(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

LEA in Private: A Privacy and Data Protection Framework for a Learning Analytics Toolbox

Christina M. Steiner, Michael D. Kickmeier-Rust, and Dietrich Albert
Knowledge Technologies Institute
Graz University of Technology, Austria
christina.steiner@tugraz.at

ABSTRACT: To find a balance between learning analytics research and individual privacy, learning analytics initiatives need to appropriately address ethical, privacy, and data protection issues. A range of general guidelines, model codes, and principles for handling ethical issues and for appropriate data and privacy protection are available, which may serve the consideration of these topics in a learning analytics context. The importance and significance of data security and protection are also reflected in national and international laws and directives, where data protection is usually considered as a fundamental right. Existing guidelines, approaches, and regulations served as a basis for elaborating a comprehensive privacy and data protection framework for the LEA's BOX project. It comprises a set of eight principles to derive implications for ensuring ethical treatment of personal data in a learning analytics platform and its services. The privacy and data protection policy set out in the framework is translated into the learning analytics technologies and tools that were developed in the project and may be used as best practice for other learning analytics projects.

Keywords: Learning analytics, ethics, privacy, data protection

1 INTRODUCTION

Learning analytics are key emerging technologies in education (Johnson, Adams Becker, Estrada, & Freeman, 2014) and their potential to optimize educational planning and processes, to inform and tailor teaching, and to inform and support learning has been highlighted by many authors (e.g., Ferguson, 2012; Greller & Drachsler, 2012; Long & Siemens, 2011). Educational institutions have always analyzed the data of their students to some extent. Learners today have access to a multitude of learning tools, applications, and resources, they enhance their learning experience in virtual or simulated environments, and they connect to others through social media. All those interactions and resources may be captured and those multi-faceted learning processes can (potentially) be analyzed using big-data analytics techniques (Pardo & Siemens, 2014).

With the advent and increasing capacity and adoption of learning analytics, an increasing number of ethical and privacy issues arise too. For example, the evolution of sensors and new technologies enables a multi-faceted tracking of learners' activities, locations etc., such that more and more data can potentially be collected about individuals, who are oftentimes not even aware of it. Data collection and use under such circumstances is, of course, ethically and legally questionable (Greller & Drachsler, 2012). Ethical issues in learning analytics include the collection of data, informed consent, privacy, de-

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

identification of data, transparency, data security, interpretation of data, data classification and management, as well as potential harm to the data subject (see e.g., Sclater, 2014b; Slade & Prinsloo, 2013). These issues have created some tension so far (Pardo, 2014). A clear and agreed upon set of guidelines with respect to the ownership of data and analytic models, rights, and responsibilities is required (Ferguson, 2012); at the moment there are no standard methods and procedures for informed consent, opting out, etc. The need for a clearly defined and uniform approach and code of conduct to appropriately deal with ethics and privacy in learning analytics is increasingly acknowledged (Berg, 2013).

LEA's BOX (www.leas-box.eu) is a research and development project funded by the European Commission that develops a learning analytics toolbox. In this paper, we outline the privacy and data protection considerations and the policy that has been formulated for the project to find a balance between learning analytics research and individual privacy. LEA's BOX takes an unprecedented initiative in learning analytics, by researching and making available a whole range of tools to educational practitioners for customizing, performing, and using competence-centred, multi-source learning analytics and open learner models. This novel and flexible modular approach to learning analytics requires a comprehensive consideration of privacy and data protection aspects, for which different sources of information have been used as a starting point. Individual existing approaches, though, do not sufficiently cover all relevant ethical values and do not harmonize the research and design perspective. For establishing requirements on the implementation of the LEA's BOX methodologies and technologies, therefore, a new, integrated framework for privacy and data protection has been defined, which may also be re-used in other learning analytics projects.

This paper is structured as follows: Section 2 summarizes ethical and privacy issues in learning analytics. Section 3 then outlines existing approaches and frameworks for dealing with these topics, and in Section 4, an overview of privacy and data protection regulations is given. Section 5 presents the LEA's BOX privacy and data protection framework, which synchronizes these resources and integrates the input from an external ethics expert. The framework comprises a set of privacy, data protection, and ethical principles, which define requirements for the project's learning analytics research, design, and development. Finally, conclusions on the presented work are outlined in Section 6.

2 ETHICAL AND PRIVACY ISSUES IN LEARNING ANALYTICS

Relevant privacy and data protection aspects and ethical issues in learning analytics can be summarized and grouped into the following overlapping areas (Campbell, DeBlois, & Oblinger, 2007; Pardo & Siemens, 2014; Sclater, 2014b; Slade & Prinsloo, 2013; Willis, 2014):

- **Privacy**: The possibility that actions and personal data are tracked causes concern for users. On the other hand, users may not be fully aware of the data being collected or exchanged when using technology services.
- **Informed consent, transparency, and de-identification of data**: This relates to the question of

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

whether an individual needs to give consent to data collection and analysis, the obligation to inform about the data being collected and analyzed, and the relevance and implication of de-identification of data.

- **Location and interpretation of data**: Learning activities today are usually spread over different tools and locations; learning analytics aims at bringing together these different data sources for a more complete picture of learning. Questions arise on the implications of using multiple and non-institutional sources, and whether the data is representative of a particular student.
- **Management, classification, and storage of data**: This area relates to questions of data management, access rights, and the measures and level of data protection needed. It also involves the issue of the temporality of data.
- **Data ownership**: This relates to the question of the ownership of the data collected, the analytics models, and the analytics output. It also links to the aspect of outsourcing and data transfers to third parties and related regulations and responsibilities.
- **Possibility of error**: Analytics results are always based on the data available and the outputs and predictions obtained may be imperfect or incorrect. Questions on the ramifications of making an error and the implications of ineffective or misdirected interventions arise because of faulty analytics results.
- **Role of knowing and the obligation to act**: Learning analytics brings new knowledge and insights about learning. Does this new knowledge entail the responsibility to act on this information, and what are the ramifications of action or inaction?

