requirements established by the EU, keeping data for only six months and not for 1 year as the initial investigation proposed. But due to the controversy and difficulties surrounding the directive the Swedish government decided to delay the legislation proposal in order to investigate it further (SVD, 2009, 2). In March 2012 the legislation was finally passed and at the beginning of May the same year service providers had to collect and store data for six months.

When the Court of Justice of the European Union declared the Data Retention Directive invalid in April 2014, the Swedish Post and Telecom Authority (PTS) announced that they would not take any legal actions against service providers who decided to stop the collection and storing of data due to the unclear situation (PTS, 2014, 1). The Swedish government therefore appointed an investigation with the purpose to clarify if the current law on data retention violated the personal integrity or any other fundamental rights.

In June 2014 the investigation concluded that the existing law on data retention did not violate any fundamental rights and is therefore valid, meaning that the law should still be applied (Justitiedepartementet, 2014). Shortly after PTS announced that they once again would take legal actions against service providers who refused to collect and store data for six months (PTS, 2014, 3).

Currently the existing legislation on data retention applies. Some service providers in Sweden, such as Banhof, decided not to follow the law and were therefore fined by the PTS (Banhof, 2014).

Denmark

The Danish implementation of the data retention directive is done through Executive Order no 988 of 28 September 2006 commonly known as "Logningsbekendtgørelsen". The directive is described in the "Order on providers of electronic communications and electronic communications services registration and storage of information about telecommunications traffic (Retention Order 1)". This order directly refers to electronic communication as it has more to do with the demands for

telecommunication and Internet service providers to register information on their customers' use of services.

4.5 Organizations in case study countries

The following chapter briefly describes the main institutions (from national data protection agency to NGOs) concerned with privacy issues in the case study countries. The collection of organizations will be helpful in case questions regarding handling of big data occurs or simple act as sources of information on a national level.

Europe – European Digital Rights (EDRi)

An important umbrella organisation including several organisations and initiatives in EU member states is the European Digital Rights (EDRi)[33]. EDRi is an international non-profit organisation and was founded in 2002 in Brussels. Currently it counts 33 members, comprising privacy organisations and civil rights organizations from 19 EU member states. The president of EDRi is also a member of the Austrian organization ‚VIBE' (see 4.5.1). As stated on their website *„Statutory membership is restricted to not-for-profit, non-governmental organisations whose goals include the defence and promotion of civil rights in the field of information- and communication technology*." [34]

Within the ‚UrbanData2Decide' case study sites member organizations of EDRi include DFRI (Sweden), GreenNet and Open Rights Group (UK) and The IT-Political Association of Denmark (IT-Pol) (Denmark), 'VIBE' (Austria).

4.5.1 Vienna, Austria

National Data Protection Authority (Datenschutzbehörde, since 01/2014 – formerly established as 'Datenschutzkomission')
The Austrian data protection authority (in German *Datenschutzbehörde*) is a governmental authority charged with data protection. The data protection authority is the Austrian supervisory authority for data protection, the equivalent of a national data protection commissioner in other countries. The data protection authority has replaced the data protection commission, which held this position until 31 December 2013. Austria was one of the first European countries with a national data protection authority – it was founded with the first data protection law and held their first meeting was held in 1979[35].

Data Protection Council ('Datenschutzrat', a monitoring body, located at the office of the federal chancellor)
The Data Protection Council functions as an advisory to national and federal governments in Austria in questions relating to data protection and privacy issues. The council's main priority is to preserve

[33] https://edri.org/
[34] https://edri.org/about/
[35] https://www.dsb.gv.at/site/6179/default.aspx

data protection in Austria and to (critically) question new regulations or changes to existing laws. The Council advises the government concerning the implementation of new legislations if the (in-)directly affect national data protection. The council is composed of representatives from political parties, Austrian Federal Chamber of Labor, The Austrian Federal Economic Chamber, Federal States, Community confederation (Gemeindebund), municipal confederation (Städtebund) and a representative from the confederacy nominated by the federal chancellor. Working for the Council is voluntary (without monetary compensation). The management of the council resides with the office of the federal chancellor.

MA26 – Vienna Municipality 'Data Protection, Right to Information & e-Government (in place since 2007)

The municipal department 'MA26' is entrusted with the data processing registry (Datenverarbeitungsregister[36]) of the municipality of Vienna. Everyone interested in registered usage of personal data by the municipality can see their files at the 'MA26'. Since the national data protection law implemented in 2000 the municipal department 'MA26' was handed over this task. In addition to data protection, the municipal department also handles information law and marital status.

