DELIVERABLE D2.3

PRIVACY AND DATA PROTECTION POLICY

Document Version:	v2.2 – 23 December 2014
Document Status:	Final
Document Type:	Report
Diss. Level:	Public
Lead Partner:	TUGraz
Delivery Date:	M10 (December 2014)

Authors:

Christina Steiner (TUGraz)
Drew Masci (UoB)
Matthew Johnson (UoB)
Ali Türker (Sebit)
Martin Drnek (SCIO)
Michael Kickmeier-Rust (TUGraz)

EXECUTIVE SUMMARY

This report documents the privacy and data protection considerations and policy in the LEA's BOX project. To find a balance between learning analytics research and individual privacy the project needs to appropriately address privacy and data protection principles and comply with relevant legal regulations. As a basis for establishing the requirements and implications for privacy and data protection three main sources of information have been used:

- a review of state of the art literature and best practice guidelines on privacy, data protection and related ethical aspects in learning analytics has been conducted.
- the relevant international data protection and privacy regulations have been studied.
- a consultation with the external ethics advisor of the project.

A range of general guidelines, model codes, and principles for appropriate data and privacy protection exist that may serve the consideration of these topics in a learning analytics context. Approaches of defining frameworks for dealing with ethics and privacy issues specifically in the field of big data, in general, and learning analytics, in particular, adapt fundamental ethical principles of privacy and data protection to these analytics domains and adjusting it to the specific application field. Since ethics and data protection do not only need to be translated into an adequate general privacy policy and information practice, but also needs to be technically reflected by according system functionality, provisions, and data structures, the use of privacy by design principles provide a valuable starting point to ensure that data protection is embedded into the design and architecture of learning analytics.

A general awareness of the importance and significance of data protection is also reflected in national and international laws and directives, where data protection is usually considered as a fundamental right. An overview on national regulations in Czech Republic, Turkey, Austria, and the United Kingdom, where LEA's BOX is carrying out pilot and evaluation studies and collecting and processing personal data. In addition, the European data protection regulations are summarized, which mainly aim in establishing common rules for data protection in member states and keeping a balance between a high level of protection of individual privacy and the movement of personal data within the European Union.

A meeting with the external Ethics advisor has been held to discuss relevant ethical issues of LEA's BOX. The meeting has fostered reflection on the project work and resulted in helpful advice on how to deal with general ethical and privacy issues in the project. Future meetings will elaborate in more detail specific ethical questions.

The consideration of these different types of resources and inputs on privacy and data protection served as a basis for elaborating an approach for dealing with privacy and data protection in LEA's BOX. A set of eight principles have been derived from the integration of privacy and data protection aspects reflected in the different resources used: Data privacy, purpose and data ownership, consent,

transparency and trust, access and control, accountability and assessment, data quality, and data management and security. These principles imply requirements for LEA's BOX on the general organisational/managerial and on the technical level. Key implications for the technical implementation of the LEA's BOX platform and services are postulated to ensure an ethical treatment of personal data in the project. (See also D2.1 for further information.)

TABLE OF CONTENTS

1. INTRODUCTION

The LEA's BOX project focuses on researching and developing novel approaches to competence-centered 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. While WP3 and WP4 are dedicated to the implementation of the analytics and visualization services, WP2 is engaged with the development of the overall LEA's BOX platform integrating and providing access to these services.

This report is the result of the work accomplished in T2.3 privacy and data protection considerations/design. In order to identify and specify privacy and data protection rules for accessing data and to define best practice guidelines for LEA's BOX, desktop research on big data protection and ethical issues and dilemmas in the context of learning analytics, as reported in the literature, has been carried out. The relevant national and EU data protection regulations have been reviewed. In addition, the project's external ethical advisor has been consulted to provide additional input and advice on ethical and legal aspects related to the use, analysis, and exploitation of learning data.

A thorough consideration of all these different sources of information shall assure that the learning analytics toolbox and platform work in accordance with national privacy policies and regulations on data protection and state of the art and best practice approaches to dealing with ethical and legal aspects. Requirements and guidelines for dealing with privacy and data protection in LEA's BOX are derived. The technical implications resulting from that for the treatment of the project's sensitive data relate to data management and security, authentication and authorisation processes etc. and complement the system design documents (cf. D2.1 delivered in M8).