3 EXISTING APPROACHES

3.1 Big Data and Ethics

Privacy and ethics have evolved into important and pressing topics not only in learning analytics but also in analytics and big data in general (e.g., Richards & King, 2014; Schwartz, 2011). “Big data poses big privacy risks,” as Tene and Polonetsky (2013, p. 251) put it. Data has become a resource of important economic and social value and the exponentially growing amount of data (from a multitude of devices and sensors, digital networks, social media, etc.) that is generated, shared, transmitted, and accessed, together with new technologies and analytics available, opens up new and unanticipated uses of information. The collection of large and multifaceted data sets and the new possibilities of their use lead to growing privacy concerns. The disclosure and use of personal data is increasingly associated with fear, uncertainty, or doubt (Dirndorfer Anderson & Gardiner, 2014). Users are concerned about privacy and that large amounts of their personal information may be tracked and made accessible for other purposes to other users (Kobsa, 2007). On the other hand, social media are deeply integrated into users’ daily lives and routines (Debatin, Lovejoy, Horn, & Hughes, 2009) and people, in fact, are willing to share many personal details via these networks. Privacy attitudes and privacy behaviours, thus, often differ (Stutzman & Kramer-Duffield, 2010), thus leading to the “privacy paradox” (Barnes, 2006). This is evident when comparing users’ self-reports about their understanding of caution regarding privacy

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

settings and their usual, unconcerned behaviour of just keeping default settings instead of updating them to their needs and preferences (Debatin et al., 2009). So, privacy attitudes and privacy behaviours do not necessarily conform — people may not act according to the privacy preferences they claim. Usually they seem unconcerned about data protection and privacy until it is breached (Spiekerman & Cranor, 2009). Importantly, users' concerns about privacy also differ depending on 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 minimization and individual control and context need to be somewhat relaxed in a big data context. They must be considered not only from an individual but also from 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.

The analytics process — regardless of the specific domain of application — aims to convert data into actionable knowledge and, in general, includes data collection (gathering information), integration and analysis (aggregating data from multiple sources and examining the data for patterns), decision making based on the information gained (acting on the results of integration and analysis stage), and review and revision of analytics models. Schwartz (2011) has 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

Besides specifying these generic principles, Schwartz in particular argues that at different stages of the analytics process different ethical considerations are relevant. Accordingly, the rules of how to tackle these challenges need to be tailored to each analytics stage — always aiming at maximizing good results and minimizing bad ones for the persons whose data is processed. In data collection, care needs to be taken about the kind of information, in particular avoiding the collection of sensitive data. For data integration and analysis, a sufficient data quality should be ensured and anonymization should be done, as appropriate. In decision making, obviously, the analytics results on which decisions are based must be reasonably accurate.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

3.2 Ethical Frameworks in Learning Analytics

Researchers have started to discuss ethical and privacy issues and principles specifically for learning analytics as a basis for advancing in this direction. Still, although many authors mention ethical issues, only a few coherent approaches exist that elaborate ethical challenges in more detail and attempt to define a framework to guide institutions, researchers, and developers in the application of learning analytics (Slade & Prinsloo, 2013).

The topics of privacy and ethics are directly related to aspects of trust and accountability (Pardo & Siemens, 2014). A rational and sensible approach to dealing with privacy and ethics is therefore needed to leverage learning analytics technologies in terms of broad practical adoption, acceptance, and growth. Reflection and deliberation on ethical questions need to be aligned with technical innovation in analytics because the slow pace of law may not able to match the speed of innovation. Nevertheless, existing approaches on ethics in learning analytics commonly and understandably ground their discussion around legal understandings of privacy (Willis, 2014).

One possible approach to elaborating the ethical issues of learning analytics is to determine, analyze, and manage the risks of implementing a learning analytics project. Stiles (2012) identifies a set of specific areas and associated risks. Data protection is considered a key risk to be addressed, including aspects of privacy, security, governance, and compliance. To ensure privacy, security, quality, and auditability of data, an appropriate level of control needs to be implemented (i.e., data and information governance, for example through policy or a checklist). Compliance with legal requirements on data privacy and security creates increased data awareness, quality, and protection (i.e., data and information compliance). The risks associated with these areas need to be appropriately addressed for the implementation and use of analytics in an educational organization.

Greller and Drachsler (2012) consider ethical and legal aspects in their generic framework for learning analytics under the dimension of “external constraints.” Apart from ethical, legal, and social constraints, they also consider organizational, managerial, and process constraints as relevant components of this dimension. These external limitations can be categorized into conventions (ethics, personal privacy, and other socially motivated constraints) and norms (restrictions by law or mandated standards and policies). This makes clear that there is a reasonable distinction but close linkage between ethics and legal regulations: Ethics deals with those measures that are morally allowed; 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 law and depend on ideological assumptions and epistemologies (Slade & Prinsloo, 2013). Since many legal regulations are based on ethics, an ethical position needs to be applied for interpreting the law (Sclater, 2014a). Kay, Korn, and Oppenheim highlight that given the mission and responsibilities of education, “broad ethical considerations are crucial regardless of the compulsion in law” (2012, p. 20).

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

Kay et al. (2012) outline that learning analytics is an area of conflict between assuring educational benefits, business interests, and competitive pressure on educational institutions, and expectations of born-digital generations of learners. They postulate four key principles for good practice with respect to ethical aspects and analytics when dealing with these conflicts:

- **Clarity**: Definition of purpose, scope, and boundaries.
- **Comfort and care**: 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, Campbell, and Pistilli (2013) refer to this area of conflict and a need for balancing between faculty expectations, privacy legislation, and an educational institution's philosophy of student development when dealing with ethical questions. They do not define specific guidelines on different ethical issues, but suggest using the Potter Box — a flexible ethical framework commonly applied in business communications — to deal with the ethical dilemma of analytics. This approach, in fact, only provides a thinking framework for analyzing a situation but does not provide one clear solution to ethical dilemmas. The Potter Box foresees four universal steps when taking ethical decisions on specific questions, as described in Table 1.

Table 1: The Potter Box.

Definition: The empirical facts of a given situation are clearly defined without judgement.	**Loyalties**: Loyalties are chosen; for example, people affected by a situation (application of learning analytics), entities acting on the gained information, responsible persons in case of failure, etc.
Values: Values representing conventions, rights, and beliefs are identified and compared (e.g., moral values, professional values). Differences in perspectives of stakeholders involved can be analyzed.	**Principles**: A set of ethical principles applicable to the situation in question (e.g., Mill's principle of utility — "Seek the greatest happiness for the greatest number") is identified and considered.

Slade and Prinsloo (2013) take a socio-critical perspective on the use of learning analytics in their article elaborating on ethical issues. They propose a framework of six principles to address ethics and privacy challenges in learning analytics:

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

- **Learning analytics as a moral practice**: Focus should not only be put on what is effective, but on supporting decisions about 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 dynamicity of data is acknowledged, thus providing only a snapshot view 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.
- **(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 (2014) analyze ethical and privacy issues in learning analytics research in educational institutions and take into account how privacy and ethics are addressed in other contexts. They identify a set of four principles that aggregate numerous issues; they are intended to serve as a basis for setting up appropriate mechanisms for meeting social, ethical, and legal requirements when developing and deploying learning analytics. The four principles are:

- **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 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.

