Learning Analytics Community Exchange

Is Privacy a Show-stopper for Learning Analytics? A Review of Current Issues and Solutions

Learning Analytics Review 6

ISSN: 2057-7494

By: Dai Griffiths, University of Bolton, UK., Hendrik Drachsler, Open University, The Netherlands, Michael Kickmeier-Rust, TU Graz, Austria; Christina Steiner, TU Graz, Austria, Tore Hoel, Oslo and Akershus University College of Applied Sciences, Norway, Wolfgang Greller, Vienna University of Education, Austria

Published: 15 February 2016

Keywords: learning analytics, privacy, ethics, data protection, data ownership.
1. Executive Summary...................................................................................... Error! Bookmark not defined.
2. Introduction .................................................................................................. 1
3. What are the dangers of Learning Analytics? ............................................. 2
   3.1. Schools ................................................................................................... 3
   3.2. Universities .......................................................................................... 4
   3.3. Workplace ............................................................................................. 5
4. Ethics and Privacy in Learning Analytics .................................................... 5
   4.1. What is Ethics? ...................................................................................... 6
   4.2. What is Privacy? ................................................................................... 10
5. Useful tools to take actions on privacy conform Learning Analytics .......... 17
   5.1. JISC - Code of practice for Learning Analytics ..................................... 17
   5.2. OU UK - Policy on Ethical Use of Student Data for Learning Analytics ......... 18
   5.3. LACE - DELICATE Checklist ................................................................. 19
6. Conclusions ................................................................................................... 21
   6.1. A fast-developing environment ............................................................... 21
   6.2. Institutional strategy in response to an environment in flux .................. 23
7. References .................................................................................................... 24
8. About ............................................................................................................ 30
1. Introduction

Learning analytics is a new field, and as with any other new field there is work to be done to apply existing ethical approaches to the new ethical quandaries which are generated. Many of the people involved are unused to wrestling with questions of ethics or privacy that are associated with the field, and are unaware of how to set about answering them. We, therefore, start by providing some illustrative examples of controversial practices or privacy issues from the different educational sectors. We continue to offer a definition what constitutes the concepts of ethics and privacy situated in the current literature. One might hope that having identified the problems, and reviewed ethical and privacy approaches, then, any problems could be resolved by carrying out the work of applying the latter to the former. The problems which learning analytics throws up, however, are particularly intractable, and resist this simple approach.

Firstly, the technical environment in education is increasingly complex. Until recently educational computing infrastructure was dominated by Virtual Learning Platforms (VLEs) and desktop applications, linked in a very attenuated way to institutional silos of student information held primarily by the student records system and library software. This landscape is rapidly changing. The infrastructure is becoming more varied and cloud-based, with most students and staff now making frequent use in their work and study time of technology that they keep in their pockets or around their wrists, or even hidden in everyday objects they use to learn, work, play or just relax. Users have become accustomed to more personalised experiences and services on the web, that are based on large amounts of data being analysed and fed back to them. Sometimes they are in control and manage data for fun use data collection gadgets for their own observational purposes, as when they measure how they slept last night or how many steps they ran before breakfast. These more varied sources of data are potentially available to institutions, and it seems only a matter of time before they are merged with the mass of data which has resulted from connecting together institutional silos and teaching systems.

Secondly, because of the range of data sources involved, learning analytics is enmeshed with multiple personal and societal issues across the range of educational activities, and partakes of the ethical and privacy questions which those issues raise. Moreover, the changes brought about by learning analytics are seen differently from various perspectives, and valued differently by the range of stakeholders involved. The ethical and privacy issues raised by learning analytics, therefore, mesh closely with wider and long-standing debates about the nature of education.

Thirdly, the rise of learning analytics is not presented simply as a more effective way to carry out educational activities, but also as a means to transform the context in which the new methods are embedded. Indeed, the use of data is currently changing both educational practice (in designed interventions as well as through the unintentional consequences of technical and managerial changes), and the wider social context. As a consequence, the frameworks of ethics and privacy which we might hope to use to assess the acceptability of learning analytics are shifting under our feet.

Given the complexity of this panorama, we do not aspire to providing definitive answers. Rather we direct the attention of the reader to the results of current work which can help those responsible to grapple with the issues. Firstly, we explore the range of questions that need to be addressed, in a
discussion informed by the coordination activities of the LACE project. Secondly, we provide an overview of frameworks and approaches that can be applied and establish what we understand as Trusted Learning Analytics that takes into account the current privacy regulations of the European Union. We conclude by identifying some of the outstanding inconsistencies between learning analytics practice and the proposed frameworks, and suggesting how institutions can respond to these.

This paper arises from the issues identified at a number of expert workshops that have been organised to address ethics and privacy in relation to Learning Analytics. It has been written with contributions from partners in the LACE and LEA’s Box projects. At LAK15, the first workshop on Ethics and Privacy in Learning Analytics (#EP4LA) was organised jointly by the EU FP7 project Learning Analytics Community Exchange (LACE) and the SURF SIG Learning Analytics, who also organised similar events at other conferences in the Netherlands (Utrecht), US, (Washington), and France (Paris). Within this review, we summarise the information and insights gained from these intensive studies into ethics and privacy for learning analytics. Its results are intended to be useful to educational policy makers, and decision makers at all levels in education.

2. What are the dangers of Learning Analytics?

It is not possible to assess the potential dangers of learning analytics without first having an idea of its scope. Reasonably enough, the scope has usually been defined in terms of learning or education. Writing in the Educause Review, Barneveld et al. (2012) distinguish between, on the one hand:

**Academic Analytics**: a process for providing higher education institutions with the data necessary to support operational and financial decision making (adapted from Goldstein and Katz).

and, on the other hand,

**Learning Analytics**: The use of analytic techniques to help target instructional, curricular, and support resources to reach specific learning goals (adapted from Bach).

Learning analytics is here identified as being on the side of the educational angels, by working to support learning rather than being enmeshed with the ethical ambiguities of finance and operational issues which cloud the running of educational institutions. By constraining the scope of the field to activities that are seen to be inherently good, these definitions can be seen as attempting to ward off ethical problems. The problem with this definition is that many learning analytics applications do not address ‘specific learning goals’, but rather seek to help institutions and students achieve such broad outcomes as ‘engagement’ or ‘retention’.

Perhaps this is why the recent Jisc Code of Practice for Learning Analytics uses a wider definition of using data about students and their activities “to help institutions understand and improve educational processes, and provide better support to learners” (Sclater 2014b). In this case, we may wonder what the scope of ‘educational process’ is. Could it, for example, include recruitment and marketing? Is there anything that an educational institution does that is not an ‘educational process’? From this perspective education is a whole, and its processes cannot be isolated in
hermetically sealed areas. So, our first concern about learning analytics is that it inevitably partakes of the ethical ambiguity of the educational system as a whole.

Secondly, as we comment in the introduction, learning analytics is not simply at the service of existing educational institutions and educational methods. Many enthusiasts for learning analytics see the approach as a way of achieving transformational change in education. In this case, the ethics of learning analytics becomes entangled with the ethical merits of the changes which are planned, and also the unplanned consequences of those changes.

Thirdly, many learning analytics applications are opportunistic, making use of the opportunities presented by bringing data together in ways which were not anticipated by those who decided to collect that data in the first place. Indeed, the principal driver behind the Big Data movement is the establishment of connections between previously separate datasets in order to gain new insights. This strategy constitutes a challenge to the existing frameworks for regulating the use of personal data. Learning analytics can be seen as a special case of this trend, and also a particularly sensitive case, given the role of education in child-care and in determining success in employment.

In summary, once the Pandora's Box of data availability has been opened, then individuals lose control of the data about them that have been harvested. They are unable to specify who has access to the data, and for what purpose, and may not be confident that the changes to the education system which result from learning analytics will be desirable. More generally, the lack of transparency in data collection and analysis exacerbates the fear of undermining privacy and personal information rights in society beyond the confines of education. The transport of data from one context to another can result in an unfair and unjustified discrimination against ban individual.

