Changing limits of privacy? An historical analysis of privacy regulation in Denmark

Paper for the IPP2012: Big Data conference at OII September 2012
By Rasmus Helles & Stine Lomborg, University of Copenhagen

Abstract

This article presents an empirical study of the developments in the regulation of privacy in Denmark through the past decade, from 2000 to 2011. During this period, technologies for harvesting, storing and analysing logs of personal data, often compiled by digital systems in so-called ‘big data’ archives, have gradually become central resources in contemporary business intelligence.

Through an historical analysis of changes in the topics decided upon by the Danish Data Protection Agency, we offer an empirical grounding of the concept of privacy and its regulatory practice. We use Bordewijk and Kaam’s (1986) classification scheme of information flows as a conceptual model to examine which kinds of information flows give rise to regulatory consolidation and change, and identify practices of personal data processing that seem to escape regulatory action.

Introduction

This article presents an empirical study of the development in regulatory practice concerning the archiving, processing and use of personal data in Denmark from 2000 to 2011. The technological development during this period includes, for instance, the digitalization of public administration in Denmark, and the rise of so-called ‘big data’ and associated techniques for massive-scale analysis of behavioural data logged by digital systems (McKinsey 2011). Such developments have profound implications for policy and regulation (Margetts 2009), in part, because they raise issues of privacy and data protection to which the regulatory infrastructure must respond.

In Denmark, personal data are regulated by the Danish Data Protection Agency [henceforward, the Agency], an independent body acting under the authority of the Act of Processing of Personal Data (The Act on Processing of Personal Data 2011) [henceforward, the Data Act]. The Agency guides citizens, companies and public organizations regarding their treatment of and operations on personal data in abidance by the law, and inspects public authorities and private companies to ensure proper conduct in these organizations’ use, storage and processing of personal data.

The storage and management of personal data in a manner that respects individual privacy is essentially about who controls the information system, data access, and the exchange of personal data between citizens, government agencies and companies. That is to say, questions of privacy regulation are intimately intertwined with power relations in information flows. This article aims to contribute to the academic understanding of privacy by using new regulatory practices of data archiving as an empirical baseline for a nuanced understanding of social definitions of privacy today.

Through an historical analysis of changes in the topics decided upon by the Agency, we offer an empirical grounding of the concept of privacy and its regulatory practice. We use Bordewijk and Kaam’s (1986) classification scheme of information flows as a conceptual model to examine which kinds of information flows give rise to regulatory consolidation and change, and identify practices of personal data processing that seem to escape regulatory action.

Theoretical background

The paper investigates the regulation of personal data in Denmark for two different, but interrelated reasons. One is descriptive, and serves the interest of documenting and identifying the most central changes in notions and regulations of privacy. The Agency is the only regulatory authority in Denmark concerned with privacy issues outside of the courts (except for cases involving the press). While obviously not a complete or linear indicator of the relative importance or prevalence of various privacy issues, cases brought before the Agency often represent issues where no case law exists, and the decisions of the Agency therefore represents a central arena for the on-going (re)-definition of privacy *vis-à-vis* new technology.

The other reason is theoretical in nature: Privacy and big data are central issues in current critical theories of new media, and have already given rise to a substantial literature concerned with general (or philosophical) implications of the macroscopic changes that the rise of big data represent. A key implication of privacy is surveillance (Bennett, Raab, and Regan 2003), that is, the right of the state to monitor its citizens, and issues of identity and discrimination associated with the automated categorization of individuals (Gandy 2006). Another related, prominent discussion concerns the right to privacy in (online) behaviour (Zimmer 2008), including the right not to have personal digital traces data mined and exploited by companies in their quest for revenue, and served to advertisers as eyeballs for personally targeted advertising (e.g. Dijck 2009; Fuchs et al. 2012). Part of the academic criticism of user commodification points out that the complex and shifting obligatory privacy settings of companies such as Facebook and Google are intentionally put in place to obscure rather than clarify to users what they allow the companies to do with their personal data.