Another recent, general approach is the code of practice for learning analytics published by Sclater and Bailey (2015). It has been developed based on an extensive literature review of legal and ethical issues in learning analytics (Sclater, 2014a) and addresses the following eight themes:

- **Responsibility:** Identifying responsibility for the data and data processing for learning analytics in an institution.
- **Transparency and consent**: Ensuring openness on all aspects of using learning analytics and meaningful consent.
- **Privacy:** Ensuring protection of individual rights and compliance with data protection legislation.
- **Validity:** Ensuring the validity of algorithms, metrics, and processes.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

- **Access:** Providing data subjects access to their data and analytics.
- **Enabling positive interventions:** Appropriate handling of interventions based on analytics.
- **Minimizing adverse impacts:** Avoiding potential pitfalls.
- **Stewardship of data:** Appropriate handling of data.

3.3 General Ethical and Privacy Guidelines or Models

The OECD guidelines are a relevant source of basic principles when seeking guidance on how to deal with privacy issues in analytics technologies and other systems (Spiekermann & Cranor, 2009; Tene & Polonetsky, 2013). In 1980, the OECD (Organisation of Economic Cooperation and Development) provided the first internationally agreed collection of privacy principles, 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 (Levin & Nicholson, 2005). The basic OECD (2013b, pp. 14–15) principles are as follows:

- **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 no 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 those 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 unauthorized 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, 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 that give effect to the above principles.

Although not binding for OECD members, the guidelines have gained legal significance and served as a basis for privacy legislation in Europe (European Parliament, 1995; Levin & Nicholson, 2005;

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

Spiekermann & Cranor, 2009). The Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (OECD, 2013b), an update of the original version from 1980, keeps the original "Basic Principles" of the guidelines, while modernizing considerations on transborder 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, health care, utilities, logistics and transport, and public administration), without any specific reference, however, to learning or academic analytics. Privacy protection is indicated as one of several areas that need public policies and practices to leverage the potential of big data. Privacy protection enabling open, secure, reliable, efficient, and cross-border flows of data is needed, while at the same time reducing privacy risks and enhancing responsible behaviour in the use of personal data.

Based on the framework of the OECD guidelines, the Federal Trade Commission of the United States (1998) has defined the Fair Information Practice Principles (FIPP), which specify concepts of fair information practice in 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:

- **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 if the user has notice.
- **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. Traditionally two approaches may be taken: opt-in or opt-out.
- **Access/Participation**: This principle relates to giving users the possibility of accessing their data and ensuring that the data is accurate and complete.
- **Integrity/Security**: Data needs to be accurate and secure; appropriate steps and safeguards must be taken to ensure that, such as using reliable data sources and cross-referencing multiple sources.
- **Enforcement/Redress**: To ensure compliance with privacy protection principles, enforcement and redress mechanisms through self-regulatory regimes, legislation creating private remedies for users, or government enforcement is required.

Ethical issues in learning analytics may also be considered in the context of the history of Internet research ethics, where the attempt of finding a balance between harms to the individual and greater scientific knowledge has been made (Slade & Prinsloo, 2013). The Association of Internet Researchers

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

provides a set of ethical guidelines for decision making (Ess & AoIR, 2002; Markham & Buchanan, 2012). These aim to provide researchers with a basis for conducting their research in an ethical and professional manner; they have been pointed to by learning analytics researchers as a valuable source for dealing with privacy issues in the application of learning analytics.

3.4 Ethics by Design

Since learning analytics involves technology, ethics and privacy concerns should not be considered from a purely legal perspective, but need to be addressed from a technological point of view (Pardo & Siemens, 2014). One way of ensuring that is to take privacy and ethics, in general, into account up front during the design process of learning analytics tools.[1] This approach, called "privacy by design," "value-sensitive design," or "ethics by design," is increasingly acknowledged in learning analytics research (e.g., Bomas, 2014; Scheffel, Drachsler, Stoyanov, & Specht, 2014) and has been taken up in LEA's BOX.

Value-sensitive design or ethics by design corresponds to the approach of incorporating ethical and legal requirements and considerations in the design and development process, i.e., making them an inherent part of the software being created (Friedman, 1997). This approach deals with design principles and guidelines so that the software itself follows ethical rules or supports humans in following ethical rules (Gotterbarn, Miller, & Rogerson, 1997; Gotterbarn, 1999). Privacy by design more concretely focuses on privacy engineering and developing guidelines for designing privacy-friendly systems (Cavoukian, 2011). Spiekermann and Cranor (2009) have carried out a privacy-requirements analysis applicable to a wide variety of systems that identifies activities typically performed by information systems and their impact on user privacy. 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 anonymization) can be distinguished (Spiekermann & Cranor, 2009).

4 PRIVACY AND DATA PROTECTION REGULATIONS

Legislation on privacy and data protection is regulated in national and international laws that address the disclosure or misuse of information held on private individuals. Regulations began to appear in countries with high Internet use (Pardo & Siemens, 2014). Examples are 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 (2000), the Australian Privacy Act and Regulation (1988, 2013), and the US Consumer Data Privacy in a Networked World (The White House, 2012). The Family Educational Rights and Privacy

[1] e.g. https://www.privacybydesign.ca/

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

Act or FERPA (2004), a US federal law, specifically applies to education, i.e., the privacy of student education records. This law allows the use of data on a need-to-know basis and provides parents with certain rights of access to their children's education records.

In parallel with legislative efforts for data protection, non-profit organizations have evolved that aim to defend users' digital rights (Pardo & Siemens, 2014); for example, the ARGE DATEN Privacy Service[2] in Austria, or the Electronic Frontier Foundation[3] and Privacy Rights Clearinghouse[4] in the US.

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" (FRA, 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 paper (an overview and comparison between international privacy laws and approaches is given, for example, in Levin & Nicholson, 2005, and Movius & Krup, 2009). Instead, only reference to the relevant European legislation shall be given, which aims to provide a unified initiative for EU members.

4.1 European Regulations

The transfer of personal data between countries in the EU is necessary in the 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.[5] The application of this European legislation is monitored by national supervisory authorities.

European data protection legislation considers the protection of personal data as 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 balance a high level of protection of individual privacy and the movement of personal data within the European Union. It applies both to data collected and processed automatically (e.g., computer database) and in non-automated ways (traditional paper files). Each member state is to apply the provisions nationally.

[2] http://www.argedaten.at/
[3] https://www.eff.org/
[4] https://www.privacyrights.org/
[5] http://ec.europa.eu/justice/data-protection/index_en.htm

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

The EU Data Protection Directive defines rules for international transfers of personal data to countries outside the EU/EEA. Such data transfer may only be done if an adequate level of protection is guaranteed and standard contractual clauses have been developed for this purpose. A specific directive for data communication has been extended in the electronic communication sector[6] (ePrivacy directive) to address the specific requirements regarding privacy and data protection for the Internet and electronic messaging services. This directive helps ensure that users can trust the services and technologies they use for electronic communication. The main regulations covered by the Directive apply to spam, ensuring the user's consent, and the installation of cookies.