To give an indication of what ethics and privacy issues are, we give below some glimpses of past controversies and posit hypothetical cases from various sectors including school, university, workplace and informal learning.

2.1. Schools

The inBloom case, 2013, USA: inBloom tried to address the problem of managing the flood of educational data that is generated in each individual school by making use of the resources of a larger consortium to offer cheaper and more efficient data processing. The white paper which set out the inBloom plan is no longer available but is summarised in (Campbell 2013). According to a sympathetic article in The Economist (The Economist 2014), the inBloom system applied a “standardised data format that gave schools complete control over what data they collected, how they were used and with whom those data were shared”. A non-profit organisation was established to manage inBloom in a transparent way and it attracted over $100 million from the Gates and Carnegie foundations. Many state governments and school managers were enthusiastic about the inBloom system and joined the consortium. However, parents’ organisations and privacy advocates were not invited to participate in the development of the system. These groups had major doubts about the privacy and security of the data about America’s pupils within the inBloom system. They feared the misuse of data, repurposing of the data for commercial interests, and did not trust the secure storage of the data after other large organisations suffered a series of cyber-attacks. The lack of representation of these stakeholders proved to be a fatal mistake, as the pressure they created rose to such a level that many consortium partners of inBloom had to pull back due to intensive
protests. Eventually, in one of the most high profile examples of the impact of privacy concerns on learning analytics, inBloom was dramatically closed down on April 21, 2014 (Kharif 2014).

**Stichting Snappet case, 2014, The Netherlands:** Public schools rented tablets with pre-installed apps for learning mathematics and languages from the educational association Stichting Snappet. By July 2014, more than 400 basic schools with children aged 7 to 9 years were using these tablets conducting more than 1.5 million operations per day. Snappet used the data from the system to classify and predict individual students’ success and provided the schools with that information to suggest educational interventions. This raised an issue at the Dutch Data Protection Authorities (CBP) that started an investigation into the activities of Snappet in 2013. In the final report of the CBP, Snappet was recognised as the legal entity responsible for the collected data. Further the collected data were identified as ‘personal data’ that needed to be treated according to the Dutch law for the protection of personal information. Moreover the CBP classified the data as ‘highly sensitive’, because they concerned young children and their individual performance at school. Although Stichting Snappet had informed the schools about the usage of the data, the Data Protection Authorities saw this information as too narrow and incomplete. Further, Snappet should have obtained the informed consent of the parents before processing the children’s data in the way they did. The detailed CBP report in Dutch can be found online (College Bescherming Persoonsgegevens 2014). In contrast to inBloom, Stichting Snappet was not shut down, as they addressed the privacy concerns of the CBP and so were able to continue their services in the boundaries of the Dutch law.

### 2.2. Universities

Students share more information than ever before, mainly through the use of social media. They are also increasingly monitored through the digital traces they leave behind when entering buildings, lending books from the library, accessing the university system via the web, and by surveillance cameras scanning every step of their physical learning path. This has interesting ethical consequences for the parties involved, especially when these data-driven changes cause a shift in established patterns of power, as between a student and a professor. Consider the following hypothetical scenario.

*A student enrolls in a costly university course with fulsome promises of success and engagement with their customers (as they refer to their students). Having failed the course the student complains that they did not receive any warning of impending failure from tutors or systems, even though the lecturer had seen the signals buried in the activity data relating to the student on the teacher’s dashboard. The student could have been warned, but was not.*

This scenario may seem exaggerated, but two current trends suggest that it may not be far-fetched. Firstly, according to the Times Higher Education Supplement “Universities are misleading prospective students by deploying selective data, flattering comparisons and even outright falsehoods in their undergraduate prospectuses” (Mathews 2014). Secondly, students who are paying increasingly large sums to purchase those services, and are starting to see this transaction not

---

1 https://nl.snappet.org/
as purchasing the opportunity to study, but rather expect delivery of services to a specified standard. At present, for example, Mälardalen University College is being sued by a US student whose course allegedly did not match the level of quality promised (Myklebust 2015). Other lecturers report that students are starting to see payment of fees as a contract to obtain a qualification in exchange for money and the performance of a set of activities (Anonymous Academic 2015). When these two trends come together, then the scenario we have set out becomes quite reasonable. To our knowledge no student has ever taken legal action of this sort in relation to learning analytics, but the scenario illustrates the ethical issues raised by not acting upon the information that one has access to. What are the ethical implications of knowing, of not knowing, and of refusing to know, in the age of Big Data?

2.3. Workplace

In today’s workplace online and just-in-time learning is a normal part of work for most employees. The computer or mobile device and the network are provided by the employer. Does that mean that everything the employee does should be monitored? Many people might answer ‘no’ to this question, following the line of reasoning behind the following scenario, taken from a web post entitled ‘LittleBrother is watching you’ by Miriam Schulman.

An employee writes a note to her boyfriend. She puts it in an envelope, affixes her own stamp, and drops it in the basket where outgoing mail is collected. The fact that the pencil and paper she used belong to her employer does not give her boss the right to open and read this letter. (Schulman 2016)

However, citing David Sobel of the US based Electronic Privacy Information Center, Schulman argues that “this is not the view taken by employers, who claim that because they own the computer, they have the right to read the e-mail it produces”. Moreover, many employers take steps to be able to monitor all of their worker’s Internet use. Employers can also buy software that enables them to rate their employees on the basis of their browsing history crosslinked with a database of thousands of web sites categorised as ‘productive’, ‘unproductive’ or ‘neutral’. In the case of Myrna Arias, her employer insisted on installing the Xora² app on her mobile phone which tracked her location 24 hours a day. When she uninstalled the app she was fired, and a law suit resulted (Superior Court of the State of California 2015). Recently, the European Court of Human Rights has confirmed the rights of employers to monitor private messages at work when using office utilities and accounts (Rawlinson 2016). Readers interested in these issues may wish to visit the Privacy Rights Clearinghouse³ which includes detailed fact sheets exploring privacy issues between employers and employees.

3. Ethics and Privacy in Learning Analytics

Learners today have access to a multitude of learning tools, applications, and resources, and they carry out learning activities in virtual or simulated environments, while connecting to others through social media. While educational institutions have always analysed the data of their students to some extent, the unprecedented quantity of data generated by these new practices has changed the

² Xora is now marketed as StreetSmart by ClickSoftware: http://www.clicksoftware.com/small-medium-business-streetsmart
nature of the analysis which can be carried out. The flood of data can now be captured and the multi-faceted learning processes that generate them can (potentially) be analysed using big-data analytics techniques (Pardo & Siemens 2014). As the methods for carrying out this analysis evolve, and new applications are deployed, an increasing number of ethical and privacy issues arise. For example, the availability of hardware sensors and biometric and location technologies enables a multi-faceted tracking of learners’ activities, movements and other parameters, so that more and more data can potentially be collected about individuals, who are often not even aware of its collection. Data collection and use under such circumstances is widely seen as ethically and legally questionable (Greller & Drachsler 2012). Ethical and privacy issues in learning analytics include the conditions for the collection or aggregation of data, informed consent, de-identification of data, transparency, data security, interpretation of data, as well as data classification and management (Slade & Prinsloo 2013). There is a need to develop a clear and agreed set of principles and guidelines with respect to the ownership of data and analytic models, rights and responsibilities (Ferguson 2012). At the moment, there are no standard methods and generally agreed procedures for informed consent, opting out etc. There is increasing recognition of the need for a shared understanding of the nature of agreement for data to be used, and a shared code of ethics to appropriately deal with the topics of ethics, privacy and learning analytics, as argued by Berg (2013) and recently shown by the LACE ‘Visions of the Future’ report (Griffiths et al. 2016).