While significant in its own right, this literature often neglects some of the more commonplace varieties of big data usage, and focuses instead on more extreme examples of obvious importance with regard to privacy, for instance, direct surveillance and social sorting (Gandy 2006). It is not immediately clear that the perspectives and concepts developed in this literature travel well to the reality that faces users, developers and administrators involved in more mundane applications of big data production and use: Although based on the same technologies and principles, recommender services such as Amazon's "people who bought this book also bought..." clearly fall in a different category than the predictive systems used to prioritize subjects in immigration control: From a legal perspective, because they deal with different classes of personal information, and from an ethical perspective, because the potential consequences of their implantation have tremendously different scope.

Since not all applications of big data are equal with respect to their legal and ethical status, we argue that the Agency has a special and important role to play. The development of a new sub-domain, such as big data, under the Data Act must involve decisions on a series of cases, encompassing the range of relevant issues, ensuring that a frame of reference for prospective companies that seek to use big data and data mining exists. It is important in this respect to note that the historical role of the Agency in Denmark has not only been one of sanction and control but has also involved a strong element of dialogue and setting of examples. Although the Agency is nominally a statutory regulator, it *de facto* acts as co-regulatory body, similar to other agencies in the Danish administration (Helles, Søndergaard, and Toft 2011, , p. 11ff.). The Agency's decisions are often intended as examples, and the formulation of decisions often involves a dialogue with stakeholders working towards establishing a viable precedent for future cases.

Furthermore, the general character of the critical debate on big data and data mining overlooks the fact that the archival practices that constitute the technological backbone of big data analysis are regulated very differently between countries: The Danish law on data protection is among the strictest in the world (together with those of the other Scandinavian countries), and therefore many international developments take a different form (see e.g. Millard and Hon 2012). Hence, some of the criticism that the literature raises against global actors (e.g. intergovernmental intelligence agencies and international commercial businesses) and the lack of clear and internationally accepted regulatory practices and

standards is actually dealt with at a national level – in the case of Denmark by continuously adjusting and redrawing boundaries of the regulatory domain of privacy in light of technological and societal developments.

Categorizing change?

The regulation of privacy with respect to new media falls under the Data Act, which regulates all forms of automated handling of personal data. The Data Act includes definitions of what constitutes personal data, which are differentiated into different levels of sensitivity, so that, for instance, information about political or religious beliefs is considered to be more sensitive than a persons' home address. Different requirements are attached to the various kinds of personal information, and the law also prohibits many actors from storing some of the most sensitive kinds of information.

A central and recurring issue in the legislation and regulation of personal data is the speed with which computer systems develop and proliferate, since new systems not only contribute to a quantitative expansion of the regulatory domain, but often also bring qualitatively new issues into existence. For example, the spread of the internet afforded the transfer of data between servers located in different jurisdictions, which triggered a wave of issues concerning the preservation of rights and jurisdiction with regard to personal data stored and transferred between or internally in global corporations (Kuner 2007, , pp. 152-153). In a similar fashion, the proliferation of use of social networking services (e.g. Facebook) has triggered a new set of concerns that most likely require legal provisions to be reviewed and updated. Obviously, the legal systems of nations and regions (e.g. the E.U.) are hard pressed to keep up with developments in the technological domain. This means that regulatory decisions sometimes have to refer to general principles in the Data Act in making decisions for new types of problems, until the law-making process catches up.

The Data Act specifies different regulations for several domains where personal data are handled (e.g. private companies and municipal administrations), since for instance information about health and social conditions are held to be essential to the activities of certain public authorities, but not to most private companies.

The level of detail concerning the different practical domains that the law deals with is quite high, which means that a substantial proportion of the activity of the Agency deals with attributing cases to the relevant sections of the law after which they are subsequently decided.

When taken in combination, the two issues outlined above (the lag in law-making relative to the speed of technological innovation and the high level of detail in existing legislation and regulation) present an analytical challenge. Since many sections of the law are only manifested a couple of times in the sample it becomes more difficult to identify trends in the data material. This is particularly challenging because we are interested in tracking the scope and direction of innovations in the regulatory practice.

Simply mapping the frequency with which different sections of the Act are used in decisions across time does not by itself produce a reliable index of changes to the regulatory practice of the Agency. Many cases involve general decisions on what constitutes different categories of personal data, and so refer to the same general sections in the law.