AKVorrat ('Arbeitskreis Vorratsdatenspeicherung')

The national working group 'Data Retention' was founded around the EU Data Retention Directive. The group has been highly critical of the Directive and collected signatures challenging the EU Data Retention Directive. In addition to a petition, the working group has been organising multiple public debates and information sessions regarding e-privacy.

ARGE Daten ('Austrian Society for Data Protection')

The Austrian Society for Data Protection 'ARGE Daten' was founded in 1983 as a working group and registered as an association in 1991. 'ARGE Daten' is one of Austria's leading privacy organizations. The organization is in close contact with universities, research institutions, as well as the industry and relevant authorities. Altogether the association counts approximately 700 members, most of them companies, public authorities, other NGOs and university staff. 'ARGE Daten' is a non-profit, non-governmental and politically independent membership corporation. Their mission statement stresses "*the protection of personal data and privacy in the age of global communication. The organization wants to achieve that information technology and telecommunication are used in a human way with social responsibility and under protection of privacy*.[37]"

VIBE.at (Organisation for Internet Users in Austria)

The organisation for Internet Users in Austria, 'VIBE' ('Verein für Internet-Benutzer Österreichs') was founded in 1999, resulting from discussions around Usenet. VIBE wants to represent the interest of Internet users with respect to authorities, Internet providers and other organisations.

[36] http://www.dsb.gv.at/site/6298/default.aspx
[37] http://www.argedaten.at/php/cms_monitor.php?q=AD-NEWS-LAST

Quintessenz
Quintessenz registered as an organisation in 1999, four years after it started out as the first eZine in german-speaking countries concerned with consequences of growing IT influence in different spheres of life. Following their motif 'data protection is a human right' they wage campaigns and raise awareness about data protection by sending out newsletters, and organizing events, for instance a yearly 'Big Brother Award'[38] ceremony held since 1999 in cooperation with privacy international and other Austrian privacy groups.

4.5.2 Oxford, United Kingdom

Information Commissioner's Office (ICO)
The UK's independent authority 'Information Commissioner's Office' (ICO) set up in is responsible to ensure information rights in the public interest, promoting openness by public bodies and data privacy for individuals. The ICO reports to the Parliament and is financed by the Ministry of Justice. As an independent National Data Protection Authority, it specifically deals with the Data Protection Act 1998 and the Privacy and Electronic Communications (EC Directive) Regulations 2003 across the UK; and the Freedom of Information Act 2000 and the Environmental Information Regulations 2004 in England, Wales and Northern Ireland and, to a limited extent, in Scotland.

Some of the tasks of ICO include: to manage a registry for data controllers, i.e. organizations that process personal data, to inform organizations about secure handling of personal data, dealing with enquiries and complaints, alongside advocacy work at the European Level. ICO is part of a working party under article 29 representing data protection authorities of the UK.

Open Rights Group
The Open Rights Group[39] (ORG), a non-profit organization was founded in 2005 by digital rights activists. ORG is a member of the European Digital Rights network; it is primarily based in London. ORG does advocacy and campaigning work with and for citizens in regards to topics such as, mass surveillance, to copyright, censorship, data protection and open data and privacy.

Responsible Data Forum
The Responsible Data Forum (RFD)[40] is developing "useful tools and strategies for dealing with the ethical, security and privacy challenges facing data-driven advocacy. RDF activities include organizing events; fostering discussion between communities; developing and testing concrete tools; disseminating useful information; and advocating for advocates and their supporters to improve the way they work with data." The RFD is collaborating with organizations such as Amnesty International, Aspiration[41] and many other international civil rights and advocacy organizations.

UK Data Archive

[38] http://www.bigbrotherawards.at/2014/vorschlagen.php
[39] https://www.openrightsgroup.org
[40] https://responsibledata.io
[41] https://aspirationtech.org/

The UK Data Archive[42] is the UK's largest collection of digital research data in the social sciences and humanities. The archive curates, organizes and provides access to data; several catalogues enable users to look through the data as well as encouraging users to deposit their data(sets) onto the platform. Currently, there are over 5,000 data collections (including national panel surveys as well as small-scale research data) available.

