British Journal of Educational Technology *Vol 45 No 3 2014* 438–450
doi:10.1111/bjet.12152

Ethical and privacy principles for learning analytics

Abelardo Pardo and George Siemens

Abelardo Pardo is a lecturer at the School of Electrical and Information Engineering at The University of Sydney, Australia. He is working on research projects exploring how technology can be used to understand and influence human behavior. He has experience in the use of digital devices in areas such as behavioral analytics, social networks, computer-supported collaboration, personalization and technology-enhanced learning. George Siemens is the Executive Director of the Learning Innovation & Networked Knowledge (LINK) Research Lab at the University of Texas Arlington. His research interests include learning in networks, data and analytics in education, social and technical sense making in complex information settings, and the ways in which technology, labour and demographic trends, and participatory culture influence the future of higher education. Address for correspondence: Dr Abelardo Pardo, Electrical Engineering Building, J03, The University of Sydney, Sydney, NSW 2006, Australia. Email: abelardo.pardo@sydney.edu.au; George Siemens, 300 Nedderman Hall, University of Texas Arlington, TX 76019. Email: gsiemens@uta.edu

Abstract
The massive adoption of technology in learning processes comes with an equally large capacity to track learners. Learning analytics aims at using the collected information to understand and improve the quality of a learning experience. The privacy and ethical issues that emerge in this context are tightly interconnected with other aspects such as trust, accountability and transparency. In this paper, a set of principles is identified to narrow the scope of the discussion and point to pragmatic approaches to help design and research learning experiences where important ethical and privacy issues are considered.

Introduction: privacy in learning environments

The use of information and communication technology has significantly changed how learning experiences are conceived and deployed. The widespread use of various digital devices together with cloud computing allows for learning scenarios not previously considered. Students are now able to access a myriad of learning resources, interact with applications focusing on a specific topic, enhance their experience in virtual environments, augment reality and connect with others through social networks. The progress of technology evolves together with the capacity to record the events occurring in a learning environment. Every interaction and resource accessed can be captured and stored. As a consequence, learning scenarios can now be analyzed using *big-data analytics techniques*. Although the use of new technology is shaping the way we learn, a more significant change may derive from the use of big-data analytics (Siemens & Long, 2011).

The ability to capture so much of the learning experience has many repercussions. For example, with the capability of *observing* students while they work on an activity, it is possible to deploy new assessment techniques that measure more accurately the right achievements. Also, a more effective feedback loop can be created where students receive information from instructors in a short amount of time. Likewise, instructors may quickly detect and intervene in situations where timely instructional support can benefit learners. Instructional designers are also able to incorporate analytics results into future design practices. All these activities rely on collecting as much data as possible during the learning process. This shift in analytics capacity also raises the issue of privacy.

While it has numerous definitions, in this document, *privacy* is defined as the regulation of how personal digital information is being observed by the self or distributed to other observers. By

Practitioner Notes

What is already known about this topic

- Learning analytics offers the possibility of collecting detailed information about how students learn.
- The ethical and privacy issues derived from these scenarios are not properly addressed.
- There is a need to clarify how these issues must be addressed from the early stages of the deployment of a learning analytics scenario.

What this paper adds

- An account of how the main legislations are advancing in the general area of privacy.
- A comparison of how other disciplines such as medicine have dealt with privacy issues when collecting private information.
- The description of a group of practical principles in which to include all the ethical and privacy-related issues present when deploying a learning analytics application.

Implications for practice and/or policy

- Designers may now take into account these principles to guide the implementation of learning analytics platforms.
- Students are now aware of the different categories of issues that must be addressed by applications collecting data while they learn.
- Instructors now have a more detailed account of the issues to address when adopting learning analytics techniques.

personal digital information, we adopt a broad definition including the information about persons captured by any means and then encoded in digital format. In the digital context, we define *ethics* as the systematization of correct and incorrect behavior in virtual spaces according to all stakeholders. In this paper, we present a description of the ethical and privacy issues as they manifest in the specific context of learning analytics research. Although the scenario shares numerous analogies with other areas such as social networks or ecommerce sites, it also has significant differences that prompt the need for an analysis of the environment in which these are deployed in academic settings. A set of principles for addressing privacy is first presented and then described within the context of the five steps of learning analytics (Campbell, DeBlois & Oblinger, 2007).

In recent years, the increasing use of digital devices, such as accessing the Internet with smartphones (Ipsos MediaCT, 2011) or the appearance of *ambience intelligence* (Rouvroy, 2008), offers an unprecedented capacity to observe users interacting with these devices. The ability to communicate with an ever-growing number of users comes together with the possibility of sharing delicate information, and therefore with the risk of violating the privacy rights of individuals. These concerns are equally valid in the context of learning analytics. For example, in a hypothetical scenario, is the improvement of the overall learning environment a valid reason to record the exact location of students within the institution and share it with peers to facilitate collaborative learning?

