Towards Developing an Educational Maxim for Privacy and Data Protection in Learning Analytics

Tore Hoel and Weiqin Chen
Oslo and Akershus University College of Applied Sciences, Norway
{tore.hoel,weiqin.chen}@hioa.no

Abstract: Privacy and data protection policies are based on legal regulations, which in turn get their justification from political, cultural, economical and other kinds of discourses. Applied to learning analytics (LA), do these policies also need a pedagogical grounding? This paper is based on an actual conundrum in developing a technical specification on privacy and data protection for LA for an international standardisation organisation. Legal and cultural contexts that make it a challenge to define universal principles for privacy and data protection are explored. If not universal principles, could consent be the point of departure for assuring privacy? In education, this is not necessarily the case as consent will be balanced by organisations’ legitimate interests and contract. The different justifications for privacy, the legal obligation to separate analysis from intervention, and the way learning and teaching works makes it necessary to argue data privacy from a pedagogical perspective. The paper concludes with three principles that is proposed to inform an educational maxim for privacy and data protection in learning analytics.

Keywords: Privacy • Data protection • Learning analytics • Data privacy

1. Introduction

Privacy and data protection measures are promoted and justified by laws and regulations. For European countries this is abundantly clear as 25 May 2018 approaches quickly, i.e. the date when the EU General Data Protection Regulation (GDPR) becomes legally binding and applicable. The discussion on how GDPR will impact computerised processing of personal data is just the pinnacle of more than 50 years of international debate on privacy. In the USA in the 1960s, “privacy” was invoked as a key term for summing up “the congeries of fears raised by the (mis)use of computers” (Bygrave, 2010). Privacy was not the only term; a “variety of other, partly overlapping concepts have been invoked too, particularly those of ‘freedom’, ‘liberty’ and ‘autonomy’” (ibid., p. 167).

When raising the discussion of privacy and data protection in a new context – i.e., learning analytics (LA) – we have to factor in the very complex global data protection scene where legal regulations are debated on a background of diverse political, cultural, economical and even philosophical ideas. The question raised in this paper is whether there also are pedagogical ideas that should be brought to bear. For example, are there specific educational requirements that will justify a practice that goes beyond what is required by law, e.g., the GDPR? If this extra requirement is found, it should ideally be summarised in an educational maxim that would resonate well enough to bridge some of the gaps we find between different legislations and cultures related to how privacy is valued or conceptualised.

This paper aims at exploring the grounds for this educational ‘extra’ that would allow us to be bold in involving the students in self-managing their own data used for learning analytics. We do this exploration on the backdrop of a heterogeneous international landscape regarding the rights of the individual and the value of privacy.

To give a practical context for this exploration, a June 2017 snapshot is provided of development work in ISO/IEC JTC1/SC36, the ISO committee working on interoperability standards for learning analytics. In the first working draft of a new technical specification it was admitted that privacy is difficult to define restrictively “as privacy is an elusive concept that means different things in different countries around the world. What is seen as an intrusion into the private life or affairs of an individual, and whether gathering of data about the individual is seen as undue or illegal varies with cultural context” (T. Hoel, personal communication, June 2017). The editors of the draft specification suggested that privacy problems should be looked at “in a LET [learning, education and training] context to be able to specify privacy and data protection principles for LET
that address specific problems and support a good learning environment for the individuals involved”. They laid down the following principle for development:

“The educational context of LA requires that the right to be informed is not interpreted restrictively; it is a pedagogical value of its own to be as open as possible about data collection and processing.”

And regarding the legal requirements of notification of the data subject about data collection, the working draft states:

“Age of the students, the educational setting, matters of authority, and other reasons could influence how notification of data collection and processing will be conceived. The educational context is, however, an opportunity to clarify privacy and data protection issues related to use of LA” (T. Hoel, personal communication, June 2017).