This document is structured as follows: Section 2 elaborates on privacy and ethical issues and dilemmas in learning analytics and proposed approaches or frameworks for dealing with these topics, as discussed in the current literature. In Section 3 an overview of privacy and data protection regulations is given, touching on existing international legislation and summarizing national as well as European regulations in particular relevant for the work in LEA's BOX. Section 4 presents a report of the first meeting with the external ethics advisor, Prof. Heinrich Römer. Section 5 elaborates on requirements with respect to privacy, data protection, and ethics, as they can be derived and integrated from the aspects discussed in the literature, from relevant laws, and based on the reflections resulting from the consultation of the ethics advisor. Finally, in Section 5 the data protection principles defined are translated into implications with respect to the development work of the LEA's BOX architecture, platform and services. We also include a small case study of privacy policies in action with tools based at SEBIT that will be integrated during the lifecycle of the project. (Section 6).

2. An overview of the literature on ethical issues, privacy, and data protection in learning analytics

Learning analytics are key emerging technologies in education (Johnson et al., 2014; see also D3.1) 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 analysed the data of their students to some extent. Learners today have access to a multitude of learning tools, application, and resources, they enhance their learning experience in virtual or simulated environments, they connect to others through social media. All those interactions and resources may be captured and those multi-faceted learning processes can (potentially) be analysed using big-data analytics techniques (Pardo & Siemens, 2014).

With the advent and increasing capacity and adoption of learning analytics also an increasing number of ethical and privacy issues arise. For example, the evolution of sensors and new technologies enables a multi-faceted tracking of learners' activities, location 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-identification of data, transparency, data security, interpretation of data, as well as data classification and management (Slade & Prinsloo, 2013). These issues have been dealt with some tension so far (Pardo, 2014). There is a need to develop a clear and agreed set of ethical guidelines with respect to the ownership of data and analytic models, rights and responsibilities (Ferguson, 2012). At the moment there are no standard methods and procedures for informed consent, opting out etc. In fact, the need for a clearly defined and uniform approach to and code of ethics to appropriately deal with the topics of ethics, privacy and learning analytics is increasingly acknowledged (Berg, 2013).

This section gives an overview on relevant work on ethics and privacy issues in big data analytics, in general, and in learning analytics, in particular, and summarises relevant more generic approaches and guidelines towards an ethics and information practice code.

2.1. Big data and ethics

Privacy and ethics have evolved important and pressing topics not only in learning analytics, though, but in analytics and big data, in general (Schwartz, 2011; PMCA, 2013). "Big data poses big privacy risks," as Tene and Polonetsky (p. 251) put it. Data has become 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 in data subjects and 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 a lot of personal details via these networks. Privacy attitudes and privacy behaviours, thus, often differ (Stutzman & Kramer-Duffield, 2009), which is called the "privacy paradox" (Barnes, 2006) and is evident when comparing users' self-reports about their understanding of caution regarding privacy settings and their actual, unconcerned behaviour of usually just keeping default settings without taking the opportunity of updating them to their needs and preferences (Debatin et al., 2009). So, privacy attitude and privacy behaviour are not necessarily conforming - people may not act according to the privacy preferences they claim. Usually they appear to be 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 minimisation and individual control and context need to be somewhat relaxed in a big data context and considered not only from an individual but also 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 at converting 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 (act 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 of 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

Beside specifying these generic principles, Schwartz in particular argues that at different stages of the analytics process different ethical considerations are relevant. Accordingly, the rules how to tackle these challenges need to be tailored to each analytics stage – always aiming at maximising good results and minimising 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 anonymisation should be done, as appropriate. In decision making it needs to be made sure that the analytics results on which decisions are based are reasonably accurate.

2.2. 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 learning analytics in this direction. Still, although many authors mention ethical issues, there are only few coherent approaches elaborating ethical challenges in more detail and attempting to define an ethical framework to guide institutions, researchers and developers in the application of learning analytics (Slade & Prinsloo, 2013).

Relevant ethical issues and dilemmas in learning analytics can be summarised and grouped into the following overlapping areas (Campbell, DeBlois & Oblinger, 2007; Pardo & Siemans, 2014; Sclater, 2014; Slade & Prinsloo, 2013; Willis, 2014):

- **Privacy**: The possibility that actions and personal data are tracked causes concerns in 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 whether an individual needs give consent to data collection and analysis, the obligation to inform about the data being collected and analysed, 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 and 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, who the owner of the data collected, of the analytics models, and the analytics output is. 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 arise and what the implications of ineffective or misdirected interventions as a result of faulty analytics results are.
- **Role of knowing and obligation to act**: Learning analytics brings new knowledge and insights about learning. The question arises, whether this gained knowledge entails a responsibility to act on this information, and what the ramifications of action or inaction are.