With the advent of technology, approaches to privacy have to be reevaluated (Nissenbaum, 2004; Oravec, 1999; Palen & Dourish, 2003; Solove, 2008). Conventional metaphors to illustrate privacy such as *Big Brother*, or the *panopticon*, are no longer adequate to understand privacy today (Haggerty & Ericson, 2000). As a consequence of changing views, the debate over privacy is increasingly polarized. On one hand, there are numerous authors and consumer advocate organi-

zations that claim that privacy is being seriously threatened by the lack of proper legislation on how information may be collected, processed, analyzed and distributed (Deng, Wuyts, Scandariato, Preneel & Joosen, 2010; Privacy Rights Clearinghouse, 1999). Government agencies, particularly those dedicated to national security, are creating legislation to control, and in some cases, capitalize on the wealth of data that are contained in distributed databases and the potential to deduce additional information. There are accounts of security breaches, which incurred in information misuse that clearly violated fundamental rights. For example, in June 2012, a hacker posted almost 6.5 million passwords of the social network LinkedIn in a forum (Souppouris, 2012). The company advised users to change the passwords immediately, but the private data of these accounts were already exposed.

On the other hand, some influential figures in the technological landscape state that privacy is no longer an issue because it is too late to restore any sort of basic guidelines to manage it properly. For instance, in 1999, the then Chief Executive Officer of Sun Microsystems Scott McNealy stated that privacy no longer existed and that users should "get over it" (Sprenger, 1999). More recently, Mark Zuckerberg, the founder of Facebook, stated that the age of privacy is over, arguing that certain information about users should be made public by default (Kirkpatrick, 2010, January).

Palen and Dourish (2003) proposed characterizing privacy in digital environments on three boundaries: discourse, identity and temporality. The discourse boundary comprises the simultaneous need to maintain some information private, and at the same time, some other information publicly available and known. There are situations in which making personal information publicly available is used to increase the level of privacy in other personal aspects. The so-called *identify boundary* is used to symbolize the tension between the private and public information from the self to others. The numerous ways in which people interact blur the line between the self and the group. The third boundary refers to the temporal nature of information. The information collected in the past can be used to infer future behavior. The tension with respect to privacy emerges from the attempt to control how the past information is used to predict future actions.

Far from settling the issue, technology has turned privacy into one of the most controversial aspects of online and digital interaction. Governments are beginning to address privacy as there is a clear need for legislation to provide guidelines on what exactly is understood by privacy and the rights that need to be preserved. Currently, a low level of *legal 'maturity'* (Kay, Korn & Oppenheim, 2012) exists as legal systems are still at the early stages of commenting on privacy, ethics and data ownership. At the same time, businesses are also addressing this issue as there is a connection between privacy and consumer trust, which is essential to maintain the social and economic benefits provided by technology (The White House, 2012). In the middle of this landscape, society seems to be evolving toward a situation in which the exchange of personal data is normal, and a delicate balance between control and limits needs to be achieved (Schwartz, 2011). Furthermore, some studies suggest that the concerns of users about privacy vary significantly depending on what is being observed, the context and the perceived value when granting access to personal information (Klasnja, Consolvo & Choudhury, 2009). This variation partially derives from the vague definition of what is considered *personal information*. Although personal data can be easily identified, the definition tends to also include the processes applied to these data and the resulting inferences. It is with the inclusion of such ample terms that the definition becomes much broader (Narayanan & Shmatikov, 2010).

Data analysis techniques have been in use for quite some time in the form of business intelligence and marketing. The generic model is to collect and process data, and support decision processes at various levels (visualizations, recommendations, automatic reasoning, etc.). The connection with privacy is clear. In principle, the more comprehensive the set of data collected, the better support is offered to these decisions. Whenever data are collected from users, a *transaction* takes place. A

business or institution obtains something valuable from a user, but what does the user get in exchange? Generally, with existing technology services, users are granted *free* access to a service such as Twitter or Facebook. Essentially, users are exchanging certain amount of privacy by access to the service. In most cases, however, they are not made aware of the data being exchanged for the service, or even worse, the service provider changes the conditions during this relationship. An increasing number of users are becoming aware of this emerging trend, as is evident with the concerns expressed when popular sites such as Twitter, Facebook or Google+ change their privacy policies or settings. Companies argue that the ultimate goal is to improve customer satisfaction, but an incorrect use of this information is perceived as a breach of contract and may prompt customer rejection (Singer, 2012).

Learning analytics is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Long, 2011). This capacity of collecting data of the events taking place in an environment can be applied to a learning experience. The application of these techniques to learning is arguably straightforward, but privacy and ethical issues persist in these environments with some additional considerations beyond what is found in how consumers use a software service. Elgesem (1999) portrays the notion of privacy in the context of various possible *channels*. A channel is an entity in which information flows among a set of stakeholders. Most of the principles that should be taken into account when dealing with privacy issues depend on the type of channel under consideration. The notion of users acting with different roles in diverse settings has been identified as one of the main factors to take into account when assessing privacy in modern contexts (Nissenbaum, 2004).