The acceptability of a learning analytics application in terms of privacy and ethics is closely related to trust, access and accountability (Pardo & Siemens 2014). In these areas policy, technology and practice are developing more rapidly than the legal framework which is intended to regulate them, and the slow pace of law-making seems unable to redress this gap. Nevertheless, existing approaches of dealing with ethics in learning analytics commonly ground their discussion within and relating to legalities and legal understanding of privacy, for lack of any other reference point (Willis 2014). The following sections provide a definition for Ethics and Privacy and an overview of relevant work on ethics and privacy issues in learning analytics.

3.1. What is Ethics?

Ethics is a moral code of norms and conventions that exists in society externally to a person, and the understanding of what constitutes ethical behaviour varies strongly over time and between cultures. Privacy, on the other hand, is an intrinsic part of a person’s identity and integrity. Research ethics have become a pressing topic in recent years, first and foremost arising from discussions around codes of conduct in the biomedical sciences such as the human genome, but also, more recently, in the shape of ‘responsible research and innovation’ (RRI) which is being promoted by the European Commission 4.

The first basic written principles for ethical research originated from the Nuremberg trials in 1949, and were used to convict leading Nazi medics for their atrocities during the Second World War (Kay et al. 2012). The basic principles derived from the development of research ethics since the Nuremberg Code (cf. Kay et al. 2012) can be summarised as:

- Voluntary participation in research;

Informed consent of the participants, and, with respect to minors, the informed consent of their parents or guardians;
Experimental results are for the larger good of society;
Not putting participants in situations where they might be at risk of harm (either physical or psychological) as a result of participation in the research;
Protected privacy and confidentiality of the information;
Option to opt-out

A recent example of an ethical debate triggered by a Big Data experiment was the Facebook contagion study (Kramer et al. 2014), where a team of researchers manipulated the newsfeed of over 650,000 Facebook users without notification or informed consent. There was a massive negative reaction to this manipulation among the user community and beyond. However, ethics is a volatile human made concept and since the Facebook episode researchers have been discussing the pros and cons of the study. Some argue that the study was indeed unethical, but that it also contributed new insights into human behaviour (Kleinsman & Buckley 2015).

3.1.1. Ethical issues and dilemmas in learning analytics
Relevant ethical issues and dilemmas in learning analytics can be summarised and grouped into the following overlapping areas (Campbell et al. 2007; Pardo & Siemens 2014; Sclater 2014a; Slade & Prinsloo 2013; Willis 2014)

- **Privacy**: Users may be concerned at the possibility that their actions and personal data are being tracked. But, users are overwhelmed in trying to control all their data streams and are often unaware of the amount of data being collected, linked and exchanged when using technology services.
- **Informed consent, transparency, and de-identification of data**: Informed consent refers to the requirement for an individual to give consent for the collection and analysis of the data which they generate. Transparency refers to the degree to which users can observe the ways in which the data they generate is used. De-identification refers to the processes which seek to make it impossible to identify the person who generated a particular piece of data.
- **Location and interpretation of data**: Learning activities today are usually spread over various tools and physical servers located in unknown locations around the world. Learning analytics aims at bringing together these different data sources to create a more complete picture of learning. Questions arise on the implications of using multiple and non-institutional sources. Who is responsible for determining which data represents a particular student, and what is the student’s own role in this?
- **Data management and security**: This area relates to the criteria to be applied in deciding questions of data storage, curation, access rights, and the measures and level of data protection provided. It also involves the issue of how long data should be stored and whether it is accessible by different data clients.
- **Data ownership**: Ownership concerns not only the data collected, but also the analytics used, and the analytics output. It also concerns outsourcing and data transfers to third parties and the responsibilities and liabilities which arise in this process.
- **Possibility of error**: Analytics results are always based on the data available and the outputs and predictions obtained may be imperfect or incorrect. Questions arise about who is
responsible for the consequences of an error, which may include ineffective or misdirected educational interventions.

- **Role of knowing and obligation to act**: Learning analytics brings new knowledge and insights about learning. Does the new knowledge gained bring with it a responsibility to act upon it? What are the ramifications of action or inaction?

### 3.1.2. Ethical Frameworks in Learning Analytics

Although ethical and privacy issues are discussed in the research literature, there are only a few coherent approaches that elaborate ethical challenges in more detail and attempt to define an ethical framework to guide institutions, researchers and developers in the application of learning analytics (Slade & Prinsloo 2013). One possible approach to elaborating the ethical issues of learning analytics is to analyse the risks of implementing a learning analytics project and to consider the management of those risks. (Stiles 2012) identifies a set of specific areas and associated risks. Greller and Drachsler (2012) have considered ethical and legal aspects in their framework for learning analytics under the dimension of ‘external constraints’. Apart from ethical, legal, and social constraints, they also consider organisational, managerial, and process constraints as relevant components on this dimension. These external limitations can be categorised into conventions, including ethics, personal privacy, and other socially motivated constraints, and norms which imply restrictions by law or mandated standards and policies. This recognises that there is a distinction, but also a close linkage, between ethics and the existing normative legal regulations: ethics deals with which measures are morally allowable; the law defines what is allowed without legal consequences (Berg 2013). In many cases ethical issues are reflected in legislation, but ethical considerations go beyond what is set in laws and depend on ideological assumptions (Slade & Prinsloo 2013) or simply on boundary negotiations with the surrounding social factors (e.g. neighbours, club members, etc.). Many legal regulations are based on ethics and derived moral codes, while in some situations an ethical position needs to be applied to interpret the law (N. Sclater & Bailey 2015). Moreover Kay et al. (2012) highlight that the mission and responsibilities of education require that “broad ethical considerations are crucial regardless of the compulsion in law” (p. 20).

Kay et al. (2012) also argue that learning analytics will experience conflict between assuring educational benefits, maintaining business interests of and withstanding competitive pressure on educational institutions, and meeting the expectations of users whose lives are immersed in digital technologies. They postulate four key principles for good practice with respect to ethical aspects and analytics when dealing with these conflicts:

- **Transparency**: definition of purpose, scope and boundaries
- **Trust**: consideration of interests and feelings of the data subject
- **Choice and consent**: information and opportunity to opt-out or opt-in
- **Consequence and complaint**: acknowledging the possibility of unforeseen consequences and mechanisms for complaint

Willis et al. (2013) refer to this area of conflict and argue that there is a need to balance the information management imperatives generated by faculty expectations, privacy protection laws, and the education institution’s philosophy of student development. They do not define specific
guidelines on different ethical issues, but suggest using the Potter Box (Potter 1972) an ethical framework commonly applied in business communications, to deal with ethical dilemmas of analytics. This approach, however, only provides a thinking framework for analysing a situation but does not provide a criteria for resolving ethical dilemmas.

Slade and Prinsloo (2013) take a socio-critical perspective, proposing a framework of six principles to address ethics and privacy challenges in learning analytics. The following bullet points provide the names of the six principles, followed by a summary of each.

- **Learning analytics as a moral practice**: Focus should not only be put on what is effective, but on supporting decisions on what is appropriate and morally necessary. The ultimate goal is understanding, not measuring.
- **Students as agents**: Students should be involved in the learning analytics process as collaborators and co-interpreters. A student-centric approach to learning analytics is recommended.
- **Student identity and performance are temporal dynamic constructs**: The dynamism of data is acknowledged, thus providing only a limited snapshot of a learner at a particular point in time in a particular context.
- **Student success is a complex and multidimensional phenomenon**: Learning progress and success consists of multidimensional, interdependent interactions and activities. The data used in learning analytics is incomplete and analytics may lead to misinterpretation or bias.
- **Transparency**: Information about the purpose of data usage, data controllers/processors, and measures to protect the data should be provided. We believe there is another type of transparency, which is more pedagogical: i.e. Information about success indicators and benchmarks. Every course curriculum now has to specify the assessment criteria (e.g. physical attendance, continuous or summative assessment, minimum criteria to pass, etc.)
- **(Higher) education cannot afford not to use data**: Information that learning analytics may provide should not be ignored by an educational institution.