Likewise, the specific subject matter of cases brought before the Agency display a huge variation across different topics (e.g. from the proper shape of log-on procedures in systems involving criminal records to the inclusion of someone on a mailing list), calling for an arbitrarily long list of categories to be included in the analysis.

Instead, we have developed a categorical system, based on the patterns of how information handled in a given system is provided, and how it is managed. This allows us to detect – on a macro scale – whether technological developments result in new questions for privacy regulation and thus for redrawing the boundaries of the regulatory domain.

The categories are derived from Bordewijk and Kaams (1986) classic text on information services. Bordewijk and Kaam propose a model for classifying computer systems according to the interplay between information flow and control that they involve. Their model organizes computer systems according to two dimensions, the source of the information handled in a given system, and the power to control how the system behaves when active (called the 'programme'). The cross tabulation of the two dimensions results in a matrix reproduced in Figure 1, listing four prototypical configurations of tele-information systems:

		Information issue by	
		Centre	Consumer
Programme Control	Centre	Allocution	Registration
	Consumer	Consultation	Conversation

Figure 1 Bordewijk and Kaams model of tele-information services (reproduced after: Bordewijk and Kaam 1986, : p.19)

In Bordewijk and Kaam's original contribution, the categorical system is designed to distinguish between different types of so-called 'tele-information systems', which at their time of writing essentially included different types of distributed client-server systems prevalent before the advent of the world wide web, such as bulletin board systems (BBS's), multi-user dungeons (MUD's), and professional dial-in systems (book keeping systems etc.).

A key advantage of their categorization is that it is fundamentally based on dimensions of information and control, which corresponds closely to fundamental principles in the Data Act (cf. Waben and Nielsen 2008), while at the same time allowing interpretive flexibility in assigning cases to categories irrespective of the specific sub-domain in the Data Act. For the purposes of our investigation, the categories were adapted to the domain of privacy concerns by interpreting information only as those kinds of information regulated in the Data Act (sensitive, semi-sensitive and ordinary personal information). Likewise, the subjects included in the original article were operationalized in relation to the privacy domain: the 'consumer' category was re-defined as the subject of personal information, and the 'centre' category was understood to be the entity controlling and/or handling the information in question.

The prototypical categories each correspond to several different sub-domains in the Data Act:

Allocution. In Bordewijk and Kaam's context, allocution essentially refers to the broadcasting of information, since both the control of what is broadcast and the process of broadcasting itself lies in the hands of the centre. In the context of privacy, we take allocution to encompass cases where personal information is made publicly available without consent from the person who is the reference point of the information in question. A central example in the context of privacy is the (often accidental) broadcasting of personal information on websites: a number of cases deal with companies or public authorities that accidentally share sensitive information about customers or citizens on their websites, for instance by publishing PDF files of power point slides containing specific information about hospital patients.

We also include in this category cases where information is passed on 'en bloc' between two centres – for instance, when registered information is made available to other parties. This involves for example the commercial exchange of databases containing customer information.

Registration. In the 'registration' category, the consumer provides information, while programme control resides with the centre (understood here as the entity responsible for the computer system in question). This includes, for example, cases concerning the security measures which a municipality must observe when storing and handling health-care information about citizens. The Data Act specifies how personally sensitive information must be handled, including mandatory guidelines for limiting the number of personnel involved in handling the information. We also include in this category the various sorts of data enrichment that typically involve statistical treatment of records, e.g. extracting use patterns from existing customer data by developing taxonomies of customers or calculating predictive scores for possible future customer actions (e.g. the propensity for a given customer to buy a given product when given a specific offer). Several of the procedures that are often labelled 'data mining' thus fall in this category, when they involve personally sensitive information, or (more broadly) the handling of customer data without informed consent.

Consultation. In the 'consultation' category, the centre provides the information, while programme control resides with the consumer. We include in this category cases where users must actively consult systems in order to extract information. This involves cases where accessing a specific sub-section of a website through typing a specific path in the browser gives access to information that is not intended for outside access (e.g. if a company or public entity accidentally allows web access to a portion of their intranet but does not provide a link to it, or when a known error in a system is left uncorrected, allowing people with sufficient technical insight to access information in the system).