The objective of this paper is to present an analysis of how privacy and ethical issues apply to the specific context of learning analytics and learning analytics research and propose guidelines to comply with the most common privacy principles emerging in various legislative initiatives. From the detailed categorization of privacy research provided by Smith, Dinev and Xu (2011), this document fits in the category of *contextual nature of the relationship between information privacy and other constructs*.

The information included in this paper may be applied to the following hypothetical scenarios:

- A researcher wants to explore the level of discourse of students when commenting issues posed as part of an online course. Students express their views in a closed platform. May the data be collected and used for publication? Is user consent needed to perform the analysis? Should other users be made aware of information related to the overall discussion? Can individuals be identified based on the derived data?
- The study of the attitudes of students when involved in teamwork is considered to improve collaboration skills. Teams are interviewed once a week by an observer and their digital traces are exhaustively collected. Should the information be made available to the team members? Is it necessary to allow students to *opt out* of the study? Should there be consequences if they choose to do so? May the institution use the collected data to create predictive models for team conflicts?
- A platform used to provide corporate training for employed adult learning records all the interactions among students. Should these data be made available to the employer?
- A researcher wants to study the correlation of students' activities across various courses the students are taking simultaneously. Should the information collected in one course be made available to teaching staff of another course? What level of anonymity is needed to publish these results? Should students be allowed to view the differences between courses?
- Researchers in massive open online courses are monitoring the development of concepts and knowledge in a discussion forum. Even though the learning is happening in an open online platform, any use of direct quotes can later be traced to specific learners. What are the

principles that guide research on open public interaction data? Should learners involved in an open course be required to give consent for data collection and analysis?

The ethical issues arising from these scenarios are simply a sample of the wide variety of potential cases. A comprehensive description of the issues in each possible scenario is beyond the scope of this paper. The suggestions provided must be taken as generic and need to be adapted to a specific context. The rest of the paper is organized as follows: Section 2 contains an account of how privacy issues are being solved in other related contexts; Section 3 includes a discussion of the main privacy principles in the context of learning analytics and a categorization of these issues; and Section 4 presents the conclusions of the analysis.

How are privacy and ethics addressed in other contexts?

A unified definition of privacy is elusive. As with the variety of scenarios emerging in which privacy has a central role, the number of concepts that are encompassed by the term is increasing (Solove, 2008). At early stages, privacy was defined in terms similar to the concept of *opacity* (Poullet, 2010). However, this vision has evolved into a combination of control and limitations. Control refers to the capability of individuals to influence the flow of their personal information. Limitations relate to the possibility of preventing others to access private data (Elgesem, 1999). Tavani (2007) identified the theories based on control and limitations as most adequate to analyze generic privacy concerns. He proposes a so-called *restricted access/limited control* theory in which these two terms are used to delimit the boundaries of certain environments. The theory is illustrated in the field of data-mining technology, which bears some proximity with learning analytics in the context considered in this document.

A description of the current legislation on privacy in the world is beyond the scope of this paper, but the first instances of this type of regulations have appeared in countries with a high penetration of Internet use. Examples of such initiatives are the Consumer Data Privacy in a Networked World in the USA (The White House, 2012), the National Privacy Principles in Australia (Australian Government, 2006), the European Union directive on protection of individuals with regard to processing of personal data (EUP, 1995), later extended with the directive on the processing of personal data and the protection of privacy in the electronic communications sector (EUP, 2002) or the Model Code for the Protection of Personal Information in Canada (Canadian Standards Association, 2001). The case of the European Union can be taken as an example of the fast pace of evolution in this area. After enacting a directive to regulate how information is processed in the electronic communication market in 1997, a second directive superseding the previous one was passed in 2002 to adapt to new market requirements. Recently, this directive underwent another round of reviews in which changes related to how *cookies* are stored in digital devices were considered (Cheverie, 2012). As a result, in May 2012, the new directive required informed consent for storing cookies in the user computer. This is an example of the rate of change that affects privacy in today's society. Even the Federal Trade Commission in the USA acknowledges this trait in the title of its report "Protecting Consumer Privacy in an Era of Rapid Change" (FTC, 2010). In parallel with these legislative efforts, nonprofit organizations have appeared to defend user digital rights. For example, the Electronic Frontier Foundation was created as early as 1990 to protect user rights in the digital world.