From this early working draft, it is clear that the authors of the standard try to carve out an educational argumentative space that would allow for certain policy principles regarding privacy. In this space one finds arguments about involvement of students, openness, and what we could term educational opportunity (‘you should teach about big data, data management, and privacy – here you have an opportunity to do so’). What are the needs and background for attempting to establish such an educational argumentative space related to privacy for LA? And how should such a set of arguments be received in the truly international setting of current LA practices? These are the questions that will drive further development of this paper.

2. A Universal Right to Privacy?

Even if educational policies often are the purview of local authorities, when we talk about educational technologies – like LA – we are dealing with global solutions that have to cater for all political and cultural climates. In this section we examine privacy and data protection in an international perspective, starting by asking if there is a universal right to privacy.

Milberg, Burke, Smith, and Kallman (1995) stated that it could be reasonably argued that protection of personal information privacy was a “hypernorm”, a principle fundamental to human existence. “If this is so, then managers have an obligation to protect personal information privacy in every system and in every country, regardless of distinctions in national levels of concern or of regulatory approaches” (Milberg et al., 1995, p. 73). However, research on the relationships among nationality, cultural values, personal information privacy concerns and information privacy regulations led Milberg et al. (1995) to conclude on a more pragmatic note: “Executives may choose to reject the ethical ‘hypernorm’ argument (…) But the threat of negative impacts on the bottom line, driven by both market forces and the legislative agenda, should be sufficient to prod them toward a more enlightened view of the personal information privacy management domain.” Further research by Milberg, Smith and Burke (2000) found that most firms took a primarily reactive approach to managing privacy “by waiting for an external threat before crafting cohesive policies that confront their information practices”.

When ideals meet stakeholders’ interests, trade-offs are inevitable. Milberg et al. (2000) find that “[a] right to privacy” has been taken to include a number of “interests” that converge and diverge, and they use targeted marketing as an example of trade-offs between the privacy interests and how society’s economic and social systems function:

“While organizations argue that they have the right to conduct business, consumers and privacy advocates often claim the right to be free of unwanted solicitations. While organizations claim the right to use information technology to improve efficiency, consumers often exhibit the desire to control the flow and dissemination of their personal information. While businesses claim the right to record information generated from their transactions, consumers increasingly want to know that this information has been gathered and stored and to control its uses” (Milberg et al., 2000, p. 36).

Trade-offs between ideals and reality may not be the best way to understand how privacy and related interests with regard to the processing of personal data are protected internationally. Alternatively, one could see how these issues are conceptualised in different countries, and how the different discourses express values that are taken up by different regulatory policies. Such an analysis is beyond the remit of this paper. However, we will bring some highlights from a study by Bygrave (2010), who explored the prospects for regulatory consensus.
Bygrave found that data protection laws in various countries “expound broadly similar core principles and share much common ground in terms of enforcement patterns” (Belgrave, 2010, p. 198). Nevertheless, “extensive harmonisation at the global level is extremely unlikely to occur in the near future” (ibid., p. 199). The reason for this lack of harmonisations is the strength of “ingrained ideological/cultural differences” (ibid., p. 199).

The differences are reflected in the nomenclature used in different parts of the world. The ISO project referred to above, used the composite term ‘privacy and data protection’ to bridge the North American and European policy discussions. Maybe ‘data privacy’ might have been an alternative compromise. We see that the GDPR uses ‘data protection’ as the main term, but also referring to ‘privacy’, though less frequently (European Commission, 2016). Even if the two concepts are closely linked they are not identical. ‘Data protection’ “is typically reserved for a set of norms that serve a broader range of interests than simply privacy protection (…) data protection is increasingly being treated in European law as a set of rights that are separate to the more traditional right to respect for privacy or private life” (ibid, p. 168-169).