The European Commission is in the process of reforming the data protection legislation to further enforce protection of personal data by updating and modernizing data protection rules.

5 THE LEA'S BOX PRIVACY AND DATA PROTECTION FRAMEWORK

The LEA's BOX project focuses on researching and developing novel approaches to competence-centred learning analytics and visualizations. Based on psycho-pedagogical knowledge representation frameworks and the review of existing learning analytics and educational data mining approaches, as well as open learner modelling techniques, conceptual research on analytics and visualization methods is carried out and translated into the technical development and integration of a toolbox of services for empowering teachers and learners. Such a novel learning analytics initiative requires a sound foundation to deal with the aspects of privacy and data protection. Adopting one of the existing approaches, as outlined in the previous sections, was not appropriate for LEA's BOX. Clearly, the approach taken in the project must go beyond pure compliance with national and international legislation[7] (cf. Section 4) but also consider relevant ethical and social aspects on a broader basis. Available guidelines and principles in the field of information practice and data management in general (cf. Section 3.3), or more specifically in the field of big data (cf. Section 3.1), are useful starting points for a more extensive ethical approach. They do not, however, sufficiently account for the specific circumstances given in an educational context, like for example settings with underage students or potential impact of learning analytics results on educational decisions and on school or academic careers. Some approaches have been defined specifically for learning analytics applications (cf. Section 3.2), but these are mainly focused on the deployment of established learning analytics in educational practice. In a research and development project like LEA's BOX though, it is important to incorporate the perspective of learning analytics as a scientific ambition, with the goal of researching and validating new analytics methods and technologies.

The aim in defining the privacy and data protection policy for LEA's BOX has been to use the different information sources on ethical and privacy aspects and best practice as a basis and to integrate them,

[6] http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32002L0058

[7] in addition to the European Directive, the privacy and data protection regulations in Austria, the Czech Republic, Turkey, and the United Kingdom in particular had to be considered.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

considering project-specific aspects, into a coherent overall framework. The requirements imposed by this framework should go beyond outlining philosophical ideals, but should actually be applied as ethical principles and feed into the design and development of the project's technologies (see Figure 1). In line with Schwartz (2011), the requirements must represent an accountable approach reflecting the specific ethical and data protection issues relevant for the project. They must also provide an appropriate framework for researching and exploring the educational possibilities of benefitting from learning analytics without sacrificing privacy (Bomas, 2014).

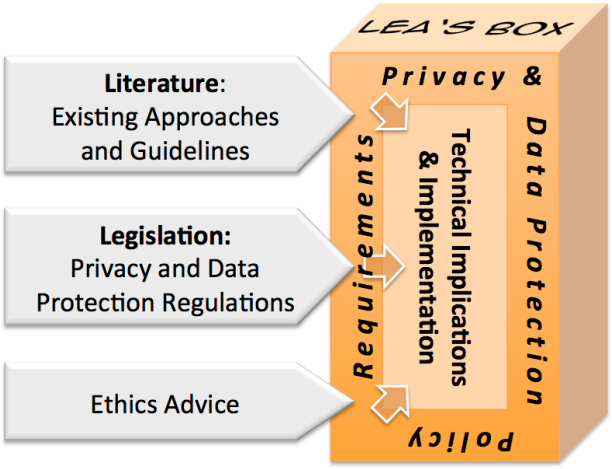

Figure 1: Privacy and data protection policy in LEA's BOX.

5.1 Ethics Advice

An ethics expert was involved in the research and development of LEA's BOX as an external privacy and ethics advisor. This expert is a representative for natural sciences on the Ethics Commission of the University of Graz, Austria. Aside from discussing the general ethical use of data, the importance and approaches to gathering data subjects' consent, and providing transparency, one particularly relevant topic evolved — the consideration of learning analytics as moral practice.[8] When researching new learning analytics approaches, as a first step the new methods and algorithms need to be tested and evaluated, which should not directly affect data subjects. This would imply that an ethical use of learning data means that the results of the analysis must not have any direct impact on the learners. Only at the second stage, after the methods have been validated, the implementation of consequences or interventions based on the analytics results should be done and evaluated. The need to ensure the validity of the data and analytics processes and their benefit to learners is also highlighted in the Jisc code of practice for learning analytics (Sclater & Bailey, 2015). This ethical perspective of validating learning analytics before using the results for decision making, in fact, conflicts with the moral value of an "obligation to act," commonly discussed in the literature (e.g., Campbell et al., 2007; Kay et al., 2012;

[8] H. Römer, personal communication, 27 November 2014.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

Willis et al., 2013), such as the idea of an ethical duty to act on the information gained from learning analytics, like information about students at risk of dropping out. This position of validating a new learning analytics approach is directly related to the consideration that learning analytics may yield results that are not perfect or valid, but may be inaccurate or even incorrect (e.g., van Harmelen & Workman, 2012). This is in line with Schwartz (2011), who claims for big data, in general, that decision making in the analytic process needs to be grounded on reasonably accurate analytic output.

5.2 Privacy and Data Protection Principles

A set of principles relating to privacy, data protection, and ethics has been identified, which form a comprehensive ethical and information practice framework for LEA's BOX. These principles have been derived from a harmonization of the aspects of data protection and privacy covered by existing guidelines and approaches, national and European regulations, complemented by the discussion points of the ethics advice. Ethical and privacy principles from these different resources have been mapped into the eight principles derived for LEA's BOX.

5.2.1 Data Privacy

The first and overarching requirement for LEA's BOX is data privacy, in line with the fundamental right to data protection as reflected in national regulations and the EU data protection directive (Rodotà, 2009). Collection and use of personal data must be fair and provide appropriate protection of privacy. Information on privacy and data protection practices should be available and easily understandable.

Users who feel that their privacy is endangered may show resistance (Greller & Drachsler, 2012). To reassure them that their data is used in an acceptable and compliant way, policies and guidelines to protect the data from abuse are needed and need to be communicated. The protection of data with respect to data collection and analysis is ensured not just by legislation but also by additional institutional privacy regulations (Campbell et al., 2007), as represented by the privacy principles at hand.

In terms of the technical development for LEA's BOX, this means nothing less than designing and building data-sensitive educational apps equal to the well-established principles of other critical online solutions, such as online banking or medical platforms. Depending on the concrete use case, this may also include using transaction numbers (TAN) for accessing delicate information. To address the concrete need for communicating safety as a means to gain trust by the users, all implemented strategies must be clearly displayed.