Conversation. For Bordewijk and Kaam, the category of conversation denotes information flows where consumers act as both information sources and programme controllers. Conversation includes exchange between two users, e.g. in the form of a telephone call. Although a telephone call clearly involves aspects of registration and consultation with respect to the computer system involved in routing the call from one phone to the other, they maintain that conversation is the most important aspect of this. Although clearly relevant to privacy discussions, the privacy of the content of technologically mediated conversations are not primarily regulated by the Data Act, but by the Penal Code, so only one case in the sample involves conversation in the more technical sense implied here. The concrete case concerns the use of biometric identification devices such as the use of fingerprints to identify the users of a transportation system, where users are registered by a scanner that extracts a check sum value based on characteristics of the print that cannot be reverse-engineered to the actual fingerprint, and stores only that value.

Method and sampling

The empirical study consists of a content analysis (Krippendorff 2004; Krippendorff and Bock 2009) of the cases brought before the Agency from 2000 to 2011, based on a sample of all the cases published by the Agency on their website (N=246). The sample is not representative of the entire collection of cases that the Agency handles. Rather, the cases selected for publication on the Agency's website represent cases of a more fundamental nature (e.g. a new subject matter, or changes to existing regulatory practice), or cases that concern important focus areas for policy development (e.g. following recent trends of political debate and policy-making concerning surveillance).

We coded all the cases according to a coding scheme based on our operationalization of Bordewijk and Kaam's (1986) conceptual model of information flows. Additionally, we coded the entire sample according to a set of basic variables to gain greater insight into variations in the data material. These variables were the date and time of the case being opened and concluded; the type of actors involved (e.g. public administration, private business, individual citizen); the occasion for addressing the Agency (e.g. complaint, request, notification); the type of media in question (e.g. standard archive, email, internet based archive); the content of the case (e.g. secure data storage, encryption, marketing, personally sensitive data, cross-referencing of databases); the ruling of the Agency, the applied sanctions, the legal paragraphs referred to, and the tags used by the Agency.

Findings

As a first step towards determining changes in the regulatory practice during the period under study, we analysed the types of media technologies involved in the sampled cases. Table 1 shows a breakdown of the prominence of various media technologies in the data set.

Table 1 Total distribution of cases according to media types

Media technology	Occurrences	Occurrence in % of total case load[1]
Digital archives	190	77,24%
Internet	132	53,66%
TV surveillance	30	12,20%
Mail	11	4,47%
Photography	5	2,03%
Other	28	11,38%

As Table 1 shows, digital archives and the internet are the two predominant media in the data. Issues of digital archives are present in more than 75 percent of the cases, whereas a little more than half of the cases involve the internet.

Table 2 Comparison of the distribution of media types over time

	Development over time	
	2000-2005	2006-2011
	Count (relative share)	Count (relative share)
Digital archives	61(50.0%)	* 129 (47.1%)
Internet	35 (28.7%)	*** 97(35.4%)
TV surveillance	8 (6.6%)	22 (8.0%)
Mail	4 (3.3%)	7 (2.6%)
Photography	1 (0.8%)	4 (1.5%)
Other	13 (10.7%)	15 (5.5%)
Total	122 (100.0%)	274 (100.0%)
Development between periods was tested using chi-square tests and is significant when marked. Legend: * sig. at α = .05, ** sig. at α = .01, *** sig. at α = .001		

Furthermore, when considering the development over time, as displayed in Table 2 above, digital archives and the internet increase in relative prominence from the first half of the decade to the second. There is a significant increase in the number of cases involving digital archives from the first to the second half of the decade. The number of cases involving the internet also increases very significantly in prominence during the period. Thus, evidently, developments in the regulatory domain follow larger technological trends.

[1] The categories are not mutually exclusive, as each case may concern a number of technologies at the same time.

Given the Agency's efforts at continuously revising the regulatory framework in light of new technological possibilities, what kind of issues are reflected in this development? To address this question, we analysed the total distribution of cases according to the four types of information flows derived from Bordewijk and Kaam (1986), displayed in Figure 2 below.