In the USA, most discourse on privacy and privacy rights tends “to focus only on the benefits these have for individuals qua individuals (…) while German jurisprudence “emphasises that the value of data protection norms lies to a large degree in their ability to secure the necessary conditions for active citizen participation in public life; in other words, to secure a flourishing democracy” (ibid., p. 171-172). While Germany has had the most comprehensive and well-established legislative platform for data protection, the USA has had an absence of comprehensive data protection legislation. Germany will have to harmonise with the other European countries when GDPR comes into effect. Globally, it is expected that GDPR will have an influence on future legislation in countries outside of Europe in the same way as the EU directive 95/46/EC has done, not least as it has placed (and GDPR will continue to place) a qualified prohibition on transfer of personal data to countries that do not provide “adequate” levels of data protection. Again, it is the free flow of information, and thus trade, that speaks the language. The formal normative basis for the data protection laws may well be derived “mainly from the catalogues of fundamental human rights” (Bygrave, 2010, p. 180), when it comes to applying these principles in international instruments, money talks. In the Asia Pacific region, for example, the approach “appears to foster data protection regimes less because of concern to protect basic human rights than concern to engender consumer confidence in business” (Bygrave, 2010, p. 188).

Hoel, Griffiths, and Chen (2017) analysed three privacy frameworks, which have inspired legal development in all parts of the world and put the frameworks and selected countries on a scale with values between a focus on the individual and a focus on the organisation (Figure 1).

![Figure 1. Individual vs. organisational focus of LA beneficiaries, privacy frameworks and countries](image)

The case studies of the LA privacy discourse in Europe and Asia (Japan, Korea, and China) (Hoel et al., 2017) showed that concerns about the rights of the individual in relation to control of data emanating from the learner are in some respect a western phenomenon. In the East, where the interest of the individual more often is projected against the interest of the group the organisation is more prominent in the discourse on who should benefit from LA.

In this section we have seen that even if the concern for data privacy is shared among the general public around the world there is a long way to go from concern, at least in the abstract, to finding a common normative basis for establishing data protection policies. The global ideological landscape does not invite to subscription of human rights ideas or other shared normative ethics principles to motivate regulatory consensus on data protection. Lately, both societal and individual arguments have made the discussion on privacy more complex.
War on terror, national security, promotion of trade and new economies are all factors that demand extensive sharing of personal data. We also see that the users of ICT services are willing to undermine their own rights as soon as they see short time benefits of opening up access to their personal data (Hazari & Brown, 2014). In the next sections we will explore how involvement of the individual could be used to justify data sharing.

3. Data Privacy by Asking for Consent

“Obtaining valid consent from data subjects in connection with the use of personal data for analysis and profiling purposes is the best insurance against violating data protection legislation. The new European Data Protection Regulation also proposes restricting the opportunities for the processing of personal data on legal grounds other than consent” (Datatilsynet, 2013, p. 49).

It is interesting that the Norwegian Data Protection Authority uses the phrase “best insurance” in their 2013 report *Big Data - privacy principles under pressure*. Risk minimisation is the word of the day now as industry and public organisations alike scramble to prepare for the advent of GDPR, setting up accountability systems, documenting what information one holds, assigning data protection officers, and taking other organisational measures. However, risk management is a different strategy than invoking rights, and such a strategy certainly chooses the organisational perspective as opposed to the individual perspective that comes with arguing from rights. So, what does it mean when The Norwegian Data Protection Authority states as their primary recommendation to meet the challenges of big data: “consent [is] still the point of departure” (Datatilsynet, 2013)? Is consent, in the context of LA, the primary point of departure?

Focussing on consent means bringing the individual into the centre of the discussion. And that means the individual as an actor with rights to decide on data management, not as object in need of protection by others. However, consent in the age of Big Data is not straight forward. The Norwegian data protection authority points to claims “that the constant demand for consent on the Internet paradoxically may result in poorer protection for the individuals” (ibid, p. 50). Now, with the new GDPR asking for wide-ranging consent may be a lesser problem. The new Regulation has strengthened the protection from giving your rights away by ticking boxes when launching software solutions. The problem with consent, we would argue, does not so much lie in hollowing out the consent mechanism as with the fact that consent is not the sole legal ground for access to personal data. And pretending it is, will confuse the individual and undermine the individual’s ability to manage one’s own data.