5.2.2 Purpose and Data Ownership

Adequate specification and documentation of the purpose of data processing must be ensured in LEA's BOX at any stage and made available. The purpose and boundaries of any learning analytics application must be clearly defined and available before processing begins since "processing personal data for undefined and/or unlimited purposes is unlawful" (FRA, 2014, p. 68). In essence, considering learning

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

analytics as a moral practice, it should aim to support learners (e.g., Slade & Prinsloo, 2013; The Open University, 2014). When researching new learning analytics methods, though, establishing and ensuring reasonable accuracy of results (i.e., creating truly actionable knowledge) represents the ethical standard to be addressed first[9] (Schwartz, 2011), before dealing with ethical questions about the responsibility to act or not act based on the new knowledge gained (e.g., Willis, 2014).

Another relevant ethical aspect is data ownership. It has been argued that there is a lack of legal clarity when considering learning analytics applications (Greller & Drachsler, 2012). Traditionally, the data collected about a person (i.e., before anonymization) belongs to the owner of the data collection tool (data client). Meanwhile, there is a trend of considering users to be the owners of the data collected about them and that institutions are borrowing data for a clearly stated purpose. In learning analytics, things get more complicated very quickly, since usually data from a whole population of learners is used to produce a prediction model. The question then arises about who the owner actually is (Pardo, 2014). Even if the raw personal data is owned by the user, what about the information derived from it? While there is no issue of copyright for raw learning data, database rights may be relevant for enhanced learning data (e.g., collations of data, prediction models). The owner of any IPR is typically the institution that has collected (and enhanced) the data (Kay et al., 2012).

The question of data ownership is also further complicated when integrating learning data from different sources, which may potentially mean different organizations/data clients. It has been argued that to fully exploit the potential of learning analytics and build a holistic picture of an individual's learning (e.g., Ferguson, 2012; Dyckhoff, 2011), data integration is needed — e.g., institutionally held student data with learning data from educational tools. In fact, the concept of data ownership may not be most appropriate and helpful; more relevant are the notions of data controller and data processor as used in data protection regulations (Sclater, 2014b). The data controller is a natural or legal person, or an authority, that processes personal data and determines the purpose of processing. The data subject has the right to be provided with information about the identity of the data controller (including contact details) and purposes of processing. A data processor is a separate legal entity, who processes personal data on behalf of the controller (FRA, 2014).

Technically, the LEA's BOX solution is to disconnect any "unproven" research results and the real-world application of various tools and learning analytics solutions. From a research perspective, we may conduct studies with leading edge ideas in the field, however, by following the highest standards for a code of conduct and with an imperative avoidance of having research data influencing real students' lives. With respect to data ownership and access rights, the LEA's BOX data architecture foresees defining access rights to data based on ownership. Concretely, for each data source defined in the system (a central learning solution like Moodle, a self-assessment, or an external learning app) the ownership can be set. By this means, a broad range of access rights can be defined. For example, a

[9] H. Römer, personal communication, 27 November 2014.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

student might not want to access the teacher data from an afternoon's learning app; on the other hand, the student might allow teachers to access the results of preparatory online quizzes, but in an anonymous way (perhaps to tailor exam preparation lessons). Of course, in certain scenarios the teacher has full access rights to data, for example when marking homework. Again, a key feature is that rules and options are visible to all users and that they cannot be hidden from the system.

5.2.3 Consent

LEA's BOX must apply appropriate techniques for gathering consent from students and parents as a legal basis for processing personal data. Informing users about the collection of their data and gathering their consent needs to be recognized as a basic ethical principle and procedure (Greller & Drachsler, 2012). It has been argued that in learning analytics there should be virtually no reason to waive informing users about the use of their data; therefore, a clear policy of informed consent must be drafted (Slade & Prinsloo, 2014). According to current privacy legislation, consent must also be implemented for the use of cookies.

Consent needs to be free, informed, specific, and given unambiguously. Sufficient information needs to be provided to the data subject to ensure that he/she is clearly informed about the object and consequences of consenting before making the decision. Information needs to be precise and easy to understand. Non-explicit consent based on inactivity (passive consent from parents) is ambiguous and should be avoided[10] (FRA, 2014). Although the European regulations do not explicitly mention the right to withdraw consent at any time, it is widely presumed that such a right exists (FRA, 2014).

Since the principle of consent refers to giving data subjects the option to agree/disagree to data collection and application, the information provided as a basis for gathering consent should establish a balance between allowing research and protecting users from potential harm. Thus, it may refer to "a broad definition of the range of potential uses to which a student's data may be put" (Slade & Prinsloo, 2014).

As opposed to research settings, within which a clear code of conduct applies, in real life applications of learning analytics, consent is often problematic. Usually, teachers and instructors do have the right and the need to access and assess performance data of learners and students have no option to opt out. Technically, our solution is to address the problem through the access rights to particular data sources, as explained above. Consent, however, is very much related to sound information about what is recorded and how data can influence one's own advancement. Thus, LEA's BOX will provide consent templates tailored to specific age groups and levels of expertise that can be linked to the setup and maintenance of access rights to the various data sources.

[10] H. Römer, personal communication, 27 November 2014.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

5.2.4 Transparency and Trust

Transparency is one of the most problematic ethical issues in learning analytics (Pardo & Siemens, 2014). While privacy legislation requires learners' consent for data collection, the principle of transparency goes beyond that. Data subjects (i.e., usually learners, but also teachers) should be given notice about what kind of data is gathered and recorded, and should be provided with information on how the analytic processing is done. Transparency also means providing information on data management procedures, on how data is dealt with after its primary purpose, and whether information is transmitted to outside an institution. Users should, however, not only be informed about how their data is used outside an educational institution, but also within the institution (Slade & Prinsloo, 2013). In addition, data subjects should also be made aware of the possible outcomes of the data application and the data protection measures taken (Willis & Pistilli, 2014).

The following information is essential to properly informing data subjects (Federal Trade Commission, 1998): 1) the entity collecting the data, 2) the uses to which the data will be put, 3) potential recipients of data, 4) the type of data collected, 5) the data collection method, 6) consequences of refusal, and 7) measures taken to ensure data quality and security. Frequently, information on consumer rights is also included. In the case of learning analytics, an appropriate and understandable description of the analytic models/procedures should be provided.[11] Data subjects should be able to understand what is happening with their data (FRA, 2014).

Informing users about what kind of data is recorded and for what purpose is not only an important ethical and legal privacy principle in LEA's BOX, but also key to fostering trust in data subjects — both for learning analytics, and for the educational institution applying it. If users trust the learning analytics technology because they understand the data application and the (potential) value and usefulness it may have to them, users' experience and acceptance is considerably enhanced (Pardo & Siemens, 2014). As a result, the application of the principle of transparency should also include information on the potential benefits (or harms) due to the data application in order to raise users' awareness and understanding of the learning analytics approach and, potentially, involve them as active agents in the implementation of learning analytics.