A comprehensive definition of *right to privacy* in learning analytics research environments is equally elusive. Even though a research team should have clear guidelines to handle student privacy, institutions are still struggling to create a coherent vision that applies to areas other than learning analytics. Although educational institutions typically have human research ethics committees or institutional review boards that already process the requests to carry research initiatives involving the collection of personal data, learning analytics poses some new boundary conditions. The notion of *subjects being put at risk* by a study, which is central for studies in other

disciplines, is not that evident in learning analytics where large, often anonymized data sets are used for analysis. On the other hand, privacy issues related to the way the collected information is handled are more relevant. For example, the use of tools to detect plagiarism raises some issues regarding the property of the information that is being handled. If a student submits a document, it could be used at a later point in time to check if a new document has been copied. Who owns such document? Is it right for a plagiarism detection tool to store these documents (Brinkman, 2012)? Because of its complexity, the use of learning analytics in academic environments has to tackle these and other issues.

In other areas such as medical research, issues of privacy and ethical use of data have been dealt with for quite some time, and therefore, there is value in analyzing the possibility of translating some of the policies used in those fields to learning analytics. Absolute confidentiality when related to medical records has been shown to have a negative impact specially when referring to genetic information and inherited diseases (Crook, 2011). In such context, the argument is that the benefits of analyzing population data allow the advance of medical research considered as *positive in general*. A similar argument could be stated about learning research. Collecting, storing and using historic data about how students learn could advance the research in the area and ultimately benefit the quality of future learning experiences; hence, absolute confidentiality should be discarded. However, as in the case of medical records, some of them are private, and therefore, clear and robust policies should be observed when managing these data. Additionally, in a learning scenario, the number of stakeholders may increase significantly. Aside from students and teachers (which could be assimilated to patients and doctors in the analogy with medical research), multiple departments within the educational institutions, companies in the educational space and governmental agencies at various levels may use the data collected to drive research initiatives or policy making. Ideally, learning institutions should converge toward a code of research ethics that should be continuously reviewed and modified by the relevant stakeholders defining the appropriate approach to privacy as it is done in other areas (Svensson & Hansson, 2007). However, the difference in the maturity level of the area of learning analytics research still requires a significant effort to delimit clearly this code and adapt it to the nuances of the area rather than simply extrapolate from other contexts. Some initial contributions to clarify this field have appeared. For example, Slade and Prinsloo (2013) have proposed a set of principles from a sociocritical perspective to guide institutions to address ethical issues in learning analytics.

The issue of medical record management has also identified long before the advent of learning analytics research the problem of *data ownership*. The observations are obtained from the users, but typically, there is an underpinning research effort being carried out by a research team and, quite frequently, supported by some funding agency. This situation can be applied to a learning research environment. Who owns the data: the institutions, the students, the companies using them for tuning various educational products? Even if we assume that the raw data collected from the users belong to them, what happens with the data derived from it? As it will be described in the following sections, this issue appears frequently in the principles underlying current regulations. Learning analytics research initiatives need to respect existing regulations while advocating for guidelines and principles that make analytics more useful to all stakeholders.

The objective of research initiatives in learning analytics is to provide techniques to understand and optimize learning and the environment in which it occurs. This optimization can be achieved through a variety of means, including personalization, which in web applications has a long trajectory of research to offer to each student the most appropriate content. Kobsa (2007) studied in detail the connection between personalization and privacy. The common point between analytics and personalization is that both are based on collecting private information about the

user and the use of those data to improve the learning experience of individuals. As a consequence, the so-called basic principles for fair use of personalization offer a good starting point to consider an analogous set of principles for analytics.

The principles described in this document are similar in spirit to those described by Schwartz (2011) in a study of the consequences of deploying analytics and the rules that need to be observed for this deployment to comply with different legal, social and cultural boundary conditions. Although the study is in the context of corporations, we believe the methodology for the study and the derived recommendations apply equally to analytics research in educational contexts. It is important to note, however, that the specific principles of privacy and ethics will vary from country to country, and in many instances, from one institution to another.

Analytics research in educational institutions

Learning analytics research assumes the use of a variety of computation-based technologies to gather and analyze events captured while students interact in a learning environment. Because of the presence of technology, privacy must be addressed from both technological and legal aspects (Le Métayer, 2009). On the technical part, there are already several proposals that take privacy into account while designing tools in a model that could be called *privacy by design* (Le Métayer, 2010; Spiekermann, 2009). Designers are encouraged to include privacy and security issues in the early stages of the design, so that the produced platforms are in a position to comply with the requirements derived from the environment in which they operate. For example, the recording of the events occurring in a learning management system needs to guarantee that the events contain meaningful information, the access can be easily restricted and, as it will be explained later in the section, users are allowed to access and request data corrections.

In some cases, the requirements emerging from privacy and security have been included as part of computer engineering programs (Voss, 2012). Privacy is an important issue in today's computing applications and therefore needs to be part of the learning outcomes of the relevant degrees. Technological issues must go hand in hand with other ethical and legal requirements. A first distinction that helps shape the context in which analytics is deployed in educational institutions is the fundamental difference between personally identified information (PII)—eg, name, address—and nonpersonally identified information (NPII). If this distinction is clear, any data management application will be easier to implement. Unfortunately, this dichotomy is disappearing due to the use of triangulation techniques that may combine several PII data such as browser type, operating system or configuration and derive another NPII factor such as the user identity (FTC, 2010; Schwartz, 2011).