In an educational setting, there are a number of stakeholders with legitimate rights to a person’s data, driven by the fact that the student has an obligation to go to school or has registered for a course, and in practice entered a contractual relation with an (business) organisation. It is not clear cut what the legal grounds for access to data are, let’s say for an administrator, a teacher, or a third party. Data about a student starts to build up from the moment the student does a web search in the course catalogue, right up to the clicks made browsing through learning resources, passing tests, and getting an exam. An educational institution is a business organisation with student records, which are not under the full control of the students. Nobody will contest that right of the institution to store and analyse data about who is registered for what course, and who ends up with what exam results. But what about the results of micro tests? There are no clear boundaries between data that are solely the student’s prerogative to manage, and data that the institution, the teacher, has a right to process (Zeide, 2017). These are issues that are subject to negotiations between parties that will base their positions on legal, moral, and pedagogical grounds.

We asked if consent is the primary point of departure within an educational context, and we have answered no. If we overlay the discussion in this section with the observations made in section 2 on the normative basis for privacy in a global perspective (Figure 2), we see that the role given to consent (and to the individual) could vary a lot in different cultural, political, legal and regional contexts. There is a need to explore more in detail how different scenarios for consent from learners in connection with use of personal data for learning analytics will play out. And we also see there is a need to explore the educational perspective on data privacy.
4. Balancing Interests for Big Data Analysis

Once leaving the abstract reflection on privacy and entering the field of practical data handling we see that the context and the purpose of data collection are important for how data privacy should be handled. As an example, let us compare how equally sensitive personal data gathered from passing through an airport, visiting a hospital, and taking part in education are handled. Public interests will trump any objection from the individual to be scanned by security cameras in the airport. In contrast, in a hospital, the individual has an absolute right to be a party to the data processing, and in extreme cases have the right to refuse to be given lifesaving treatment. Health and education are quite similar in that the individual is very much “part of the treatment”, and therefore consent should be sought; however, there are differences. Some education is compulsory. If not consent is justification for processing personal data, there must be others, i.e., contract, legal obligations, vital interests, public interest, or legitimate interest of the controller (Article 7 of the EU Data Protection Directive soon to be replaced by GDPR).

Figure 3 describes how a decision to ask for consent is a balancing act weighing different interests considering the different justifications for collecting and processing personal data.

In education, especially in the new data-driven practices involving use of online platform and sensor data, we do not think the data controller will be justified never to ask for consent invoking legal obligation or vital interests (the right side of the continuum in Figure 3). Contract or legitimate interests (e.g., business reasons) on the other hand, would be convenient to invoke, to allow data collection and processing without too much interference of the individual. However, if demands from the students to be involved are getting strong also business reasons will drive the balance to the left in Figure 3.

We would assume that educational institutions will justify their data processing either by consent or legitimate interests, e.g., stated in a contract. What are the limits to using legitimate interests, and are there any reasons related to LA that would speak against consent as a default justification for collecting data from learning activities?
In terms of legitimate interests, Cormack (2016a) sums up how European law specifies requirements for this justification to be used:

Where personal data are processed for legitimate interests, there must be a clearly stated purpose, the processing must be necessary for that purpose, the impact and risk for the individuals whose data are processed must be minimised, and any remaining impact or risk must be justified by a balancing test against the claimed interest. Interests, even though legitimate, cannot justify processing that involves an inappropriate risk to the individuals whose data are processed. (Cormack, 2016a)

The schools and universities need to know what they want to achieve with data analysis, otherwise they do not pass the ‘necessity’ test: information that is not necessary for the declared purposes should not be collected (Cormack, 2016). And there is no way out for the institutions to turn to the students and ask for a blanket acceptance of collecting data. The students need to know what they are asked about, to be able to balance the benefits and risks of the proposal. Data-driven techniques, where the ideas of possible interventions first appear after the data are collected and processed do not give much in terms of specific purpose descriptions for justifying the process before it is started.