In LEA's BOX, the key element of transparency and trust is the OLM, the Open Learner Model feature. While in many learning analytics applications, the results of complex analyses are opaque, the OLM tries to make transparent the logic of analyses and the ways that results have been accumulated in simple, user-centric ways. First, results are presented in a simple, easy to understand fashion; if a particular user is interested in learning more about the results and their basis, he/she can drill deeper into the analyses and get feedback about the individual data sources that contributed to a particular result. In addition, the OLM offers negotiation features that allow students to influence and negotiate — and perhaps complain — about the results of analyses (see Bull 2012; Bull and Kay, 2010 for details). Such features

[11] H. Römer, personal communication, 27 November 2014.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

empower learners to be an active, serious part of the assessment and appraisal process, which in the end leads to a stronger trust and belief in the system.

5.2.5 Access and Control

In addition to gathering users' consent and providing transparency of when and how data is collected and analyzed, data subjects should be given control of their own data. This forms the fifth principle of our framework. Access and control mean that users should be given access to the data collected about them, and the opportunity to correct that data if necessary. The principle of access and participation is reflected in legislation as a right of the data subject. While giving access is completely in line with the idea of transparency, the aspect of modifying data is somewhat challenging in learning analytics and only applies to certain types of data — i.e., data from plain observations, but not necessarily summaries or results obtained from data. Procedures for correction or deletion of personal data, if inaccurate, misleading, or out-dated, need to be provided to users.

In fact, some authors have even claimed that in order to establish a culture of participation, learners must be considered agents sharing responsibility for the accuracy, maintenance, and currency of their student data; they may even be actively involved in the implementation of learning analytics and help in shaping interventions (Slade & Prinsloo, 2013; The Open University, 2014). This requires a plan and clear communication with learners.

Dashboards and open learner models are approaches of visualizing learning analytics data and results. They are often an inherent part of learning analytics approaches as instruments for reporting and fostering reflection (Bull & Kay, 2010; Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). These visual approaches provide users with access to the data whenever and for however long they want, thus offering transparency to data subjects on the data collected about the learning process (Pardo & Siemens, 2014). More recent approaches, such as negotiated user models, reflect the idea of student control, since the open learner model is used to interactively negotiate and potentially update the content of the learner model. In LEA's BOX, research on open learner model communication and negotiation can be considered an application of the ethical principle on access and participation.

Access and control over data need to be governed by technically implementing appropriate authentication mechanisms and establishing an access right structure. Simple and understandable procedures for indicating inaccurate data, for updates or corrections, and for verifying information need to be implemented in the management and maintenance of data files. Technically, in LEA's BOX this point is very much related to the OLM; at each point in time, users can access the data, as aggregated for the OLM, and the resulting analyses. Based on the user rights and ownership rules described above, users may add or undock certain information, building the basis for the learner model.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

5.2.6 Accountability and Assessment

Principles of data protection can only work when there are appropriate mechanisms to enforce and redress them (FRA, 2014). The institution, department, or person responsible or accountable for a learning analytics application and its proper functioning needs to be identified. In LEA's BOX, a clear structure of the responsibilities of individual partners and persons has been established from the outset of the project.

In addition, the learning analytics process should be evaluated in order to refine data collection, management, and analysis (Pardo & Siemens, 2014). The overarching goal of learning analytics is to better understand learning processes and to optimize and support learning and teaching. This can only be achieved by ensuring the correctness of the data and the analytics algorithms. Constantly reviewing and adjusting analytics methods will increase the accuracy of results and the suitability of the learning analytics process to maximize its impact (Pardo, 2014; van Harmelen & Workman, 2012). The importance of the review and revision stage in analytics is also highlighted by Schwartz (2011) who also refers to assessing the impact of using analytics based on stakeholder trust. In LEA's BOX, the continuous assessment, refinement, and enrichment of learning analytics methods and tools is the basis for ongoing improvement. In addition to this validation and elaboration of data processing, impact on learners and teachers (e.g., in terms of acceptance) will be addressed in a series of pilot and evaluation studies.

In particular, at the very heart of the project lies collecting as broad a set of information from various sources as possible and interpreting this information in a cautious, probabilistic manner. Each data source contributes — by design — a certain weight or probability to a total view. All computations and analyses are made with the fundamental premise that data sources or analyses may be wrong (or done on the wrong basis). And all analyses can be accessed and negotiated by users. This is very much in line with the idea of a probabilistic competence-based knowledge space theory (Albert & Lukas, 1999; Heller, Ünlü, & Albert, 2013) that aims to separate latent competencies and observable performance. However, we must highlight that it is a delicate and sometimes impossible process to convey the scientifically and practically complex ideas and theories upon which analyses are based. This is specifically true for younger students. Practically, we suggest clearly and continuously highlighting this particular responsibility in guidelines and manuals for users, especially teachers/instructors.

5.2.7 Data Quality

According to different ethics frameworks, an appropriate quality of data needs to be ensured (e.g., Federal Trade Commission, 1998; OECD, 2013b; Pardo & Siemens, 2014). Data needs to be representative, relevant, accurate, and up-to-date. Information that is not up-to-date cannot be assumed reliable in reflecting the status of a learner; it may thus lead to wrong conclusions (The Open University, 2014). Sharing responsibility for the accuracy and maintenance of personal data between the educational institution and the learner (compare "Access and Control") is considered reasonable for ensuring an adequate level of data quality.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

Especially when gathering and combining data from multiple sources, care needs to be taken to ensure reliability. The data collected may provide an incomplete picture of the learning process, as it only represents a snapshot in time and context. Bias and stereotyping need to be guarded against by constantly taking into account the incomplete and dynamic nature of individual learning and experience (Slade & Prinsloo, 2014).

Besides an adequate quality of raw learning data, data must be used wisely for carrying out integration and analysis in LEA's BOX. Any interpretation, enhancement, or manipulation of data with the aim of extracting meaning must be grounded in sound techniques, the analytics models must be transparent and available for review and testing.

Technically, the system is designed to separate competence/domain models, user models, sources of evidence (the "data" actually), and the cautious stochastic link between them. All models may be erroneous, so it is important to avoid any overestimation of a particular model or data source. All learning analytics algorithms and visualizations in LEA's BOX account for this prime foundation.

5.2.8 Data Management and Security

In general, personal data needs to be treated and managed in a sensitive and ethical way in LEA's BOX. Data must be kept protected and secure at different levels and by adequate measures in accordance with applicable jurisdictions. Accountability, thus, requires safeguards for data protection; compliance of data processing with data protection regulations needs to be demonstrated (FRA, 2014).

Appropriate measures need to be taken to protect the data against unauthorized access, loss, destruction, or misuse. This includes a clearly defined policy of who is authorized to access the data, to which parts of the data and the application, and which kinds of data operations are allowed (Pardo & Siemens, 2014). Processes for redress need to be provided to users in case of any unauthorized access or use of personal data. Preservation and storage of data needs to be aligned with national and EU regulations.