Fostering trust

User behavior in online platforms has been analyzed extensively due to the maturity of generic analytics or personalization techniques. Several studies have identified *trust* as one of the most important traits to improve user experience on Internet, analyzing how users behave when facing an online application and the type of features that contribute to increase user trust (Koehn, 2003). Additional studies have extended this approach by considering how trust is taken into account when users make decisions about privacy and how it affects the perception of the overall user experience (Knijnenburg & Kobsa, 2012). Issues such as proximity with the rest of the users or letting the user know exactly what is being tracked appear as two of the most influencing factors. Trust is an equally important issue within a learning institution, and therefore, it also needs to be understood through the lens of learning analytics.

Social networks, in particular Facebook, were initially targeting the student population. Now, even with widespread use, the student population is still a significant percentage of users of these

networks. Toch, Wang and Cranor (2012) published a survey about the privacy risks derived from the personalization based on social networks. Users publish information in these sites that, when used to personalize content, may potentially embarrass users in front of friends or colleagues. Lewis, Kaufman and Christakis (2008) analyzed how students reacted to privacy in the context of social networks. The personal choices with respect to security settings of a cohort of students were studied, and it was concluded that the more active students are in these platforms, the more aware they are of privacy issues and likely to have a private profile (not accessible to the open public). Martin (2011) carried out another relevant study on which almost 1000 students answered a set of questions about various scenarios in which privacy decisions were shown. The results pointed at the relationship with other individuals and the protection offered by the application for the data to be transferred as the two factors that influence the amount of data disclosure.

These studies also point out that the deployment of learning analytics in educational institutions inherits the intricacies of other analytics environments, but at the same time must be done by taking into account the specific traits of the students.

Principles of learning analytics deployment

The principles described in this section have been conceived after reviewing numerous proposals related to analytics in general, government frameworks and regulatory directives. Specific effort has been undertaken to group similar issues under a common generic principle from a pragmatic angle. Each principle may generate additional considerations. While a detailed treatment of all principles is beyond the scope of this document, the prominent categories are detailed below. The aim of these principles is to provide a baseline for educational institutions that are currently using analytics to reflect on their level of compliance and envision possible improvements with respect to privacy. For the sake of simplicity, these stakeholders have been divided into three categories: students (or their tutors if underage), instructors and administrative staff. There are cases in which data are shared with other external organizations, but from the point of view of the following analysis, they will be considered as administrative staff. The overarching assumption for these principles is that they must be observed in order to comply with legal requirements, but also with social requirements that might not be properly reflected in current laws. The discussed principles are transparency, student control, security, and accountability and assessment.

Transparency

This is arguably the principle that includes the largest number of concerns. Transparency can be applied to virtually every stage of learning analytics. In broad terms, all three stakeholder groups should have access to the description of how the analytics process is carried out and should be informed of the type of information that is being collected, including how it is collected, stored and processed. Most of the current laws require student consent for the data to be collected, but the principle of transparency can be taken beyond this requirement. As detailed in the previous section, one factor that influences the level of information disclosed is knowing whether the data are going to be transferred to another entity. Analytics in learning institutions can take this aspect as an advantage. By guaranteeing that the data are not going to be given to any other institution, a higher level of trust may develop among the students.

Data management procedures must be clearly outlined by any institution conducting analytics. An aspect that seems to affect the level of trust is how organizations handle historic data. How long will the data be kept? Will the data be used after students graduate? Institutions must carefully assess these aspects of analytics. The guarantee that personally identifiable data are deleted when a student leaves an institution may also have a positive impact on student trust, but at the same time, keeping student data will be helpful for the university to refine its analytics

models, track the development of student performance over multiple years and cohorts or simply for internal or external quality assurance processes.

Transparency can also be extended to aspects that are typically not well reflected in laws. For example, as pointed out by Elgesem (1999), the European Union directives state that data must be collected for *specific, explicit and legitimate purposes*. Applying the transparency principle to this statement, institutions should also describe to stakeholders the processes by which the data are handled and the results obtained. An important part of the value of analytics resides on how data are manipulated. For example, in January 2012, Google announced a major update in their privacy policies. The change was not on the type of data being captured, but on extending how the data from different applications were combined (Bosker, 2012). An effective way to show the legitimacy of a process is by exposing the process itself. This issue poses some ethical and technical difficulties. Should institutions make public the software they use for data analysis? What if they are using proprietary tools? There is no set model to tackle these issues, but institutions certainly can employ certain measures that convey to all stakeholders that the transparency principle is being applied as broadly as possible. For example, an institution could publish a list describing the instances in which data are being used and cases in which data are *not* used. This list can be refined with feedback from the stakeholders.