The students need to be actively involved, as we see when LA is set up to personalise learning. Cormack makes it clear that legitimate interests cannot be used to justify any activity where the intention is to personalise a service or otherwise affect individual users, “since this would contradict the requirement that the impact on individuals be minimised” (Cormack, 2016a).

Once the organisation has identified patterns in data that enable it to identify and design such an intervention, however, it should also have sufficient information to seek valid consent from those individuals who may be affected by it. Whereas at the time the data were collected the results of data-driven analysis and their consequences could not be foreseen or explained to individuals, now they can. Consent can now be fully informed. Offering a choice between personalised and generic versions of the service should increase the likelihood that consent to personalisation is freely given. (Cormack, 2016a)

The constraints of the law and the intrinsic qualities of data-driven practices that LA is part of seem to drive LA implementers towards what Sclater (2017) has called a hybrid approach: using legitimate interests for analysis and consent for intervention. Cormack (2016a, 2016b) has argued that the solution, which came up in the discussion of consent related to the developed of GDPR, termed “downstream” consent should be applied: “consent can also be requested ‘downstream’, when the purpose of the processing changes” (Article 29 Data Protection Working Party, 2011). Upstream there is the analysis of the data, trying to identify patterns; downstream are the interventions to be taken when one knows what the problem is, it is still not acted upon, and one is able to communicate clearly to the student options that the student could agree to.

The approach proposed by Cormack (2016a, 2016b) dividing the monolithic ‘big data’ process into two stages (analysis to find patterns; and intervention to identify and affect relevant individuals) opens up a need for examining the educational specific consequences and opportunities when applied on LA. This is the focus of the last part of this paper.

5. Pedagogical Opportunities arising from LA Data Privacy

In theory, a separation of LA into two processes, analysis and intervention seems simple. Analysis justified by legitimate interests is the prerogative of the institution; students are first involved when clear actions can be outlined with opt-in and opt-out options to consent to. To perceive this as two distinct processes with no overlapping stakeholders and no interfering sub-processes seems to be too far from real life in education. What about the teachers, are they part of the analysis process? What about access to data? Does data for analysis come only from institutional systems, like Student Information Systems, Learning Management Systems? How does the institution get access to data from non-institutional and informal learning settings, e.g., mobile and cloud learning platforms outside the control of the school or university, social media, other sensor data relevant for learning?

Contrasting the hybrid model of analysis and intervention with the LA process model developed by ISO/IEC JTC1/SC36 (Figure 4) we see that three important sub-processes precede the analysis stage. In order to be able
To do analysis one needs to decide upon which learning activities to monitor; to collect the data that serve as proxies for the activities under study; and as a last important step also imbued with a host of privacy and data protection issues, decide upon how the data should be stored and processed before analysis.

![LA process model developed by ISO/IEC JTC1/SC36 (ISO, 2016)](image)

**Figure 4.** LA process model developed by ISO/IEC JTC1/SC36 (ISO, 2016)

To exclude these introductory processes from exchange with the students under the pretext that the analysis of these activities is within the legitimate interest of the institution feels strange. On the other hand, it might well be that conversations about what is going on prior to and during analysis are part of activities that are crossing different professional and educational discourses with associated norm sets. Learning analytics is different from traditional academic analytics, which does not aim at actionable insights feeding back to the individual learner. Therefore, analysis cannot only be an administrative task, or a pure research activity. And with teachers on board doing analysis, this is definitely also a pedagogical activity, which involved the learners. To see how it involves learners, and how it is different from intervention, we first need to look at what characterises intervention.

It is the risks to the learner, caused by the institution acting upon the knowledge from analysis that make it necessary to ask the learner to consent to processing of personal data, giving him or her the opportunity to opt out when the nature of the proposed intervention is clarified. Even if these deliberations have a legal flair to them, they are mainly of pedagogical nature. The worst scenario from a student’s perspective is probably illegal: that predictive profiling could be subject to automated processing leading for example to exclusion (Hoel & Chen, 2016). Most likely, interventions would be to present the learner with different alerts and prods (e-mails or messages); visualisations showing progress, position relative to different student cohorts, etc.; and recommendations for what to read next, what tests to take next, etc. Some of these interventions will be executed by machine, but most likely the majority will involve interaction between the students and the teachers.