In line with this principle of data management and security, the effective governance and stewardship of data should be ensured and a clear and transparent structure of data shall be established in LEA's BOX. Security thereby needs to involve measures on the managerial and technical levels (Federal Trade Commission, 1998; FRA, 2014). On the managerial level, internal organizational rules should be established that cover, for example, regular information to employees about data security rules, obligations of confidentiality, a clearly defined structure of responsibilities and competencies in data processing and transfer, training on effective security precautions, and so on. Technical measures for data security relate to having the right equipment (hardware and software) in place, encryption in data transmission and storage, using passwords to limit access, data storage on secure servers, et cetera.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

6 CONCLUSION

Stephanie Moore (2008) highlighted that ethics is a critical aspect, however hard to tackle, because it is full of variability, contradicting viewpoints, and squishy definitions. Specifically in the context of designing, developing, and deploying educational software and in the context of researching learning, individual beliefs, values, and preferences influence the scientific work. Our data protection and privacy framework provides the foundation for a proper code of conduct; in particular, it requires that technology and tools developed in the project, and any third-party technology used, are in line with these foundations. The principles defined form the requirements for LEA's BOX, which are then translated into concrete technical implications and actual implementation. Thus, we transfer the respective principles into an approach of "ethics by design."

Despite the ethical challenges of learning analytics in general, and in the context of a research project developing novel tools and algorithms in particular, "education cannot afford not to use [big] data," to put it in the words of Sharon Slade and Paul Prinsloo (2013, p. 34). In the context of this complex and sensitive field, this paper cannot claim to be complete; for example, critical further aspects concern tracking IP addresses, accessing individual data as done by many Smartphone apps (e.g., GPS location), the identifiability of users among each other, or access to webcams or chat functions.[12] Still, the established framework provides the project's "personal" code of conduct, strengthens our "personal" awareness, and derives a number of concrete technical requirements. The framework is also considered relevant to learning analytics on a larger scale and may be adopted as a starting point and theoretical basis for other learning analytics initiatives. While the way in which these principles are actually applied and implemented may take different forms, compliance with current laws and regulations must be ensured at any stage of the project as a main requirement of privacy and data protection. The principles defined in our framework must be aligned with the very specific context of the concrete learning analytics application in question. As Tene and Polonetsky put it, the "levers … must be adjusted to adapt to varying … conditions" (2013, p. 242).

Overall, it is very difficult (if not impossible), to translate a general ethical mind-set into direct recommendations for technical designs and architectures. In summary, we can highlight the following aspects: 1) use up-to-date security standards and proper data encryption and anonymization strategies, 2) design clear and well-documented ownership rules and access rights to data and display them up front and in a suitable way for all user groups. In addition, 3) allow users, specifically learners, to influence and intervene with the system and its analyses, and 4) design all algorithms, analyses, and visualizations with the awareness that they may be wrong. In the end, such measures must be communicated to optimize trust and credibility.

[12] A critical introduction in the context of online gaming is given for example in the *iX Developer* journal, volume 1/2015.

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

7 ACKNOWLEDGMENTS

The work presented in this paper is supported by the European Commission (EC) under the Information Society Technology priority of the 7th Framework Programme for R&D under contract no 619762 LEA's BOX. This document does not represent the opinion of the EC and the EC is not responsible for any use that might be made of its content.

8 REFERENCES

Albert D. & Lukas J. (1999). *Knowledge spaces: Theories, empirical research, applications*. Mahwah: Lawrence Erlbaum Associates.

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday, 11.* Retrieved from http://firstmonday.org/article/view/1394/1312

Berg, A. (2013). *Towards a uniform code of ethics and practices for learning analytics* [Web log post, August 21]. Retrieved from https://www.surfspace.nl/artikel/1311-towards-a-uniform-code-of-ethics-and-practices-for-learning-analytics/

Bomas, E. (2014). *How to give students control of their data.* [Web log post, August 29]. Retrieved from http://www.laceproject.eu/blog/give-students-control-data/

Bull, S. (2012). Preferred features of open learner models for university students, In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), Intelligent tutoring systems (pp. 411–421), Berlin/Heidelberg: Springer-Verlag. http://dx.doi.org/10.1007/978-3-642-30950-2_53

Bull, S., & Kay, J. (2010). Open Learner Models. In R. Nkambou, J. Bordeau, & R. Miziguchi (Eds.), Advances in intelligent tutoring systems (pp. 318–338). Berlin: Springer. http://dx.doi.org/10.1007/978-3-642-14363-2_15

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics. *Educause Review, 42*(4), 40–57. Retrieved from https://net.educause.edu/ir/library/pdf/PUB6101.pdf

Cavoukian, A. (2011). *Privacy by design: The 7 foundational principles. Implementation and mapping of fair information practices*. Ontario, Canada: Information and Privacy Commissioner of Ontario. Retrieved from http://www.ipc.on.ca/images/Resources/pbd-implement-7found-principles.pdf

Debatin, B., Lovejoy, J. P., Horn, A.-K., & Hughes, B. N. (2009). Facebook and online privacy: Attitudes, behaviors, and unintended consequences. Journal of Computer-mediated Communications, 15, 83–108. http://dx.doi.org/10.1111/j.1083-6101.2009.01494.x

Dirndorfer Anderson, T., & Gardiner, G. (2014). What price privacy in a data-intensive world? In M. Kindling & E. Greifeneder (Eds.), *Culture, Context, Computing: Proceedings of iConference 2014* (pp. 1227–1230). Illinois: iSchools. Retrieved from https://www.ideals.illinois.edu/handle/2142/47417

Dyckhoff, A. L. (2011). Implications for learning analytics tools: A meta-analysis of applied research questions. *International Journal of Computer Information Systems and Industrial Management Applications, 3*, 594–601.

Ess, C., & AoIR (2002). *Ethical decision-making and Internet research: Recommendation from the AoIR Ethics Working Committee.* (Report from the Association of Internet Researchers).

Council Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L 281/31

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

Federal Trade Commission. (1998). *Privacy Online: A report to Congress*. Retrieved from the Federal Trade Commission website https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4, 304–31. http://dx.doi.org/10.1504/IJTEL.2012.051816

Family Education Rights and Privacy Act (FERPA,) 34 C.F.R.§ 99.34 (2004). Retrieved from https://www.law.cornell.edu/cfr/text/34/part-99

European Union Agency for Fundamental Rights (FRA). (2014). *Handbook on European data protection law.* Luxembourg: Publications Office of the European Union. Retrieved from http://fra.europa.eu/en/publication/2014/handbook-european-data-protection-law

Friedman, B. (1997). *Human values and the design of computer technology*. Cambridge, MA: Cambridge University Press.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society, 15*, 42–57.