4.5.3 Malmö, Sweden

The Swedish Data Protection Authority (DPA)

A public authority responsible for the supervision of personal data processing. The main purpose of the DPA is to ensure that the processing of personal data does not violate any individual's personal integrity. The DPA is therefore responsible to make sure that existing laws on personal data applies, handle question and complaint's on personal data from private individuals and detect and prevent new threats to personal integrity (Datainspektionen 2015, 2).

The Pirate Party (PP)

A political party founded in 2006. The PP is interested in questions concerning the individual's personal integrity, both online and in society. The PP therefore advocates a society where neither states nor companies are allowed to conduct surveillance without suspicion of a crime (Piratpartiet, 2015). In the European Parliament elections 2009 the PP received enough votes to gain two seats in the parliament, but lost both of them in the election 2014. The PP does not have, and have not had, any seats in the Swedish Parliament.

4.5.4 Kopenhagen, Denmark

Danish Data Protection Agency

The Danish Protection Agency mainly ensures that the European Directive on Processing of Personal Data (Directive 95/46/EC) is abided by law. The agency provides guidance and information to companies, authorities as well as citizens regarding the processing of personal information. The agency also deals with complaints by citizens or it "*can also take up cases of its own initiative – own-initiative cases – if, e.g. due to a citizen enquiry or newspaper article, the agency suspects a violation of the regulations of the Act on Processing of Personal Data*[43]".

The IT-Political Association of Denmark (IT-Pol)

The IT-Political Association of Denmark[44] is an NGO organization who advocate for privacy, openness, and Internet freedom. IT-Pol is also a member of European Digital Rights (EDRi). IT-Pol drives politicians in dialogues on IT, gives presentations on conferences, schools and so forth. Main issues IT-Pol is dealing with include censorship, privacy, copyright and software patents.

[42] http://www.data-archive.ac.uk/home
[43] http://www.datatilsynet.dk/english/
[44] http://itpol.dk/presentation-of-it-pol

4.6 Checklist for Privacy Preservation

The following checklist offers questions that should enable decision-makers to reflect upon legal, social and ethical aspects of privacy preservation. All of the three aspects as discussed in the report are crucial in order to decide upon which and how to collect and publish data. Whereas legal aspects (e.g. EU Directive on Data Protection) are more clearly defined and easier to navigate, other privacy aspects remain more context-sensitive, including the granularity of data in order to avoid possible ramifications for individuals or groups. This checklist not only applies to single data sets but also to possible combinations of data, for instance geospatial data that can lead to negative ramifications for individuals or entire neighborhoods.

References to previous chapters are added to some questions in order to understand the linkages to the privacy discussions in the report.

Collection and Type of Data

- Which type of data are you working with?

☐ Open Data
☐ Closed Data
☐ Social Media Data

- Are you operating in line with the implementation of the EU legislation on Data Protection, Data Retention and E-Privacy in your respective country? [See chapter 4.4.2 – 4.4.6]
  Yes / No

- In case of using closed data, have the data owners given their informed consent?
  Yes / No

- If no consent has been given by data subjects, have data been anonymized?
  Strongly agree / agree / don't know / disagree / strongly disagree

Sensitivity of Data

- Were the social media data (e.g. Tweets, Facebook likes) openly accessible? [see chapter 3.2]
  Yes / No

- Are you operating with sensitive (personal) data? [see chapter 4.1]
  Strongly agree / agree / don't know / disagree / strongly disagree

- Is it public or private data? [see chapter 4.1]
  Yes / No

- Does the analysis of data (potentially) harm individuals or groups? [see chapter 4.1]
  Strongly agree / agree / don't know / disagree / strongly disagree

- Does the combination of data sets (potentially) harm individuals or groups?
  Strongly agree / agree / don't know / disagree / strongly disagree

- Is the granularity of data abstract enough? I.e. is any data published retraceable to individuals or exact addresses/ households?
  Strongly agree / agree / don't know / disagree / strongly disagree

5 CONCLUSIONS

A discussion about privacy in the context of big data and urban decision making proves to be complex. In order to disentangle the multifacetedness of privacy this report looked at various privacy aspects in the social, ethical and legal realms. In the context of the 'UrbanData2Decide' project a manifold analysis of privacy starting from conceptualizations of privacy to legal directives by the European Union is important as the project touches upon privacy concerns on multiple levels. The collection of data from different data sources (including big data and social media data) raises ethical as well as legal questions concerning the reuse, combination and publication of data (e.g. stigmatization of a neighborhood through data visualization of local crime rates).