Pardo and Siemens (Pardo & Siemens 2014) have analysed ethical and privacy issues in learning analytics research in educational institutions and have also taken into account how privacy and ethics are addressed in other contexts. They identify a set of four principles that aggregate numerous issues and are intended to serve as a basis for setting up appropriate mechanisms for meeting ethical and legal requirements when developing and deploying learning analytics. When applying these principles, this needs to be done in due consideration of legal and social requirements. We again provide the names of the principles with a short summary for each:

- **Transparency**: All stakeholder groups in learning analytics, i.e. learners, teachers, educational administrators, should be provided with information on what type of data is collected and how it is processed and stored.
- **Right to access**: Security of data needs to be guaranteed. Access rights need to be clearly defined for a data set.
- **Student control over data**: This refers to giving users the right of users to access the data collected about them and, if necessary, to correct it.
• **Accountability and assessment:** The analytics process should be reviewed and for each aspect of the learning analytics scenario the responsible entities should be identified.

Ethical issues in learning analytics may also be considered in the context of the history of Internet research ethics, where there has been an attempt to find a balance between harm to the individual and greater scientific knowledge (Slade & Prinsloo 2013). The Association of Internet Researchers provides a set of ethical guidelines for decision making about internet research (Ess 2002; Markham & Buchanan 2012). These are aimed at providing researchers with a basis for conducting their research in an ethical and professional manner, and the guidelines have also been indicated by learning analytics researchers as a valuable source for dealing with privacy issues in the application of learning analytics.

### 3.1.3. Ethics by Design

Since learning analytics involves technology, ethics and privacy aspects concerns should not be considered from a purely legal perspective, but also need to be addressed from a technological point of view (Pardo & Siemens 2014). One way of ensuring this is to take privacy and ethics into account during the design process of learning analytics tools. This approach is referred to as ‘privacy by design’, ‘value-sensitive design’ or ‘ethics by design’ and it has been started to be acknowledged and taken up in learning analytics research, for example (Bomas 2014; Scheffel et al. 2014). Value-sensitive design, or ethics by design, involves addressing ethical and legal requirements and considerations during the design and development process, so as to incorporate them into the software being created (Friedman 1997). Design principles and guidelines are applied with the intention of ensuring that the software follows ethical rules, or supports humans in doing so (Gotterbarn et al. 1997; Gotterbarn 1999).

### 3.2. What is Privacy?

The right to privacy is a basic human right and an established element of the legal systems in developed countries. Already in the nineteenth century, Warren and Brandeis (1890) wrote an article on ‘The Right to Privacy’, in which they proposed that privacy is the ‘right to be let alone’, and focused on the protection of individuals. This right is often debated in the context of the popular press with regards to royals and celebrities. The concept of privacy as the right to be let alone was further developed by Westin (1968) who made it clear that new technologies change the balance of power between privacy and societal technologies. From this, Westin went on to specify privacy as the ‘right of informational self-determination’ and as a vital part for restricting government surveillance in order to protect democratic processes. According to Westin [ibid.], each individual is continually engaged in a personal adjustment process in which they balance the desire for privacy with the desire for disclosure and interaction with environmental conditions and social norms. Flaherty (1989) took the informational self-determination further and claimed that networked computer systems pose a threat to privacy. He first specified 'data protection' as an aspect of privacy, which involves "the collection, use, and dissemination of personal information". This concept forms the foundation for fair information practices used by governments globally. Flaherty promoted the idea of privacy as information control. Roessler (2005) later operationalised the right to privacy across three dimensions:

1. **Informational privacy**
2. Decisional privacy
3. Local privacy.

It is important to note that privacy is not the same as anonymity or data security. They are related concepts that have an effect on privacy, but do not represent privacy as such.

3.2.1 State-of-the-art on Privacy in Learning Analytics

Privacy and ethics have evolved as an important and pressing topics not only in learning analytics, but also in analytics and big data, in general (Schwartz 2011; PMCA 2013). “Big data poses big privacy risks,” as (Tene & Polonetsky 2013, p.251) put it. Data has become resource of important economic and social value, and the vast amount of data now being shared, transmitted and accessed, opens up new and unanticipated uses of information when combined with new analytics methods. The collection of large and multifaceted data sets and the new possibilities of their use lead to growing privacy concerns among data subjects. The disclosure and use of personal data is increasingly associated with fear, uncertainty, or doubt (Dirndorfer Anderson & Gardiner 2014). Importantly, users’ concerns about privacy also differ according to the kind of data being collected, the context, and the perceived value of disclosing personal data (Pardo & Siemens 2014). In their article, Tene and Polonetsky (2013) elaborate on fundamental principles of privacy codes and legislation and argue that the principles of data minimisation and individual control and context need to be somewhat relaxed in a big data setting and considered not only from an individual but also a societal perspective (e.g. public health, environmental protection), while at the same time emphasizing transparency, access, and accuracy. The authors also discuss the distinction between identifiable and non-identifiable data and consider de-identification methods (anonymization, pseudonymization, encryption, key-coding) as an important measure for data protection and security.

Schwartz (2011) developed a set of ethical principles for analytics based on a series of interviews with experts in the field of data privacy, legislation, and analytics. These include a set of overarching ethical standards:

- Compliance with legal requirements
- Compliance with cultural and social norms
- Accountable measures tailored to identified risks
- Appropriate safeguards to protect the security of data
- Responsible limits on analytics in sensitive areas or with vulnerable groups.

In addition to specifying these generic principles, Schwartz argued that different ethical considerations are relevant at different stages of the analytics process. Accordingly, the rules regulating how these challenges are tackled need to be tailored to each analytics stage – always aiming at maximising the good results and minimising the bad outcomes for the people whose data is processed. In data collection, care needs to be taken about the kind of information gathered, in particular avoiding the collection of sensitive data. For data integration and analysis a sufficient data quality should be ensured and appropriate anonymisation should be carried out. In decision making it should be ensured that the analytics results on which decisions are based are reasonably accurate.
3.2.1. Privacy Guidelines and Models

In 1980, the OECD provided the first internationally agreed collection of privacy principles (OECD 2013b), aiming at harmonizing legislation on privacy and facilitating the international flow of data. The set of eight basic guidelines mirrored the principles earlier defined by the European Convention for the Protection of Individuals with Regard to the Automatic Processing of Personal Data and addressed (Levin & Nicholson 2005). The basic principles set out in the OECD Privacy Framework (OECD 2013b, p. 14-15) are summarised below:

- **Collection limitation**: There should be limits to the collection of personal data. Data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.
- **Data quality**: Personal data should be relevant to the purposes for which they are to be used, and to the extent necessary for those purposes. Data should be accurate, complete and kept up-to-date.
- **Purpose specification**: The purposes for which personal data are collected should be specified not later than at the time of data collection. Subsequent use should be limited to the fulfilment of those purposes or compatible purposes.
- **Use limitation**: Personal data should not be disclosed, made available or used for purposes other than specified – except with the consent of the data subject or by the authority of the law.
- **Security safeguards**: Personal data should be protected by reasonable security safeguards against loss or unauthorised access, destruction, use, modification, or disclosure.
- **Openness**: There should be a general policy of openness about developments, practices and policies with respect to personal data. Information on the existence and nature of personal data, purpose of their use, and the identity and location of the data controller should be available.
- **Individual participation**: Individuals should have the right to obtain confirmation of whether or not data relating to them is held and to have communicated to them the data, to be given reasons if a request is denied, and to challenge data relating to them and to have the data erased, rectified, completed or amended.
- **Accountability**: The data controller should be accountable for complying with measures which give effect to the above principles.