Student control over data

Student control over data is now present in most privacy regulations, but it may vary significantly on the way it is implemented. It is related to the transparency principle in the sense that, in order for students to have control over the data that are collected, they need to know what are collected, when, how and how they are manipulated. Thus, all these aspects should be transparently offered to the students. The open student model has been proposed as an approach to make not only information collected but also the models derived from that data available to learners (Bull & Kay, 2007). This approach has not been widely adopted. A significant part of how an educational institution handles this principle is captured in some form of *terms of use*. Terms of use should be carefully stated as to delimit the processes that are being applied and, at the same time, convey the idea that students must remain in control of their data at every moment.

The main aspect derived from this principle is the right of users (in this case students or tutors) to access and correct the data obtained about them. Offering students the possibility of accessing the collected data falls entirely into the transparency principle. Allowing them to correct such data is more controversial. This requirement clearly poses certain challenges and applies only to certain types of data. If students are shown a summary of the collected data set, do they see the source data as to decide if they want it modified? Should students be allowed to amend the dates and times in which they were connected to a university platform? With the exception of data that are derived from plain observation of an event, everything else should be open to being modified. Educational institutions may convey a high sense of compliance with the transparency principle by simplifying student control over data through clear procedures for data access and correction. For example, a platform in which students see the data collected over time with some visualization scheme, an intuitive distinction for data that can be changed and a simple procedure to request such change would increase the perception of transparency.

The presence of this right in various regulations has prompted the appearance of initiatives in which users are given new alternatives when controlling their anonymized private data. For example, the Portable Legal Consent (The Economist, 2012) proposes users to grant access to their data to any institution that complies with a set of conditions, the most important of them being that the results must also be publicly available. Educational institutions could adopt a similar approach. Students provide their consent to access the anonymized data provided the results are made publicly available to any educational organization.

Right of access

The collected data need to be under a clearly defined set of rights of access. When these rights are not properly observed, the consequences for user trust could be drastic. There have been a large number of episodes where security is breached and highly sensitive data of users have been exposed. Educational institutions must pay special attention to this principle as the effect of exposing sensitive data to the public may have a profound impact on all stakeholders. Because of the variety of tools and users that may gain access to the data, a detailed access policy is recommended. This policy can be complex and should be taken into account at the early design stages of the analytics framework. The policy should clearly identify the type of operations allowed in the data and also which users have access to which areas of the application. Conversely, users must be clearly informed of the type of information they are manipulating. For example, tutors typically have access to reports derived from data analysis. These reports should be treated as confidential information with certain restrictions. Educational institutions already manipulate this type of information (student records), but analytics extends this scenario to a larger variety of data and contexts.

Security appears also when considering the trade-off between access and benefit. In one extreme, total restriction to access only data produced in your immediate context seriously affects its potential for improvement. For example, students at risk of dropping a course due to changes in their economic status cannot be identified by accessing data collected in a single course. This situation can be extended to include other stakeholders such as part-time students, human resources departments or even go beyond the institutional boundaries to future or current employers. Although learning analytics as a body of research may benefit enormously if data are made public for researchers to analyze, some boundaries need to be defined. In this regard, learning analytics is no different to other scientific areas such as educational data mining or medical research. In fact, there are already some ongoing initiatives to promote the exchange of this information between different realms (Verbert, Manouselis, Drachsler & Duval, 2012).

A technique often used to lower the reluctance to sharing data is anonymization. The process consists in removing from the observed events anything that allows the identification of the user from which the data were obtained. Although these techniques seem a step in the right direction, it is highly sensitive to the type of data. In the case of data collected by educational institutions, the anonymization process might even require semantic analysis to detect those portions that could be used for identification (Teutsch, Piat & Reffay, 2009). Lately, the level of skepticism on anonymization has been increasing to the point where some authors summarize the problem saying that data can be either useful or perfectly anonymous, but never both (Narayanan & Shmatikov, 2010; Ohm, 2010).

Institutions should first make sure the security principle is present and fully honored in their approach and then consider the possibility of contributing certain subsets of their collected data to the rest of the community.

Accountability and assessment

Accountability is a principle that affects all the aspects in a learning analytics scenario. Each aspect must have a person, body, department or institution identified as responsible for the proper functioning of its related components. Accountability is a principle that translates into robustness of the overall process. For example, within a university, data security is the responsibility of the information technology infrastructure department. By identifying responsible entities, the overall analytics process is reviewed and responsibility areas are identified.

Identifying entities that are accountable for specific data and analytics areas is accompanied by the assessment principle. By assessment, we refer also to the responsibility of the institution to

constantly evaluate, review and refine the data collection, security, transparency and accountability. This is especially relevant when considering that laws and regulations applying to learning analytics are changing at a rate significantly higher than other areas. Also, and derived from the principle of adopting learning analytics in terms that are socially acceptable, a continuous refinement cycle conveys the stakeholders the compromise with such principle.