Social and ethical aspects discussed in chapters 2 and 3 emphasized the importance to reflect upon the context in terms of type of data (e.g. validity of tweeds or data from national survey) and the setting in which information is collected (e.g. professional/private). The analysis of social media data for instance needs to reflect upon the context of tweets, e.g. background and possible intentions of the user, as well as shortcomings due to the digital divide and coming up with strategies to obtain additional information through other (non-digital) channels. In addition discussion about the complexity of Internet research (chapter 3) paid attention to the different actors (ICT specialists and programmers; providers and maintainers of infrastructure; Internet users and researchers) involved in the production of tools, designing of platforms and their privacy regulations and contributing content to the platforms. The discussion about Internet research as ethical practices further unraveled important facets of the value sensitive design approach, such as the incorporation of ethical values in technological design (e.g. informed consent of web browser cookies) as well as cultural differences of privacy values and concerns.

The growth of social media use, online communication via online blogs and smartphone users led to a new wave of privacy concerns. These concerns raise challenging questions as to what possible responsibilities users themselves have. An approach towards self-management of privacy is seen rather critically by as managing personal information in the future in the context of big data analytics is far too complex to manage for most individuals. As discussed in the social and ethical guidelines (chapters 4.1 and 4.2) the consideration of ethical values translated into technological design.

In addition the three main ethical guidelines in software engineering (ACM Code of Ethics; AITP Code of Ethics; Software Engineer's Code of Ethics) provide sets of standards of ethical conduct in the professional domain. Besides ethical guidelines, also the description of legal guidelines (see chapter 4.4) shall serve as a reference text for privacy questions faced throughout the 'UrbanData2Decide' project. The three main EU directives important for the project, which regulate the most important dealings with data in EU member states, are: the EU Directive on Data Protection, E-Privacy Directive and Data Retention Directive. An analysis of their implementation in the four case study countries proved important, as some laws differ within EU countries. Thus, the collection and use of big or open data in the case study countries, needs to bear national differences in mind.

6 REFERENCES

Adam, A. (2008) The Gender Agenda In Computer Ethics. *In The Handbook of Information and Computer Ethics*. Kenneth E. Himma, Herman T. Tavani.

Allen- Castellitto, A. (1999) Coercing Privacy. *William and Mary Law Review* 40, p.723– 757.

Banhof (2014) *Bahnhof aktiverar "plan B": erbjuder fri anonymisering*, www.banhof.se, published 2014-11-16, retrieved 2015-04-20

Barth, A. (2008) Design and Analysis of Privacy Policies. A Dissertation submitted to the Department of Computer Science, Stanford University.

Bellman, S., Johnson, E. J., & Kobrin, S. J. (2004) International differences in information privacy concerns: A global survey of consumers. *Information Society*, 20(5), p. 313–324.

Buchanan, E., Ess, C. (2008) Internet Research Ethics: The Field and Its Critical Issues. In *The Handbook of Information and Computer Ethics*. Kenneth E. Himma, Herman T. Tavani.

Boyd D., Crawford, K. (2012) Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon, Information, *Communication & Society* 15(5), p. 662-679. Online: http://www.tandfonline.com/doi/abs/10.1080/.VB8Tz_l_uCk

Copp, D. (2005) *The Oxford handbook of ethical theory*. Oxford University Press.

Crawford, K., Schultz, J. (2014) *Big Data and Due Process*: Toward a Framework to Redress Predictive Privacy Harms*, Boston College Law Review* 55 (1), p.93-128.

Datainspektionen (2015:1) Personuppgiftslagen, www.datainspektionen.se, retrieved 2015-04-20.

Datainspektionen (2015:1) Datainspektionens uppdrag, www.datainspektionen.se, retrieved 2015-04-20.

David, K., & Patterson, D. (2012) *Ethics of Big Data: Balancing risk and innovation*. Sebastopol: O'Reilly Media, Inc.

DeCew, In Pursuit of Privacy, Jeff Weintraub (1997) The Theory and Politics of the Public/PrivateDistinction, in: Jeff Weintraub & Krishan Kumar, eds., Public and Private in Thought and Practice: Perspectives on a Grand Dichotomy,Chicago: University of Chicago Press,1–42.