The OECD guidelines were not binding for OECD members, but have gained legal significance and served as a basis for privacy legislation in Europe (European Parliament 1995; Levin & Nicholson 2005; Spiekermann & Cranor 2009). The Guidelines were updated in 2013 (OECD 2013b). The revision keeps the original ‘Basic Principles’ of the guidelines, while modernising considerations on trans-border data flows and strengthening privacy enforcement. The updated guidelines focus on the practical implementation of privacy protection through an approach grounded in risk management. Furthermore, the need for greater efforts to address the global dimension of privacy through improved interoperability is acknowledged.

Currently, the OECD is working on privacy-related issues in the context of large-scale data use and analytics. In a preliminary report (OECD 2013a) on the broader topic of ‘data-driven innovation as a new source of growth’ different sectors of data use and analytics are elaborated (online...
advertisement, healthcare, utilities, logistics and transport, and public administration), however without any specific reference to learning or academic analytics. Privacy protection is indicated as one of several areas that needs public policies and practices to leverage the potential of big data. Privacy protection enabling open, secure, reliable, efficient, and also cross-border, flows of data on the one hand, and reducing privacy risks and enhancing responsible behaviour in the use of personal data is called for.

Based on the framework of the OECD Guidelines, the Federal Trade Commission of the United States have defined the Fair Information Practice Principles (FIPP), which specify concepts of fair information practice in the electronic marketplace. These cover five core principles of privacy protection, which many other guidelines and reports on fair information practice have in common, and are therefore relevant for information practice in dealing with personal information, in general (Federal Trade Commission 1998):

- **Notice/Awareness**: Users need to be informed before personal data is collected from them. Giving notice is necessary in order to enable the data subject to consciously decide whether he/she wants to disclose personal information, and to what extent. This principle is considered the most fundamental one, since the other principles are only meaningful provided that the user has noticed them.
- **Choice/Consent**: This principle refers to giving data subjects options as to how personal data collected from them may be used, e.g. secondary use. Thereby, traditionally two approaches may be taken, opt-in or opt-out.
- **Access/Participation**: This principle relates to giving users the possibility to access their data and to ensure that the data is accurate and complete.
- **Integrity/Security**: Data needs to be accurate and secure and appropriate steps and safeguards need to be taken to ensure that, e.g. using reliable data sources, cross-referencing multiple sources.
- **Enforcement/Redress**: To ensure compliance to privacy protection principles, there need to be enforcement and redress mechanisms through self-regulatory regimes, legislation creating private remedies for users, or government enforcement.

Privacy and data protection is regulated in national and international information legislation, which addresses the disclosure or misuse of information held on private individuals. Regulations started to appear in countries with the expanding spread and use of the Internet (Pardo & Siemens 2014). Examples include the European Union Directive on the protection of individuals with regard to processing of personal data and the free movement of such data (European Parliament 1995), the Canadian Personal Information Protection and Electronic Documents Act (Government of Canada 2004), the Australian Privacy Act and Regulation (Australian Government 1988), and the US Consumer Data Privacy in a Networked World (The White House 2012). The Family Educational Rights and Privacy Act (US Government 2004) is a US federal law that specifically applies to the protection of the privacy of student education records. This law allows the use of data on a need-to-know basis and provides parents certain rights to access to their children's education records.

In parallel with legislative efforts towards data protection, non-profit organisations have emerged that have the mission of defending user digital rights (Pardo & Siemens 2014), for example the ARGE
DATEN Privacy Service in Austria\(^5\) or the Electronic Frontier Foundation\(^6\) and Privacy Rights Clearinghouse in the US\(^7\). Paraphrasing from The European Agency for Fundamental Rights (2014), the following distinctions between data types are generally made in legislation:

- **Personal data:** information relating to data subjects who are identified or identifiable
- **Sensitive data:** a special category of personal data, i.e. data relating to natural persons concerning their racial or ethnic origin, political opinion, trade-union membership, religious or philosophical beliefs, data concerning health or sex life; this type of data is also called data deserving special protection
- **Anonymised data:** data that does no longer contain any identifying elements. In contrast, **pseudonymised** data contains identifiers, and the personal information is pseudonymised by replacing the identifiers by a pseudonym or by encryption

There is a general awareness of the importance and significance of data protection, and this is reflected in many national and international documents, where data protection is considered a fundamental right (Rodotà 2009). Nevertheless, “the right to data protection is not an absolute right; it must be balanced against other rights” (European Agency for Fundamental Rights 2014, p.21), i.e. it needs to be considered and implemented always in relation to its function in society. Providing a comprehensive description of the legislation initiatives on privacy and data protection of personal data is beyond the scope of this report (an overview and comparison between international privacy laws and approaches is given, for example, in (Levin & Nicholson 2005) and (Movius & Krup 2009).

### 3.2.2. European Regulations

The transfer of personal data between countries in the EU is necessary in day-to-day business of companies and public authorities. Since conflicting data protection regulations of different countries might complicate international data exchanges, the EU has established common rules for data protection (European Commission 2015). The application of this European legislation is monitored by national supervisory authorities. European data protection legislation considers the protection of personal data to be a fundamental right. Current EU law is the 1995 Data Protection Directive (European Parliament 1995), which applies to countries of the European Economic Area (EEA; i.e. all EU countries plus Iceland, Liechtenstein and Norway). The directive seeks to keep a balance between a high level of protection of individual privacy and the movement of personal data within the European Union. It applies to data that is collected and processed automatically (e.g. computer databases) and in non-automated ways (e.g. traditional paper files). This directive refers to the national law applicable and indicates that each Member State shall apply the national provisions it adopts pursuant to this Directive to the processing of personal data.

The Directive was extended by the ePrivacy Directive (European Commission 2002), which addresses the specific requirements for privacy and data protection in the context of information and communication technologies, and especially the internet and electronic messaging services. The Directive aims to ensure that users can trust the services and technologies they use for electronic

---


\(^6\) [https://www.eff.org/](https://www.eff.org/)

\(^7\) [http://www.privacyrights.org/](http://www.privacyrights.org/)
communication. The main regulations apply to spam, ensuring the user’s consent, and the installation of cookies.

### 3.2.3. European Data Protection Regulations under review

The European Commission’s 2012 proposal for the reform of data protection rules redefined personal data as “any information relating to an individual, whether it relates to his or her private, professional or public life.” It also states the right for people to “have easier access to their own data and be able to transfer personal data from one service provider to another more easily” (European Commission 2012) as well as a right to be forgotten. The European Commission is currently in process of establishing a reform of its data protection legislation, to enforce protection of personal data by updating and modernising data protection rules. It is planned that process should be concluded by the end of 2015, allowing for a formal adoption of the Directive in early 2016, to be followed by a two-year transitional period (EDPS 2015).

The final text of the regulations has not yet been published, but the policies promoted by the EU watchdog in this field, the European Data Protection Supervisor, may give an indication of the approach to be adopted. The supervisor sees a need to redress “the imbalance between innovation in the protection of personal data and its exploitation, making safeguards effective in our digitised society” (EDPS 2015, p.8) knowing there are data-driven technologies underway that will challenge the principles of data protection. When these technologies converge with artificial intelligence, natural language processing and biometric systems, empowering applications with machine-learning ability for advanced intelligence we will need accountable business practices, innovative engineering, empowered individuals and future-proofed data protection rules.