The topics of privacy and ethics are directly related with aspects of trust and accountability (Pardo & Siemens, 2014). A rational and sensible 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 with 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 of dealing with ethics in learning analytics commonly and understandably ground their discussion within and relating to legalities and legal understanding of privacy (Willis, 2014).

One possible approach of elaborating the ethical issues learning analytics is to determine and analyse the risks of implementing a learning analytics project and how to manage them. Stiles (2012) identifies a set of specific areas and associated risks. Data protection is considered as a key risk to be addressed, including the aspects of privacy, security, governance, and compliance. To ensure privacy, security, quality, auditability of data an appropriate level of control needs to be implemented (i.e. data and information governance – for example through policy, 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 organisation.

Greller and Drachsler (2012) have considered ethical and legal aspects in their framework for learning analytics under the dimension of 'external constraints'. Apart from ethical, legal, and social constraints, they also consider organisational, managerial, and process constraints as relevant components on this dimension. These external limitations can be categorised into conventions, like ethics, personal privacy, and other socially motivated constraints, and norms which imply 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 which measures are morally allowable; the law defines what is allowed without legal consequences (Berg, 2013). In many cases ethical issues are reflected in legislation, but ethical considerations go beyond what is set in laws and depends on ideological assumptions and epistemologies (Slade & Prinsloo, 2013). Much of legal regulations are based on ethics, and in particular situations an ethical position needs to be applied for interpreting the law (Sclater, 2014). Kay, Korn, and Oppenheim (2012) highlight that given the mission and responsibilities of education, “broad ethical considerations are crucial regardless of the compulsion in law” (p. 20).

Kay et al. (2012) outline that learning analytics is in the area of conflict between assuring educational benefits, business interests of and competitive pressure on educational institutions, and expectations of the born digital generations of learners. They postulated 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 the area of conflict and a need for balancing between faculty expectations, privacy legislation, and an educational institutions 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 analysing 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 making any judgements.	**Loyalities**: Loyalities 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 analysed.	**Principles**: A set of ethical principles (e.g. Mill’s principle of utility – ‘Seek the greatest happiness for the greatest number’) is identified and considered that are applicable to the situation in question.

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:

- **Learning analytics as a moral practice**: Focus should not only be put on what is effective, but on supporting decisions on what is appropriate and morally necessary. The ultimate goal is understanding, not measuring.
- **Students as agents**: Students should be involved in the learning analytics process as collaborators and co-interpreters. A student-centric approach to learning analytics is recommended.
- **Student identity and performance are temporal dynamic constructs**: The 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) have analysed ethical and privacy issues in learning analytics research in educational institutions and have also taken into account how privacy and ethics are addressed in other contexts. They identify a set of four principles that aggregate numerous issues and are intended to serve as a basis for setting up appropriate mechanisms for meeting ethical and legal requirements when developing and deploying learning analytics. When applying these principles, this needs to be done in due consideration of legal and social requirements. 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 of users to access the data collected about them and, if necessary, to correct it.
- **Accountability and assessment**: The analytics process should be reviewed and for each aspect of the learning analytics scenario the responsible entities should be identified.

2.3. General ethical and privacy guidelines and models

The OECD guidelines have been indicated as 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 Co-Operation and Development) provided the first internationally agreed collection of privacy principles[1], aiming at harmonizing legislation on privacy and facilitating the international flow of data. The set of eight basic guidelines mirrored the principles earlier defined by the European Convention for the Protection of Individuals with Regard to the Automatic Processing of Personal Data and addressed (Levin & Nicholoson, 2005). The basic OECD principles are (OECD, 2013, p. 14-15):

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

[1] http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

The OECD guidelines were not binding for OECD members, but have gained legal significance and served as a basis for privacy legislation in Europe (European Parliament, 1995; Levin & Nicholoson, 2005; Spiekermann & Cranor, 2009). The Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (OECD, 2013), an update of the original version from 1980, which constituted The revision keeps the original "Basic Principles" of the guidelines, while modernising 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, OECD is working on privacy-related issues in the context of large-scale data use and analytics. In a preliminary report (OECD, 2014) on the broader topic of 'data-driven innovation as a new source of growth' different sectors of data use an analytics are elaborated (online advertisement, health care, utilities, logistics and transport, and public administration), however without any specific reference to learning or academic analytics. Privacy protection is indicated as one of several areas that needs public policies and practices to leverage the potential of big data. Privacy protection enabling open, secure, reliable, efficient, and also cross-border, flows of data on the one hand, and reducing privacy risks and enhancing responsible behaviour in the use of personal data is called for.