De Zwart, Melissa; Humphreys, Sal; Van Dissel, Beatrix (2014) Surveillance, big data and democracy: lessons for Australia from the US and UK, UNSW Law Journal. Online:

http://www.unswlawjournal.unsw.edu.au/sites/default/files/final_t3_de_zwart_humphreys_and_van_dissel.pdf

Donovan, A., Finn, R., Wadhwa, K. (2014) Report on legal, economic, social, ethical and political issue. Deliverable D2.1. BYTE 'Big data roadmap and cross-disciplinarY community for addressing socieTal Externalities, EU Project, 7th Framework [derived from: www.byte-project.eu]

Etzioni, Amitai (1990) 'The Limits of Privacy', Basic Books. NY.

Floridi, L. (2008) Foundations of information ethics. In: The Handbook of Information and Computer Ethics. Himma, K., Tavani, H. (Eds.).

Friedman, B., Kahn Jr, P. H., Borning, A., & Huldtgren, A. (2013) Value sensitive design and information systems. In: Early engagement and new technologies: Opening up the laboratory, p. 55–95. Springer.

Greenleaf, G. (2012) Global Data Privacy Laws: 89 Countries, and Accelerating: *Privacy Laws & Business International Report, Issue 115, Special Supplement [retrieved on 30.03.2015* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2000034]

Hilton, T. (2000) Information systems ethics: A practitioner survey. *Journal of Business Ethics*, *28*(4), p. 279–284.

Himma, K., & Tavani, H. (2008) *The Handbook of Information and Computer Ethics*. Kenneth E. , Herman T. Tavani.

Hoy, M.G., Milne, G. (2010) Gender Differences In Privacy-Related Measures For Young Adult Facebook Users. *Journal of Interactive Advertising*, 10(2), p. 28-45.

Jenkins, D. G., & McCauley, L. A. (2006) GIS, SINKS, FILL, and disappearing wetlands: unintended consequences in algorithm development and use. In *Proceedings of the 2006 ACM symposium on applied computing*. ACM. p. 277–282.

Johns, M. D., Chen, S.-L., & Hall, G. J. (2004) *Online social research : methods, issues, & ethics*. New York: P. Lang. Retrieved from http://www.loc.gov/catdir/toc/fy043/2002025388.html

Justitiedepartementet (2014), Datalagring, EU-rätten och svensk rätt, Ds 2014:23, www.regeringen.se, published 2014-06-13, retrieved 2015-04-20.

Kang, J. (1998) Information Privacy in Cyberspace Transactions. *Stanford Law Review*, 50(4), p.1193–1294.

Krupa, Y., Vercouter, L. (2010) Contextual Integrity and Privacy Enforcing Norms of Virtual Communities. *Security and Privacy, 2006 IEEE Symposium, p. 15 - 198.*

Madrigal, A. (2012) The Philosopher Whose Fingerprints Are All Over the FTC's New Approach to Privacy. The Atlantic, Online Magazine [derived 20.03.2015 http://www.theatlantic.com/technology/archive/2012/03/the-philosopher-whose-fingerprints-are-all-over-the-ftcs-new-approach-to-privacy/254365/]

Milne, G.R., Culnan, M.J. (2004) Strategies for Reducing Online Privacy Risks: Why Consumers Read (or Don't Read) Online Privacy Notices. *Journal of Interactive Marketing*, 18(3), p. 15-29.

Nissenbaum, H. (2004) Privacy as contextual integrity. *Wash. L. Rev.*, *79*, 119.

Nissenbaum, H. (2010) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Palo Alto, CA: Stanford University Press.
Piratpartiet (2015), Piratpartiet i korthet, www.piratpartiet.se, retrieved 2015-04-20.

PTS (2014:1) PTS kommer inte i nuläget att vidta åtgärder utifrån datalagringsreglerna, www.pts.se, published 2014-04-10, retrieved 2015-04-20.

PTS (2014:2) Angående lagring av trafikuppgifter m.m. för brottsbekämpande mål, www.pts.se, published 2014-06-16, retrieved 2015-04-20.

PTS (2015) Electronic Communications Act, www.pts.se, retrieved 2015-04-20m

Post, R. (1989) The Social Foundations Of Privacy: Community and Self In The Common Law Tort. *California Law Review*, 77 (5), p. 957-1010.

Radder, H. (2004) Pragmatism, ethics, and technology.

Regeringen (2014) Personuppgiftslagen, www.regeringen.se, published 2004-03-22, updated 2014-10-03, retrieved 2015-04-20.

Schultz, R. A. (2006) *Contemporary issues in ethics and information technology*. IGI Global.

SFS (2013) Lag (2003:389) om elektronisk kommunikation, www.riksdagen.se, retrieved 2015-04-20

SVD (2009:1) Sverige stäms för datalagring, www.svd.se, published 26 May 2009, retrieved 2015-04-20.