Conclusions

The evolution of technology, the increasing use of new digital devices, the adoption of technology in educational environments and the potential impact of analytics when applied to learning experiences are changing the educational landscape. Learning analytics can help institutions make decisions at all levels including students, instructors and administrations. Analytics has been used for some time in other areas such as business intelligence, and there are numerous studies suggesting the principles under this technology should be deployed in corporations to deal with privacy issues.

Educational institutions pose a new scenario with specific requirements. Students interact very intensively with the university (or its computational platforms) during a concrete time frame, carrying out very specific tasks, and produce highly sensitive data during the process. These special conditions prompt the need for a revision of the privacy principles with respect to analytics and their application in educational settings.

In this paper, an analysis of privacy issues when deploying analytics in an instructional environment has been presented. Four principles have been identified to categorize the numerous issues derived from privacy: transparency, student control over the data, security, and accountability and assessment. By discussing the various aspects within each category, institutions have mechanisms to assess their initiatives and achieve compliance with current laws and regulations as well as with socially derived requirements.

Additional information

The following locations provide information about research ethics:

- Central Office for Research Ethics Committees (COREC), UK (http://www.corec.org.uk/links.htm)
- The Ethics Application Repository (TEAR), New Zealand (http://tear.otago.ac.nz/)
- The Interagency Advisory Panel on Research Ethics (PRE), Canada (http://www.pre.ethics.gc.ca/english/)
- Social Sciences Research Ethics, UK (http://www.lancaster.ac.uk/researchethics/)
- Office for Human Research Protection, USA (http://www.hhs.gov/ohrp/)

References

Australian Government (2006). *Private sector information sheet 1A—national privacy principles*. Sydney: Office of the Privacy Commissioner.

Bosker, B. (2012). Google privacy policy changing for everyone: so what's really going to happen? *The Huffington Post*. Retrieved March 24, 2014, from http://www.huffingtonpost.com/2012/02/29/google-privacy-policy-changes_n_1310506.html

Brinkman, B. (2012). An analysis of student privacy rights in the use of plagiarism detection systems. *Science and Engineering Ethics*, *19*, 3, 1255–1266.

Bull, S. & Kay, J. (2007). Student models that invite the learner in: the SMILI:() open learner modelling framework. *International Journal of Artificial Intelligence in Education*, *17*, 2, 89–120.

Campbell, J. P., DeBlois, P. B. & Oblinger, D. G. (2007). Academic analytics. *Educause Review*, *42*, 1–24. EDUCAUSE White Paper.

Canadian Standards Association (2001). *Model code for the protection of personal information* (pp. 7–8). Ontario: Canadian Standards Association. (No. CAN/CSA-Q830-96).

Cheverie, J. (2012). Data protection reform legislation in the European Union. *EDUCAUSE Blog*. Retrieved March 24, 2014, from http://www.educause.edu/blogs/cheverij/data-protection-reform-legislation-european-union