Two recommendations by the EDPS point to the current debate on ethics and privacy within LA. The first relates to consent for data use. It is clear that people do not feel they have control over the data they provide online. The Data Protection Eurobarometer survey in June 2015 (Eurobarometer 2015) shows that two-thirds of European citizens have this concern. Furthermore, more than six out of ten citizens do not trust online businesses; and seven out of ten are concerned that their information being used for a different purpose from the one it was collected for. One in seven say that their explicit approval should be required in all cases before their data is collected and processed (EDPS 2015). The Data Protection Supervisor proposes that “individuals should be able to exercise more effectively their rights with regard to any information which is able to identify or single them out, even if the information is considered ‘pseudonymised’” (EDPS 2015, p.5). The EPDS argues that “Consent is one possible legal basis for processing, but we need to prevent coercive tick boxes where there is no meaningful choice for the individual and where there is no need for data to be processed at all. We recommend enabling people to give broad or narrow consent, to clinical research for example, which is respected and which can be withdrawn”. However, the EPDS does not see consent as an easy way to resolve the current problems. (EDPS 2015, p.5). The EDPS is also of the opinion that the “EU should preserve, simplify and operationalise the established notion that personal data should only be used in ways compatible with the original purposes for collection” (EDPS 2015, p.5).

Another recommendation regards the algorithmic logic on which an action following analytics rests. The EDPS is of the opinion that the “reform should reverse the recent trend towards secret tracking...
and decision making on the basis of profiles hidden from the individual. The problem is not targeted advertising or the practice of profiling, but rather the lack of meaningful information about the algorithmic logic which develops these profiles and has an effect on the data subject.” (EDPS 2015, p.8).

Another body which is seeking to influence national policy on privacy is the World Economic Forum, which has published the report ‘Unlocking the Value of Personal Data: From Collection to Usage’. The report characterises evolving practice and perspectives in data management and summarises some the changes in a figure reproduced as Figure 1: New perspectives on the use of data below. (World Economic Forum 2013, p.7) . The new perspectives identified by the World Economic Forum usefully identifies some of the social and economic pressures which are bearing on data privacy, and which are making informed consent an increasingly difficult area.

![Figure 1: New perspectives on the use of data](image)

### 3.2.4. Privacy by Design

Privacy by design focuses on the engineering of systems that support or enforce privacy, and on the development of guidelines for designing and developing privacy-friendly systems (Cavoukian 2011)). Spiekermann and Cranor (2009) have carried out a privacy requirements analysis that is applicable for a wide variety of systems and identify system activities typically performed by information systems and their impact on user privacy (see Table 1: Information system activities and their impact on aspects of user privacy below for an overview). This impact depends on how the system activities are performed, what type of data is used and who uses it, and which privacy spheres are affected. Guidelines are provided on how notice, choice, and access can be implemented as fair information practices and users can be informed about them. Relating to these guidelines, in ethics by design a ‘privacy-by-policy’ approach (focus on implementation of notice and choice principles) and a ‘privacy-by-architecture’ approach (focus on minimizing collection of identifiable personal data and anonymisation) can be distinguished (Spiekermann & Cranor 2009).
System Activity | Relevant Aspects for User Privacy
--- | ---
Data transfer | • transparency on data transfers (within organisation, to third parties)  
• controlled transition of data  

Data storage | • protection from unauthorised access  
• transparency and control over personal data  
• awareness of data storage activities, persistent and transient storage  

Data processing | • awareness of transformation of data  
• information on secondary use of data  
• outsourcing of data for processing  

Table 1: Information system activities and their impact on aspects of user privacy. Source: (Spiekermann & Cranor 2009)

A related trend is Contextual Integrity (Nissenbaum 2009), a concept that has arisen in recent years to provide guidance on how to respond to conflicts between values and interests, and to provide a systematic setting for understanding privacy. It is not proposed as a full definition of privacy, but as a framework for evaluating the flow of information between agents (individuals and other entities) with a particular emphasis on explaining why certain patterns of flow provoke public outcry in the name of privacy (and why some do not). Contextual Integrity defines a context specified by roles, activities, norms, and values that interact with one another. The actors in this context are: senders, receivers, subjects and the attributes are data fields.

4. Useful tools to take actions on privacy conform Learning Analytics
In order to use educational data for Learning Analytics in an acceptable and compliant way, and to overcome the fears connected to data aggregation and processing, policies and guidelines need to be developed that protect the data from abuse and ensure trustworthy treatment.

4.1. JISC - Code of practice for Learning Analytics
Jisc describes itself as “the UK higher education, further education and skills sectors’ not-for-profit organisation for digital services and solutions” (Jisc n.d.). One of its principal initiatives is the Open Learning Analytics Architecture (Sclater 2015), which is the foundation for a service to be provided to the UK higher education sector. Jisc has decided that an essential first step in moving towards this service is to formulate a ‘Code of practice for learning analytics’ (Sclater 2014b; P. Sclater & Bailey 2015). This four page document aims to set out the responsibilities of educational institutions to ensure that learning analytics is carried out responsibly, appropriately and effectively. The following bullet points set out the eight points of the guidelines, together with the first sentence of the explanatory text, which summarises each point.

- **Responsibility**: “Institutions must decide who has overall responsibility for the legal, ethical and effective use of learning analytics.”
- **Transparency and consent**: “Institutions will define the objectives for the use of learning analytics, what data is necessary to achieve these objectives, and what is out of scope. ...”
- **Privacy**: Access to student data and analytics should be restricted to those identified by the institution as having a legitimate need to view them.”
• **Validity**: “It is vital that institutions monitor the quality, robustness and validity of their data and analytics processes in order to develop and maintain confidence in learning analytics and ensure it is used to the benefit of students.”

• **Access**: “Students should be able to access all learning analytics performed on their data in meaningful, accessible formats, and to obtain copies of this data in a portable digital format.”

• **Enabling positive interventions**: “Institutions should specify under which circumstances they believe they should intervene when analytics suggests that a student could benefit from additional support.”

• **Minimising adverse impacts**: “Institutions recognise that analytics can never give a complete picture of an individual’s learning and may sometimes ignore personal circumstances. Institutions will take steps to ensure that trends, norms, categorisation or any labelling of students do not bias staff, student or institutional perceptions and behaviours towards them, reinforce discriminatory attitudes or increase social power differentials.”

• **Stewardship of data**: “Data for learning analytics will comply with existing institutional data policies and the DPA (Data Protection Act)”

We discuss the privacy aspects of the Jisc guidelines in greater detail in section 7.1.

**4.2. OU UK - Policy on Ethical Use of Student Data for Learning Analytics**

The Open University of the UK has developed a policy for its own use of learning analytics. As one would expect, this is both more specific and more detailed than the Jisc policy, which is aimed at the sector in general. The policy was published in 2014, and it remains a reference point for the sector, the OUUK policy is composed of eight principles, which we reproduce below. For more details please see the Policy on Ethical use of Student Data for Learning Analytics (The Open University 2014b), the Frequently Asked Questions (FAQs) on the policy (The Open University 2014a), and the guide to ‘Using information to support student learning’ (Open University n.d.).

• **Principle 1**: Learning analytics is an ethical practice that should align with core organisational principles, such as open entry to undergraduate level study.

• **Principle 2**: The OU has a responsibility to all stakeholders to use and extract meaning from student data for the benefit of students where feasible.

• **Principle 3**: Students should not be wholly defined by their visible data or our interpretation of that data.

• **Principle 4**: The purpose and the boundaries regarding the use of learning analytics should be well defined and visible.

• **Principle 5**: The University is transparent regarding data collection, and will provide students with the opportunity to update their own data and consent agreements at regular intervals.

• **Principle 6**: Students should be engaged as active agents in the implementation of learning analytics (e.g. informed consent, personalised learning paths, interventions).

• **Principle 7**: Modelling and interventions based on analysis of data should be sound and free from bias.
Principle 8: Adoption of learning analytics within the OU requires broad acceptance of the values and benefits (organisational culture) and the development of appropriate skills across the organisation.

4.3. LACE - DELICATE Checklist

Drachsler & Greller (2016) developed the eight point DELICATE checklist to support a new learner contract, as the basis for a trusted implementation of Learning Analytics. The checklist is derived from the intensive studying of the legal texts mentioned above and several round-table talks of experts, e.g. at the six EP4LA workshops. It has been put into its final shape by the LACE project and in close cooperation with its associated partners across Europe.