SVD (2009:2) Regeringen skjuter på datalagring, www.svd.se, published 2009-10-21, retrieved 2015-04-20.

Sheehan, K. B. (1999) An Investigation of Gender Differences in On-Line Privacy Concerns and Resultant Behavior. *Journal of Interactive Marketing*, 13(4), p. 24-38.

Solove, D. (2007) I've Got Nothing to Hide and Other Misunderstandings of Privacy. *San Diego Law Review, 44,* p. 745 -772.

Solove, D. (2012) Privacy Self-Management and the Consent Dilemma. *Harvard Law Review, 126,* p. 1180-1903.

Solove, D. (2002) Conceptualizing Privacy. *California Law Review,* 90(4), p. 1087-1155.

Stahl, B. C., Eden, G., Jirotka, M., & Coeckelbergh, M. (2014) From Computer Ethics to Responsible Research and Innovation in ICT: The transition of reference discourses informing ethics-related research in information systems. *Information & Management*.

Stone, G. (2006) Commentary, Freedom and Public Responsibility. In: *Chicago Tribune,* May 21, 2006, p. 11.

Taddicken, M. (2014) The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure. *Journal of Computer-Mediated Communication*,19, p. 248-273.

Tavani, H. (2008) Informational Privacy: Concepts, Theories and Controversies. In: The Handbook of Information and Computer Ethics. (Eds.) Himma, K., Tavani, H. John Wiley & Sons Inc., New Jersey.

Tong, R. (1999) Feminist Ethics. *Stanford Encyclopedia of Philosophy. [available online: http://plato.stanford.edu/archives/fall1999/entries/feminism-ethics/]*

TRUSTe (2006) Consumers have a false sense of security about online privacy – Actions inconsistent with attitudes. Conducted by TNS, December 2006. Available from http://www.prnewswire.com/news-releases/consumers-have-false-sense-of-security-about-online-privacy---actions-inconsistent-with-attitudes-55969467.html

Van den Hoven, J. (2001) Privacy and the varieties of informational wrongdoing. *Readings in Cyberethics. Jones and Bartlett, Sudbury, MA*, p. 430–442.

Van den Hoven, J. (2008) Moral methodology and information technology. In: The handbook of information and computer ethics. Himma, K., Tavani, H. (Eds.).

Van den Hoven, J. (2008) Information technology, privacy, and the protection of personal data, In: Information technology and moral philosophy. van den Hoven, J. Weckert, J. (eds.), Cambridge, Cambridge University Press, p. 331 – 322.

Vedder, A. (2008) Responsibilities for Information on the Internet. The Handbook of Information and Computer Ethics, Himma, K., Tavani, H. (Eds.)

Warren, S., Brandeis, L. (1890) The Right To Privacy. *Harvard Law Review,* 4(5), p. 193 - 196.

Waldo, J., Lin, H., Millett, L. (2007) Engaging Privacy and Information Technology in a Digital Age. The National Academies. [available online: http://www.nap.edu/catalog/11896.html]

Westin, A. (2003) Social and Political Dimensions of Privacy. *Journal of Social Issues*, 59(2), p. 431-453.

Wolfie, C. (2004) Kommerzielle Digitale Überwachung im Alltag. Erfassung, Verknüpfung und Verwertung persönlicher Daten im Zeitalter von Big Data: Internationale Trends, Risiken und Herausforderungen anhand ausgewählter Problemfelder und Beispiele. Studie im Auftrag der Bundesarbeitskammer, AK Wien.

Yao, M. Z., Rice, R. E., & Wallis, K. (2007) Predicting user concerns about online privacy. *Journal of the American Society for Information Science and Technology*, 58(5), p. 710–722.

7 ANNEX

7.1 Acronyms and Abbreviations

OECD Organisation for Economic Co-operation and Development

EU European Union

ACM Association for Computing Machinery

AITP Code of Ethics Association of Information Technology Professionals

7.2 Glossary of Terms

Personal Information

Personal Information can be understood as sensitive or intimate information (medical condition), any information about a person (income), as well as only personally identifying information (address). According to the European Union, "*personal data shall mean any information relating to an identified or identifiable natural person ('data subject')".*

Open Data

A piece of data is open if anyone is free to use, reuse, and redistribute it - subject only, at most, to the requirement to attribute and/or share- alike.[45]

[45] http://opendefinition.org