In conclusion, both analysis (and the preceding sub-processes) and intervention will involve extensive interaction with the students around topics that are mainly related to their learning tasks. We have difficulties seeing that questions of data access and handling are dealt with inside a secluded administrative and research logic without involvement of students and teachers, and their virtual learning agents. That being said, we see the value of keeping the separation between analytical and intervention concerns, being forced to execute the balancing test, weighing the benefits and risks of collecting and processing personal data. We believe that different normative models could live side by side. The legal model tells you to wait asking for consent until the individual has a chance to make an informed choice based on alternative proposals for intervention. The research model tells you to ask for permission to gather information, follow the fair processing principles, and keep the data safe. The administrative model tells you to use anonymised aggregated data and follow strict legal procedures when dealing with personal information. Most importantly, the pedagogical model tells you to support the student’s own learning and use every opportunity to enhance the learning experience by bringing in relevant tasks and material. Data for learning analytics is such an opportunity.

### 6. Conclusions

We introduced this study with the challenges faced by an international group of standards experts trying to motivate international norms for privacy and data protection in the context of learning analytics. How do you find a common ground for policy development when we have countries where all learning activity data seem to be available for analysis (e.g., China), and countries that are reluctant to allow library data to be analysed because of privacy issues, and questions whether learning analytics is legal in the first place (e.g., Norway)
(Hoel & Chen, 2017; Hoel, Chen, & Griffiths, 2017)? It would help to build consensus about privacy and data protection policies if these also could be argued from an educational perspective, not only from universal or individual rights perspective.

In this paper we have demonstrated that privacy as a ‘hypernorm’ yields when pressured by corporate, commercial, or national security interests. Likewise, consent as general justification for collection and processing of personal data is not applicable in an educational setting unless the process is carefully staged, separating analysis from intervention. The discussion of justifications for accessing and processing learning activity data has shown that we from the very beginning are within a space of negotiations using a variety of justifications based on ethics, law, national policies, and pedagogies. Therefore, we should in LA make an effort of making the pedagogical justification for privacy policies more explicit.

We would claim that there is scope for international consensus on educational privacy policies, based on the following principles:

1. Privacy and data protection in learning analytics is achieved by negotiating data sharing with each student.
2. Openness and transparency is essential. How the educational institution will use the data and act upon the insights of analysis should be clarified in close dialogue with the students.
3. Big data is a field of study that will impact all society; and therefore, in the context of learning analytics, this subject field should be used as an opportunity to engage students when negotiating privacy and data protection policies.

These principles have strong grounding in discourse and practice on privacy. The first principle is in accord with the theory of contextual integrity proposed by Nissenbaum (2014).

The theory of contextual integrity is a theory of privacy with respect to personal information because it posits that informational norms model privacy expectations; it asserts that when we find people reacting with surprise, annoyance, indignation, and protest that their privacy has been compromised, we will find that informational norms have been contravened, that contextual integrity has been violated. (Nissenbaum, 2014, p. 25).

The second principle is in accordance with the best practice guidelines we now see published by educational institutions informing about how LA will be implemented (Sclater, 2016; Open University UK, 2014).

The third principle connects to the discussion on 21st century skills and competences for new millennium learners (Ananiadou & Claro, 2009). In Norwegian education, for more than a decade digital literacy has been defined as one of the central competences needed in the future, and the ability to use digital tools was defined as a basic skill (Selton-Green et al. 2009; Krumsvik 2008, 2009). Understanding how student data are used is part of digital literacy.

We would suggest that these principles are further developed and expressed in a LA privacy maxim for education. In the future, we will extend our analysis to gather more data on how different countries develop privacy policies in education. This paper has shown that there is a need to understand how data privacy policies for LA connects to pedagogical practices. This educational argumentative space will be subject to further studies.

7. References