The DELICATE checklist contains eight action points that should be considered by managers and decision makers planning the implementation of Learning Analytics solutions either from their own institution, or from any external provider. The eight points are: 1. Determination, 2. Explain, 3. Legitimate, 4. Involve, 5. Consent, 6. Anonymise, 7. Technical aspects, 8. External partners. Figure XY below shows the full overview of the checklist and all its relevant sub questions.

The DELICATE checklist should be understood as a supportive instrument for any educational institution to demystify the ethics and privacy discussions around Learning Analytics. It is inspired by medical checklists that are used to increase awareness and monitoring of critical medical procedures in the operating theatre. DELICATE is envisioned as a checklist for all ‘data operating theaters’ with access to confidential student data to remind them about specific privacy aspects. The DELICATE checklist can support the design for value and privacy conform Learning Analytics that can benefit all stakeholders and keep control with the users themselves and within the established trusted relationship between them and the institution.

The eight-point DELICATE checklist offers a concise approach to educational managers and policy makers who are moving towards a data-driven educational institution. Working through the DELICATE checklist involves all key stakeholders of the Learning Analytics process and enables a trusted environment for the deployment of Learning Analytics.
Is Privacy A Show-Stopper For Learning Analytics? A Review Of Current Issues And Solutions

The DELICATE Checklist
Implementing trusted Learning Analytics in education

**DETERMINATION** – Why you want to apply Learning Analytics
- What is the added value (Organisational and data subjects)
- What are the rights of the data subjects (e.g., EU Directive 95/46/EC)

**EXPLAIN** – What are the objectives and boundaries
- What data will be collected for which purpose?
- How long will this data be stored?
- Who has access to the data?

**LEGITIMATE** – Why you are allowed to have the data?
- Which data sources you have already (aren’t they enough)
- Why are you allowed to collect additional data?

**INVOLVE** – Involve all stakeholders and the data subjects
- Be open about privacy concerns (of data subjects)
- Provide access to the personal data collected (about the data subjects)

**CONSENT** – Make a contract with the data subjects
- Ask for a consent from the data subjects before the data collection
- Define clear and understandable consent questions (Yes / No options)
- Offer the possibility to opt-out of the data collection without consequences

**ANONYMISE** – Make the individual not retrievable
- Anonymise the data as far as possible
- Aggregate data to generate abstract metadata models (Those do not fall under EU Directive 95/46/EC)

**TECHNICAL** – Procedures to guarantee privacy
- Monitor regularly who has access to the data
- If the analytics change, update the privacy regulations (new consent needed)
- Make sure the data storage fulfills international security standards

**EXTERNAL** – If you work with external providers
- Make sure they also fulfill the national and organisational rules
- Sign a contract that clearly states responsibilities for data security
- Data should only be used for the intended services and no other purposes

The LACE Project is supported by the European Commission
Seventh Framework Programme under grant 619424.
5. Conclusions

5.1. A fast-developing environment

In this paper we have seen that ethics and privacy are both high profile issues within the field of learning analytics. We have discussed a number of initiatives which are addressing the issues which are arising, but are many pending contradictions and issues which remain unresolved and confused, both in policy and at the level of the individual.

As we have described in section 4.1.3, the contextual integrity approach to data (Nissenbaum 2009) emphasises that the acceptability of the use of data depends on where it was generated and where it is being used. From a similar perspective the European Data Protection Supervisor (EPDS) has stated that the “EU should preserve, simplify and operationalise the established notion that personal data should only be used in ways compatible with the original purposes for collection” (EDPS 2015, p.5). This is not an EU directive, but it illustrates a policy perspective which is entirely at odds with the Big Data business model that explicitly aims to collect and integrate as many data sources as possible and gain new insights from those data through overarching mining and analyses.

One of the basic principles of research ethics since the Nuremberg Code has been informed consent of the participants (cf. Kay et al. 2012). This is watered down a little in the OECD principles, which state that “Data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject” (OECD 2013b, p. 14-15). Learning analytics practice, however, has shifted significantly from these principles, tending to challenge the concept of data minimization (focused collection) and consent requirements (Tene & Polonetsky 2013). The Jisc Code of Practice sets out the requirements for educational institutions as follows:

Institutions will define the objectives for the use of learning analytics, what data is necessary to achieve these objectives, and what is out of scope. The data sources, the purposes of the analytics, the metrics used, who has access to the analytics, the boundaries around usage, and how to interpret the data will be explained clearly to staff and students. Institutions should also clearly describe the processes involved in producing the analytics to students and staff or make the algorithms transparent to them. Students will normally be asked for their consent for personal interventions to be taken based on the learning analytics. (N. Sclater & Bailey 2015)

This makes is clear that, according to the Jisc guidelines, institutions do not have to ask students for permission to gather, hold and analyse their data. Consent is reframed to refer to permission to take action on the results of data analysis, a quite different matter from obtaining permission to gather data. In taking this position Jisc does no more than recognise current practice, and indeed the proposal that institutions should normally obtain consent to take action on the results of analytics is more than most educational institutions currently do to obtain consent from their students. This is not entirely surprising given the historical context, in which education institutions have never requested permission to hold data about students’ performance in examinations, class activities, library use or attendance. This data has always been seen as simply being essential for the functioning of the institution, and to the benefit of all concerned, including students. The Jisc guidelines also refer to this perspective:
It is vital that institutions monitor the quality, robustness and validity of their data and analytics processes in order to develop and maintain confidence in learning analytics and ensure it is used to the benefit of students. Institutions should ensure that:

- Inaccuracies in the data are understood and minimised
- The implications of incomplete datasets are understood

Similarly Principle 2 of the *Open University Policy on Ethical use of Student Data for Learning Analytics* states that “The OU has a responsibility to all stakeholders to use and extract meaning from student data for the benefit of students where feasible”. The implication of these guidelines from Jisk and OUUK is that there is an ethical duty on the institution to gather the best data that it can about its learners, to ensure that the service that it provides is as good as it can be. This is explicitly stated in the OUUK FAQs on their policy: “Unfortunately, it is not possible, at present, to have your data excluded. In order to have a complete dataset, the University will use all student data to analyse patterns of behaviour” (The Open University 2014a). The implication seems to be that if individuals were given the right to opt-out, then this could be seen as unethical, because opting-out reduces the efficacy of learning analytics which can improve the education of others.

It can also be argued, however, that the quantity of information held by institutions about their learners has changed the nature of the institution’s holding of data to such a degree that past practice is no longer a reliable guide. Moreover, if the requirement of the individual to expose their data for the greater good of society were applied more widely, this could lead to a society in which all citizens were permanently monitored, with troubling ethical and political implications. How these matters will be resolved is as yet unclear. They are part of a much wider social and political struggle.

The powerful economic forces of the Internet giants such as Google, Microsoft, Facebook and Amazon, leverage access to big data find themselves in an uneasy relationship with state security agencies which make use of many of the same techniques. For example the head of the National Security Agency (NSA) is reported as stating about access to Internet communications “I don’t want a back door. I want a front door. And I want the front door to have multiple locks. Big locks” (Nakashima & Gellman 2015). Google, on the other hand, has shown its distrust of surveillance by encrypting all Gmail messages (Moscaritolo 2014), even though it has in the past collaborated with the NSA (Ellen Nakashima 2010), and in the opinion of some activists is still doing so (Schneier 2014). The indiscriminate access to data of both the state surveillance and the Internet giants is opposed by privacy campaigners, such as Edward Snowden, and by the designers of technological systems that can enhance privacy, such as openPDS project run by Sandy Pentland at MIT, and the Hub of All Things funded by the UK Research Council and the d-cent project funded by the EU. Privacy and ethics in learning analytics have their own dynamic, and are informed by the traditions of education, but their future will surely be strongly informed by the wider ethical, political, economic and technological debate.