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

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

be enforcement and redress mechanisms through self-regulatory regimes, legislation creating private remedies for users, or government enforcement.

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 (Slade & Prinsloo, 2013). The Association of Internet Researchers provides a set of ethical guidelines for decision making about internet research (Ess & AoIR, 2002; Markham & Buchanan, 2012). These are aimed at providing researchers a basis for conducting their research in an ethical and professional manner and have also been indicated by learning analytics researchers as a valuable source for dealing with privacy issues in the application of learning analytics.

2.4. Ethics by Design

Since learning analytics involves technology, ethics and privacy concerns may not purely considered from a legal perspective, but need to be addressed from technological point of view (Pardo & Siemens, 2014). One way of ensuring that is to take privacy and ethics, in general, into account already during the design process of learning analytics tools. This approach is called ‘privacy by design’, ‘value-sensitive design’ or ‘ethics by design’ and it has been started to be acknowledged and taken up also in learning analytics research (e.g. Bomas, 2014; Scheffel, Drachsler, Stoyanov, & Specht, 2014).

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 support humans to follow ethical rules (Gotterbarn, Miller, & Rogerson, 1997; Gotterbarn, 1999). Privacy by design, more concretely focuses on privacy engineering and developing guidelines for designing and developing privacy-friendly systems (Cavoukian, 2011). Spiekermann and Cranor (2009) have carried out a privacy requirements analysis that is applicable for a wide variety of systems and identify system activities typically performed by information systems and their impact on user privacy (see Table 2 for an overview). This impact depends on how the system activities are performed, what type of data is used and who uses it, and which privacy spheres are affected. Guidelines are provided on how notice, choice, and access can be implemented as fair information practices and users can be informed about them. Relating to these guidelines, in ethics by design a ‘privacy-by-policy’ approach (focus on implementation of notice and choice principles) and a ‘privacy-by-architecture’ approach (focus on minimizing collection of identifiable personal data and anonymisation) can be distinguished (Spiekermann & Cranor, 2009)

Table 2: Information system activities and their impact on aspects of user privacy.

System Activity	Relevant Aspects for User Privacy
Data transfer	- transparency on data transfers (within organisation, to third parties) - controlled transition of data
Data storage	- protection from unauthorised access - transparency and control over personal data - awareness of data storage activities, persistent and transient storage
Data processing	- awareness of transformation of data - information on secondary use of data - outsourcing of data for processing

2.5. WRAP UP

Overall, there are a range of more or less generic approaches of defining guidelines, model codes, and principles for appropriate data and privacy protection (cf. Section 2.3). Learning analytics and big data, in general, tend to challenge some of the principles defined in such collections of principles and best practice guides, like the concept of data minimization (focused collection) and consent requirements (Tene & Polonetsky, 2013). Approaches of defining frameworks for dealing with ethics and privacy issues specifically for the field of big data (cf. Section 2.1), in general, or learning analytics (cf. Section 2.2), in particular, translate fundamental ethical principles of privacy and data protection to these analytics domains and adjusting it to the specific application field. But even guidelines proposed especially for learning analytics are usually generic and need to be aligned with the very specific context of a concrete learning analytics application in question. Tene and Polonetsky (2013) talk about “levers that must be adjusted to adapt to varying … conditions” (p. 242). Since ethics and data protection do not only need to be translated into an adequate general privacy policy and information practice, but also needs to be technically reflected by according system functionality, provisions, and data structures, the use of privacy by design principles (cf. Section 2.4) provide a valuable starting point to ensure that data protection is embedded into the design and architecture of learning analytics.

In general, organisations researching and providing learning analytics technologies and educational institutions adopting learning analytics need to set up mechanisms and policies to address ethical and privacy issues in learning analytics in a context-dependent and appropriate manner. When specifying a set of policies and principles on the ethical use of educational data for learning analytics, the ethical frameworks proposed in the literature and existing ethical and privacy guidelines as summarised above provide a useful starting point. In addition, guarantees and guidelines need to be provided that are in line with current national and international regulations (see Section 3). Grounding on these different resources, a well-defined ethical code may be established, which leaves little room for ambiguity – and this is what is actually needed in learning analytics (Berg, 2013). An example is demonstrated by The Open University (2014), which has set out a policy on ethical use of student data for learning analytics. This policy specifies eight principles which are implemented and shall provide a university-wide guide with regard to the ethical use of learning analytics for analysing student data and identifying interventions for student support.