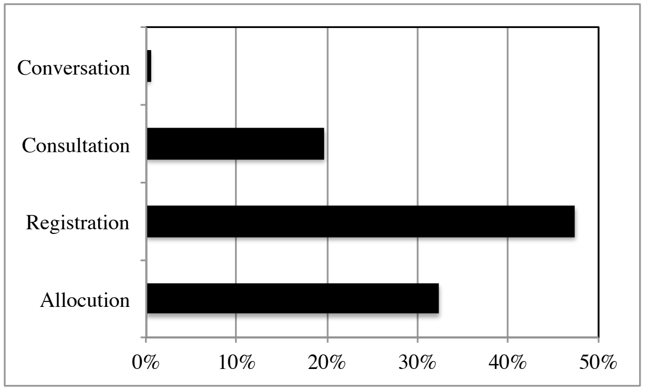

Figure 2 Distribution of sample cases

Figure 2 shows that almost fifty percent of the cases published by the Agency concern issues of registration. Allocution accounts for about one third of the sample and is thus also prominent in establishing regulatory practice, whereas consultation accounts for about twenty percent of the cases. As mentioned earlier, only one case in the sample concerns conversation.

Considering technological developments over the past decade, it is probable that regulatory practice has undergone changes in terms of the prominence of registration vis-à-vis consultation and allocution. For instance, one would expect an increase in the 'registration' category because the digitalization of intra-organizational activities as well as business-to-client relationships entails a wealth of new opportunities for creating and performing operations on digital archives. To examine possible changes in the practices of regulating privacy, we analysed the relative share over time of the four different patterns of information flow, as displayed in Table 3.

Table 3 Comparison of the distribution of cases over time according to communication categories

	Development over time	
	2000-2005	2006-2011
	Count (relative share)	Count (relative share)
Allocution	17 (22.3%)	52 (38.0%)*
Registration	41 (54.0%)	60 (43.8%)
Consultation	17 (22.3%)	25 (18.2%)
Conversation	1 (1.3%)	0 (0.0%)
Total	76 (99.9%)	137 (100.0%)
Development between periods was tested using chi-square tests and is significant when marked. Legend: * sig. at $\alpha = .05$		

As Table 3 shows, the relative prominence of registration and consultation decreases over time, whereas allocution increases. However, only the development concerning allocution is statistically significant.

Although the relative prominence of registration in the sample decreases with more than ten percentage points, this development is not statistically significant. Hence, we cannot conclude that any regulatory change in regard to registration has taken place. This is somewhat unexpected, considering the aforementioned significant increase in cases concerning the collection and use of digital archives. This suggests that cases concerning what can be registered, who can create personal data registers, and what forms of operations they can perform, are not becoming more important for defining the development of the regulatory domain over time. Moreover, when taking a close look at the actual cases in the registration category, even in the second half of the decade under study, the cases very seldom concern issues of data handling, including data mining. That is, issues pertaining to what personal data registers are used for in or between organizations are not central to defining the regulatory domain of privacy in Denmark.

The relative share of cases in the sample involving consultation is decreasing over time, but again, the finding is not statistically significant. The lack of increase of regulatory attention to issues of consultation is surprising, especially when considering the digitalization of various domains of public administration and private services in Denmark. This digitalization involves not only digital registers of citizens' fiscal affairs, banking, and healthcare, but also an expanded use by citizens of these digital public services, and a broad diffusion in of integrated office systems (e.g. SharePoint) that allow users in an organization to share data across systems. With these digital systems follow security breaches and thus incidents of unintended access to third parties, and cases about security breaches account for the majority of cases involving consultation. However, the Agency does not proactively adjust the regulatory domain to other issues of consultation, although the Agency reactively attempts to keep up with technological developments in Danish society.