The confusion at the policy level is reflected at the personal level. It is clear that some users have concerns about privacy and about the potential for their personal information to be tracked and

---

10 [http://hubofallthings.com/](http://hubofallthings.com/)
11 [http://dcentproject.eu/](http://dcentproject.eu/)
made accessible to other users for unknown purposes. On the other hand, social media are deeply integrated into users’ daily lives and routines (Debatin et al. 2009) and people demonstrate a willingness to share many personal details on these networks. Privacy attitudes and privacy behaviours, thus, often differ (Stutzman & Kramer-Duffield 2010). This dichotomy is evident when comparing users’ self-reports about how they understand cautions about privacy settings and their actual, unconcerned behaviour of usually just keeping default settings without taking the opportunity to adjust them to their needs and preferences (Debatin et al. 2009). So it seems that people often do not act according to the privacy opinions that they express, and often they appear to be unconcerned about data protection and privacy until it is breached (Spiekermann & Cranor 2009).

5.2. Institutional strategy in response to an environment in flux
What is the institution to do in response to this uncertainty? Is it the case, as the title of this paper suggests, that privacy issues will be a show-stopper for learning analytics. In our judgement this is unlikely. Irrespective of the ethical and political concerns about data management, there is huge economic, social and political momentum behind the big data business model, and this momentum is reflected within the domain of education, for good or for ill. Nevertheless, the cautionary scenarios which we have set out in section 3 indicate why care is needed in implementing analytics, as mistakes can be costly. The trust of students and staff is essential to institutions, and the protection of privacy should not been seen as a burden but rather as a valuable service that can be offered to build trusting relations with our stakeholders (Hoven 2008). To achieve this, support for privacy should be embedded in learning analytics tools, in order to increase the trust that data subjects have in these systems. To build this kind of relationship, a high degree of openness is needed, combined with reassuring explanations referencing relevant legislation like the EU Directive 95/46/EC, and a conscious adjustment of the formal and informal contracts between learners and their educational providers.

5.2.1. The role of guidelines
We have discussed a range of more or less generic approaches of defining guidelines, model codes, and principles for appropriate data and privacy protection. In a rapidly evolving environment there is a need for new approaches to the definition of frameworks for dealing with ethics and privacy. Organisations researching and providing learning analytics technologies, as well as educational institutions adopting learning analytics, need to set up mechanisms and policies to address ethical and privacy issues in learning analytics in a context-dependent and appropriate manner. While these guidelines will not settle debates around privacy and ethics, they can help to ensure that institutions make a conscious and informed decision on policy, that this is in line with the regulatory environment, and that all members of the institution understand the policy and the reasons for it. It will also be necessary for the education sector as a whole, and institutions in particular, to be more explicit in distinguishing their practice from commercial enterprises, and make explicit their mission of care and support for individuals. The achievement of a common understanding on these issues is an important achievement in an educational institution, which depends on trust, and is also under close regulatory supervision.

The ethical frameworks proposed in the literature, and the existing ethical and privacy guidelines summarised above, can provide a useful starting point for the creation of policies and principles for
the ethical use of data in learning analytics. Building on these different resources, a well-defined ethical code may be established, which leaves little room for ambiguity – and this is a pressing need in learning analytics (Berg 2013). The Open University UK ‘Ethical use of Student Data for Learning Analytics Policy’ (The Open University 2014b) is a trail-blazing document in this respect, specifying eight principles that provide a university-wide guide for the ethical use of learning analytics to analyse student data and identify interventions for student support.

5.2.2. Privacy and ethics by design

Learning analytics guidelines, however, are usually generic, and need to be aligned with the very specific contexts of particular learning analytics applications. Tene and Polonetsky (2013) discuss this in terms of “levers that must be adjusted to adapt to varying ... conditions” (p. 242). To achieve this alignment, ethics and data protection guidelines not only need to be translated into an adequate general privacy policy and information practice, but also be technically reflected in system functionality, provisions, and data structures. Consequently the use of privacy by design principles (cf. Section 4.2.4) provides a valuable starting point to ensure that data protection is embedded into the design and architecture of learning analytics, using methods such as value-sensitive design specified by Friedman (1997). This proposes that ethical analysis and reflection need to take place when and where they can make a difference for the design and governance of technology, starting early on in the design and development process, and close to where the technology is being shaped and designed. We recommend that learning analytics projects should follow a value-sensitive design process, which enables ethical and privacy values to be considered on the same level as functional requirements. In this way ethical considerations are seen not as an unfortunate constraint, but rather as a means to help develop a system that achieves its aims not only in a technical but also in an ethical and humane manner.

6. References

Anonymous Academic, 2015. My students have paid £9,000 and now they think they own me. The Guardian. Available at: http://www.theguardian.com/higher-education-network/2015/dec/18/my-students-have-paid-9000-and-now-they-think-they-own-me.


Bomas, E., 2014. How to give students control of their data. LACE project blog. Available at: http://www.laceproject.eu/blog/give-students-control-data/ [Accessed February 1, 2016].


Campbell, L., 2013. inBloom to implement Learning Registry and LRMI. CETIS. Available at: http://blogs.cetis.org.uk/imc/2013/02/08/inbloom-to-implement-learning-registry-and-lrmi/
Is Privacy A Show-Stopper For Learning Analytics? A Review Of Current Issues And Solutions


and on the free movement of such data. Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.1995.281.01.0031.01.ENG.


Jisc, About us. *Jisc web site*. Available at: https://www.jisc.ac.uk/about [Accessed February 1, 2016].


Kharif, O., 2014. Privacy Fears Over Student Data Tracking Lead to InBloom’s Shutdown. *Bloomberg Businessweek*.


Mathews, D., 2014. Rosy prospectuses “misleading” students. *Times Higher Education*. Available at:
Is Privacy A Show-Stopper For Learning Analytics? A Review Of Current Issues And Solutions


ISSN:2057-7494


The Economist, 2014. Withered inBloom. The Economist “Schumpeter” Business and management
Is Privacy A Show-Stopper For Learning Analytics? A Review Of Current Issues And Solutions


The Open University, 2014a. Ethical use of Student Data for Learning Analytics Policy FAQs. Available at:

The Open University, 2014b. Policy on Ethical use of Student Data for Learning Analytics. Available at:


World Economic Forum, 2013. Unlocking the Value of Personal Data: From Collection to Usage, Available at:
7. About ...

7.1.1. Acknowledgement
This document was produced with funding from the European Commission Seventh Framework Programme as part of the LACE Project, grant number 619424.

7.1.2. About this document
(c) 2016, Dai Griffiths, Hendrik Drachsler, Michael Kickmeier-Rust, Christina Steiner, Tore Hoel, Wolfgang Greller.

Licensed for use under the terms of the Creative Commons Attribution v4.0 licence. Attribution should be “by Dai Griffiths, Hendrik Drachsler, Michael Kickmeier-Rust, Christina Steiner, Tore Hoel, Wolfgang Greller., for the LACE Project (http://www.laceproject.eu)”.

For more information, see the LACE Publication Policy: http://www.laceproject.eu/publication-policy/. Note, in particular, that some images used in LACE publications may not be freely re-used.

The Learning Analytics Review is published by the LACE project at the University of Bolton, Bolton, UK. Please cite this document including the issue number and the ISSN of the Learning Analytics Review, which is 2057-7494.

The persistent URL for this document is: http://www.laceproject.eu/learning-analytics-review/privacy-show-stopper

7.1.3. About LACE
The LACE project brings together existing key European players in the field of learning analytics & educational data mining who are committed to build communities of practice and share emerging best practice in order to make progress towards four objectives.

| Objective 1 – Promote knowledge creation and exchange |
| Objective 2 – Increase the evidence base |
| Objective 3 – Contribute to the definition of future directions |
| Objective 4 – Build consensus on interoperability and data sharing |

http://www.laceproject.eu @laceproject