Crook, M. A. (2011). The risks of absolute medical confidentiality. *Science and Engineering Ethics*, *19*, 1, 107–122.
Deng, M., Wuyts, K., Scandariato, R., Preneel, B. & Joosen, W. (2010). A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, *16*, 1, 3–32.
Elgesem, D. (1999). The structure of rights in directive 95/46/EC on the protection of individuals with regard to the processing of personal data and the free movement of such data. *Ethics and Information Technology*, *1*, 283–293.
EUP. (1995) Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. European Union: European Parliament.
EUP. (2002) Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector. European Union: European Parliament.
FTC (2010). *Protecting consumer privacy in an era of rapid change* (pp. 1–122). Washington: Federal Trade Commission.
Haggerty, K. D. & Ericson, R. V. (2000). The surveillant assemblage. *British Journal of Sociology*, *51*, 4, 605–622.
Ipsos MediaCT. (2011). *Mobile internet & smartphone adoption* Vol. 2011 (pp. 1–12). Germany: Ipsos MediaCT.
Kay, D., Korn, N. & Oppenheim, C. (2012). Legal, risk and ethical aspects of analytics in higher education. *CETIS Analytics Series*, *1*, 6, 1–30.
Kirkpatrick, M. (2010, January). Facebook's Zuckerberg says the age of privacy is over. *Readwrite Web*. Retrieved March 24, 2014, from http://readwrite.com/2010/01/09/facebooks_zuckerberg_says_the_age_of_privacy_is_ov
Klasnja, P., Consolvo, S. & Choudhury, T. (2009). Exploring privacy concerns about personal sensing. In H. Tokuda, M. Beigl, A. Friday, A. J. Bernheim Brush & Y. Tobe (Eds). *International conference pervasive* (pp. 176–183). Berlin: Springer Verlag.
Knijnenburg, B. P. & Kobsa, A. (2012). *Making decisions about privacy: information disclosure in context-aware recommender systems. Public policy*. Irvine: University of California.
Kobsa, A. (2007). Privacy-enhanced web personalization. In P. Brusilovsky, A. Kobsa & W. Nejdl (Eds). *The adaptive web* (pp. 628–670). Berlin: Springer.
Koehn, D. (2003). The nature of and conditions for online trust. *Journal of Business Ethics*, *43*, 1/2, 3–19.
Le Métayer, D. (2009). A formal privacy management framework. *Formal Aspects in Security and Trust*, *5491*, 162–176.
Le Métayer, D. (2010). Privacy by design: a matter of choice. In S. Gutwirth, Y. Poullet & P. De Hert (Eds). *Data Protection in a Profiled World* (pp. 1–6). Berlin: Springer Verlag.
Lewis, K., Kaufman, J. & Christakis, N. (2008). The taste for privacy: an analysis of college student privacy settings in an online social network. *Journal of Computer-Mediated Communication*, *14*, 1, 79–100.
Martin, K. E. (2011). TMI (too much information): the role of friction and familiarity in disclosing information. *Business & Professional Ethics Journal*, *30*, 1–2.
Narayanan, A. & Shmatikov, V. (2010). Myths and fallacies of "personally identifiable information.". *Communications of the ACM*, *53*, 6, 24.
Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington Law Review*, *79*, 1, 101–139.
Ohm, P. (2010). Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review*, *57*, 1701–1777.
Oravec, J. A. (1999). Integrating privacy studies into. *Journal of Information Technology for Teacher Education*, *8*, 1, 55–70.
Palen, L. & Dourish, P. (2003). Unpacking "privacy" for a networked world. In G. Cockton & P. Korhonen (Eds). *Proceedings of the conference on Human factors in computing systems—CHI '03* (p. 129). New York: ACM Press. doi: 10.1145/642633.642635.
Poullet, Y. (2010). About the e-privacy directive: towards a third generation of data protection legislation? *Data Protection in a Profiled World* (pp. 3–30). Berlin: Springer Verlag.
Privacy Rights Clearinghouse (1999). *Privacy rights handbook* (p. 335). New York: Harper Perennial.
Rouvroy, A. (2008). Privacy, data protection, and the unprecedented challenges of ambient intelligence. *Studies in Ethics, Law, and Technology*, *2*, 1, 1–54.
Schwartz, P. M. (2011). Privacy, ethics, and analytics. *IEEE Security & Privacy*, *9*, 3, 66–69.
Siemens, G. & Long, P. (2011). Penetrating the fog: analytics in learning and education. *Educause Review*, *48*, 5, 31–40.
Singer, N. (2012). E-tailer customization: convenient or creepy? *New York Times*, New York.

Slade, S. & Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist*, *57*, 10, 1510–1529.
Smith, H. J., Dinev, T. & Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS Quarterly*, *35*, 4, 989–1015.
Solove, D. (2008). *Understanding privacy*. Cambridge, MA: Harvard University Press.
Souppouris, A. (2012). LinkedIn investigating reports that 6.46 million hashed passwords have leaked online. *The Verge*. Retrieved March 24, 2014, from http://www.theverge.com/2012/6/6/3067523/linkedin-password-leak-online
Spiekermann, S. (2009). Engineering privacy. *IEEE Transactions on Software Engineering*, *35*, 1, 67–82.
Sprenger, P. (1999). Sun on privacy: "get over it". *Wired Magazine*. Retrieved March 24, 2014, from http://www.wired.com/politics/law/news/1999/01/17538
Svensson, S. & Hansson, S. O. (2007). Protecting people in research: a comparison between biomedical and traffic research. *Science and Engineering Ethics*, *13*, 1, 99–115.
Tavani, H. T. (2007). Philosophical theories of privacy: implications for an adequate online privacy policy. *Metaphilosophy*, *38*, 1, 1–22.
Teutsch, P., Piat, F. & Reffay, C. (2009). Anonymizing and sharing corpora of online training courses. *Workshop interaction analysis and visualization for asynchronous communication* (pp. 1–6).
The Economist. (2012, April 28). Consent 2.0. A better way of signing up for studies of your genes. *The Economist*. New York.
The White House (2012). *Consumer data privacy in a networked world* (p. 62). Washington: US Government, The White House.
Toch, E., Wang, Y. & Cranor, L. F. (2012). Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-adapted Interaction*, *22*, 1–2, 203–220.
Verbert, K., Manouselis, N., Drachsler, H. & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society*, *15*, 3, 133–148.
Voss, G. (2012). Gaming, texting, learning? Teaching engineering ethics through students' lived experiences with technology. *Science and Engineering Ethics*, *19*, 3, 1375–1393.