Allocution is the only communication category that develops significantly and thus suggests changes of established regulatory practice However, the increasing prominence of allocution reflects a remarkable bias in our sample. As evident from Table 3, allocution plays a minor role in the first part of the decade, a role that is not much different in the final two years of our study. However, allocution peaks in the period from 2007 to 2009. During this period, the Agency posted a number of similar cases, with very similar rulings concerning the accidental publication of personal data on the websites of various types of organization. For instance, a Danish university incidentally published the civil registration numbers of students and was reprimanded by the Agency, which then initiated investigations of – and found – similar breaches in other universities' websites. The fact that the Agency published many such cases indicates a wish to establish a precedent by pushing the same argument concerning personal data processing over and over again. When controlled for this sample bias, allocution no longer involves a significant development.

In summary, the regulatory practice of privacy in Denmark displays a somewhat contradictory trajectory. On the one hand, the analysis documents that a regulatory shift has taken place concerning media types, from the first to the second half of the decade. The significantly increased prominence of internet and digital archives in our sample suggests that the regulatory domain responds to the technological development. On the other hand, the lack of development in the relative share of the four communication categories seems to suggest that despite dealing with new technologies, the Agency still addresses the same kinds of questions and thus struggles to keep up with the developments in the use of personal data archives that follow from the technological development.

Discussion

The general characteristics of the information flows and power relationships reflected in the sampled cases remain largely unchanged across the period of study, suggesting that the overall composition of the domain regulated by the Agency is the same now as it was 10 years ago (see Table 3).

On the one hand, this persistence in the distribution of cases across time is what might be expected of most public bodies acting in accordance with relatively fixed (or at least slowly evolving) legal provisions. On the other hand, however, the changes in the technological landscape surrounding the activities of the

Agency have been profound, and some of these clearly involve transformations that directly concern the legal limits for can be done with personal information and by whom.

One clear candidate for these kinds of technological shifts is the assortment of statistical and analytical procedures that are often labeled data mining (Linoff and Berry 2011), frequently applied in customer relation management and various forms of (direct) marketing efforts. Data mining is normally based on the customer databases of large companies, which often involve not only basic customer information such as name and address, but also detailed information on purchasing history, economic information and customer/company communication etc. While no overview of the total data mining activities in Danish companies exists, it is clear that such techniques are increasingly used in various forms. However, our analysis of the activities of the Agency shows no indication of this development being reflected in regulatory practice. As demonstrated, the relevant communication category, registration (see Table 1), has seen no statistically significant change over the past decade (see Table 3).

In fact, only one case in the entire sample concerned actual data mining practices. The case involved the company Experian (a credit monitoring company), which was developing a system for scoring the predicted likelihood of prospective loan-takers to default on their payments. Arguments presented in the final (negative) decision by the Agency clearly demonstrate that such practices do in fact involve a number of relevant privacy concerns, and that the case clearly falls within the jurisdiction of the Agency and is regulated under the Data Act.

There are several, likely explanations for the absence of predictive data mining from the activities of the Agency. One in particular follows from the principles underlying their classification as registration. Registration involves cases where users provide information, but programme control is handled by the centre. This means that the process and resulting actions are rarely transparent or even detectable from the point of view of users or customers. Often, the results of data mining analysis take the form of scores, indicating the degree to which a customer is likely to act in a given way (e.g. accept an offer of a service at a given price) or a designation in a classificatory scheme of a company's user base. This rarely involves any direct input from customers, but happens entirely on the basis of information already present in the database.

The difficulty in detecting the use of personal information for predictive marketing and similar purposes means that regulators are faced with the difficult and cost-intensive task of proceeding by physical inspection of candidate companies, rather than reacting to complaints or inquiries from the public. This, further, is in line with the rise in cases involving allocution: many of these cases concern the accidental publishing of personal information on websites, which is

detectable to a broad share of the population, resulting in reports from concerned or involved citizens to the Agency.

Conclusion

Our findings indicate that while regulatory practice has developed substantially to encompass new media types, decisions continue to fall proportionally in the same categories across the period understudy.

Two key questions seem to define the regulatory framework: the first concerns what can be registered (and by whom), as demonstrated by the wealth of cases concerning registration; the second concerns the proper storage and secure access to personal data archives and is demonstrated in the caseload concerning both allocution and consultation. At the same time, however, there is a notable lack of focus in the case sample on questions about what personal data archives may be used for. Hence, issues concerning the integration of data archives, data mining with a view to commercial gain, and other similar uses of personal data do not receive the Agency's attention, either because the Agency does not consider such issues to be of primary importance to privacy regulation, or because these uses of personal data are simply difficult to detect and thus regulate.