Gotterbarn, D. (1999). How the new software engineering code of ethics affects you. *IEEE Software, 16*, 58–64.

Gotterbarn, D., Miller, K., & Rogerson, S. (1997). Software engineering code of ethics. *Communications of the ACM, 40*, 110–118. http://dx.doi.org/10.1145/265684.265699

Heller, J., Ünlü, A., & Albert, D. (2013). Skills, competencies, and knowledge structures. In J.-C. Falmagne, D. Albert, C. Doble, D. Eppstein, & X. Hu (Eds.), *Knowledge spaces: Applications in education* (pp. 229-242). Berlin: Springer.

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2014). *NMC Horizon Report: 2014 Higher Education Edition.* Austin, TX: The New Media Consortium.

Kay, D., Korn, N., & Oppenheim, C. (2012, November). Legal, risk and ethical aspects of analytics in higher education*, JISC CETIS Analytics Series, 1*(6). Retrieved from http://publications.cetis.ac.uk/2012/500

Kobsa, A. (2007). Privacy-enhanced web personalization. In P. Brusilovski, A. Kobsa, & W. Nejdl (Eds.), The adaptive web: Methods and strategies of web personalization (pp. 628–670). Berlin: Springer. http://dx.doi.org/10.1007/978-3-540-72079-9_21

Levin, A., & Nicholson, M. J. (2005). Privacy law in the United States, the EU and Canada: The allure of the middle ground. *University of Ottawa Law & Technology Journal, 2,* 357–395.

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*, 30–40. Retrieved from http://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Markham, A., & Buchanan, E. (2012). *Ethical decision-making and Internet research: Recommendations from the AoIR Ethics Working Committee (Version 2.0).* Available at http://aoir.org/reports/ethics2.pdf

Moore, S. L. (Ed.) (2008). Special issue: Practical approaches to ethics for colleges and universities. *New Directions for Higher Education*, *142*, 1–7.

Movius, L. B., & Krup, N. (2009). U.S. and EU privacy policy: Comparison of regulatory approaches. *International Journal of Communication, 3,* 169–178.

OECD. (2013a). *Exploring data-driven innovation as a new source of growth: Mapping the policy issues raised by "big data."* In Supporting Investment in Knowledge Capital, Growth and Innovation, OECD Publishing. http://dx.doi.org/10.1787/9789264193307-12-en

OECD. (2013b). *The OECD Privacy Framework*. OECD Publishing. Retrieved from http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

Pardo, A. (2014). Designing learning analytics experiences. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 15–38). New York: Springer. http://dx.doi.org/10.1007/978-1-4614-3305-7_2

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*, 438–450. http://dx.doi.org/10.1111/bjet.12152

Personal Information Protection and Electronic Documents Act (2000, c.5). Retrieved from the Department of Justice Canada website: http://laws-lois.justice.gc.ca/eng/acts/P-8.6/

Privacy Act 1988. No. 119 1988. (Cth). Retrieved from the Australian Federal Register of Legislation website: https://www.legislation.gov.au/Details/C2012C00414

Privacy Regulation 2013 (Cth). Retrieved from the website: Office of the Australian Information Commissioner website: https://www.oaic.gov.au/privacy-law/privacy-act/privacy-regulations

Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest Law Review, 49*, 393–432.

Rodotà, S. (2009). Data protection as a fundamental right. In S. Gutwirth, Y. Poullet, P. de Hert, C. de Terwangne, & S. Nouwt (Eds.), *Reinventing data protection*? (pp. 77–82). Dordrecht: Springer. http://dx.doi.org/10.1007/978-1-4020-9498-9_3

Scheffel, M., Drachsler, H., Stoyanov, S., & Specht, M. (2014). Quality indicators for learning analytics. *Educational Technology & Society, 17*, 117–132.

Schwartz, P. M. (2011). Privacy, ethics, and analytics. *IEEE Security and Privacy, 9*, 66–69. http://dx.doi.org/10.1109/MSP.2011.61

Sclater, N. (2014b). *Code of practice for learning analytics: A literature review of the ethical and legal issues*. Jisc. Retrieved from http://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-_Literature_Review.pdf

Sclater, N. (2014c, October 29). Notes from Utrecht workshop on ethics and privacy issues in the application of learning analytics. *Effective Learning Analytics.* Jisc. [Web log post]. Retrieved from http://analytics.jiscinvolve.org/wp/2014/10/29/notes-from-utrecht-workshop-on-ethics-and-privacy-issues-in-the-application-of-learning-analytics/

Sclater, N., & Bailey, P. (2015). *Code of practice for learning analytics*. Jisc. Retrieved from http://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57,* 1509–1528. http://dx.doi.org/10.1177/0002764213479366

Spiekerman, S., & Cranor, L. F. (2009). Engineering privacy. *IEEE Transactions on Software Engineering, 35,* 67–82. http://dx.doi.org/10.1109/TSE.2008.88

Stiles, R. J. (2012). *Understanding and managing the risks of analytics in higher education: A guide*. EDUCAUSE. Retrieved from http://net.educause.edu/ir/library/pdf/EPUB1201.pdf

Stutzman, F., & Kramer-Duffield, J. (2010). Friends only: Examining a privacy-enhancing behavior in Facebook. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10),* 1553–15262. http://dx.doi.org/10.1145/1753326.1753559

Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property, 11*, 239–273.

Open University (2014). *Policy on ethical use of student data for learning analytics*. Retrieved from http://www.open.ac.uk/students/charter/essential-documents/ethical-use-student-data-learning-analytics-policy

The White House. (2012*). Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy*. Washington, DC: The White House. Retrieved from https://www.whitehouse.gov/sites/default/files/privacy-final.pdfvan

(2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, *3*(1), 66–90. http://dx.doi.org/10.18608/jla.2016.31.5

Harmelen, M., & Workman, D. (2012). Analytics for learning and teaching. *Jisc CETIS Analytics Series, 1*(3). Retrieved from http://publications.cetis.ac.uk/2012/516

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57,* 1500–1509. http://dx.doi.org/10.1177/0002764213479363

Willis, J. E. (2014). Learning analytics and ethics: A framework beyond utilitarianism. *EDUCAUSE Review Online.* Retrieved from http://www.educause.edu/ero/article/learning-analytics-and-ethics-framework-beyond-utilitarianism

Willis, J. E., III, & Pistilli, M. D. (2014, April). Ethical discourse: Guiding the future of learning analytics. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/ethical-discourse-guiding-futurelearning-analytics

Willis, J. E., Campbell, J. P., & Pistilli, M. D. (2013). Ethics, big data, and analytics: A model for application. *EDUCAUSE Review Online*. Retrieved from http://er.educause.edu/articles/2013/5/ethics-big-data-and-analytics-a-model-for-application