The lack of regulatory activity with respect to newly arisen practices involving personal data does not necessarily point to a lack of specific regulatory authority, but (more importantly) to a need for innovation of the regulatory practice. As mentioned, the role of the Agency includes a dialogic component, in which new issues for privacy regulation are negotiated through a dialogic process of regulatory decision-making. This role is a central element in the regulatory culture of the Danish political system, and the continuation of this culture may require the adoption of new, proactive regulatory instruments, or at least that the existing level of inspection and control are intensified. In the specific case of regulation of data mining, this follows from the opaque way in which results may be used – it can simply be hard for citizens to detect if data about themselves or past behavior has been collected and used by other parties. In a more general perspective, the call for a more proactive regulation follows from the general diffusion of personally sensitive information into new kinds of systems handled by new kinds of actors.

Given that much of regulatory action still lies with national authorities, and that the regulation of personal information differs substantially across boarders, our findings suggest comparative research as a potentially fruitful avenue. Results from comparative studies of privacy regulation can shed light on the relative speed in changes to regulatory practices between countries, and may also help identify useful, regulatory innovations from other countries.

References

The Act on Processing of Personal Data. 2011. http://www.datatilsynet.dk/english/the-act-on-processing-of-personal-data/read-the-act-on-processing-of-personal-data/compiled-version-of-the-act-on-processing-of-personal-data/.

Bennett, Colin, Charles Raab, and Pricilla Regan. 2003. "People and Places. Patterns of individual identification with intelligent transportation systems." In *Surveillance as Social Sorting. Privacy, risk, and digital discrimination*, ed. David Lyon. London: Routledge. 153-175.

Bordewijk, J., and B. Kaam. 1986. Towards a new classification of tele-information services. *Intermedia* 14 (1): 16-21.

Dijck, Jose van. 2009. Users like you? Theorizing agency in user-generated content. . *Media, Culture & Society* 31 (1): 41-58.

Fuchs, Christian, Kees Boersma, Anders Albrechtslund, and Marisol Sandoval, eds. 2012. *Internet and Surveillance: The Challenges of Web 2.0 and Social Media*. London and New York: Routledge.

Gandy, Oscar H. 2006. "Data mining, surveillance, and discrimination in the post-9/11 environment." In *The new politics of surveillance and visibility*, eds. Richard V. Ericson and Kevin D. Haggerty. Toronto: University of Toronto Press. vi, 386 p.

Helles, Rasmus, Henrik Søndergaard, and Ida Toft. 2011. *Does media policy promote media freedom and independence? The case of Denmark*. Athens: Mediadem.

Krippendorff, Klaus. 2004. *Content analysis. An introduction to its methodology*. 2. ed. Thousand Oaks, Calif.: Sage Publications.

Krippendorff, Klaus, and Mary Angela Bock. 2009. *The content analysis reader*. 1. edition ed. Thousand Oaks, Calif.: Sage Publications.

Kuner, Christopher. 2007. *European data protection. Corporate compliance and regulation (2. edition)*. Oxford, UK: Oxford University Press.

Linoff, Gordon, and Michael J. A. Berry. 2011. *Data mining techniques : for marketing, sales, and customer support. 3rd edition*. New York: Wiley.

Margetts, Helen Z. 2009. The internet and public policy. *Policy & Internet* 1 (1): 1-21.

McKinsey. 2011. *Are you ready for the era of 'big data'?* : McKinsey Global Institute.

Millard, Christopher, and W. Kuan Hon. 2012. Defining 'personal data' in e-social science. *Information, Communication & Society* 15 (1): 66-84.

Waben, Henrik, and Kristian Korfits Nielsen. 2008. *Lov om behandling af personoplysninger med kommentarer. 2. udg*. København: Jurist og Økonomforbundets Forlag.

Zimmer, Michael. 2008. The Externalities of Search 2.0: The Emerging Privacy Threats when the Drive for the Perfect Search Engine meets Web 2.0. *First Monday* 13 (3).