3. PRIVACY AND DATA PROTECTION REGULATIONS

Legislation on privacy and data protection is regulated in national and international information privacy and data protection laws, which address the protection prohibiting the disclosure or misuse of information held on private individuals. Regulations started to appear in countries with high spread and use of the Internet (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 (Government of Canada, 2004), the Australian Privacy Act and Regulation (Australian Government, 1988, 2013), or the US Consumer Data Privacy in a Networked World (The White House, 2012). The Family Educational Rights and Privacy Act (US Government, 2004), an US federal law, is a legislation that specifically applies to education, i.e. the protection of the privacy of student education records. This law allows the use of data on a need-to-know basis and provides parents certain rights to access to their children's education records.

In parallel with legislative efforts to data protection, non-profit organisations evolved that aim at defending user 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

[2] http://www.argedaten.at

[3] https://www.eff.org/

[4] https://www.privacyrights.org/

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 deliverable (an overview and comparison between international privacy laws and approaches is given, for example, in Levin and Nicholson (2005) and Movius and Krup (2009)). Instead, in the following subsections an overview of the relevant regulations for LEA's BOX are given. This includes national regulations of the project partners' countries, where evaluation and pilot studies and thus, the collection of data are planned – i.e. Czech Republic, Turkey, Austria, and United Kingdom. In addition, an overview of the European legislation is given, which aims at providing a unified initiative for EU members and also governs the transfer from a Member State to third countries.

In general, in legislation the following distinctions between data types are made (FRA, 2014):

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

3.1. NATIONAL REGULATIONS

The following subsections provide an overview on national regulations with respect to data protection and privacy. Concretely, the national regulations of the Czech Republic, Turkey, Austria, and the United Kingdom, as relevant for the LEA's BOX project, have been analysed and the legislation related to collection of personal data, data security, the use and transmission of data is summarised.

3.1.1 Regulations in Czech Republic

Processing of personal data in the Czech Republic must be in accordance with Act No. 101/2000 Coll. on the Protection of Personal Data. Personal data can be processed only with explicit consent of the data subject. While giving consent the data subject must be informed about the purpose of the data processing and what particular data the consent covers, about which controller the consent is given to and for which period of time.

Personal data can be processed by

> *controller of personal data, that shall mean any entity that determines the purpose and means of personal data processing, carries out such processing and is responsible for such processing. The controller may empower or charge a processor to process personal data, unless a special Act provides otherwise; (§ 4, letter j)*
>
> *processor* of personal data, that shall mean any entity processing personal data on the basis of the aforementioned Act or authorisation by a *controller* (§ 4, letter k) – in that case the controller must conclude an agreement with the processor – in particular, the agreement shall explicitly stipulate the scope, purpose and period of time for which it is concluded and must contain guarantees by the processor related to technical and organisational securing of the protection of personal data. (§ 6) *Processor* must be registered by the Office for Personal Data Protection.

Scio company is registered by the Office for Personal Data Protection, therefore can conduct not only as a "controller", but also as a "processor", i. e. is authorized to process personal data of third parties.

Personal data may be preserved only for a period of time that is necessary for the purpose of their processing. After expiry of this period, personal data may be preserved only for purposes of the state statistical service, and for scientific and archival purposes. When using personal data for these purposes, it is necessary to respect the right to protection of private and personal life of the data subject from unauthorised interference and to make personal data anonymous as soon as possible. (§ 5, letter e)

Czech Personal Data Protection Act is saying that free transfer Personal data cannot be limited, if the data are transferred to the Member State of the European Union. This Personal data transfer can be realized only by subjects with appropriate registration by The Office for personal data protection and all obligations of this Act are fulfilled. Company Scio has a appropriate registration in the sense of this Act (no. of registration: 00022741).

Personal data may be transferred to third countries if the prohibition of restriction of the free movement of personal data is ensuing from an international treaty to the ratification of which the Parliament has given his assent and which is binding the Czech Republic, or if the personal data are transferred on the basis of decision of an institution of the European Union. The Office in the Official Journal publishes information about such decisions.

Where the condition pursuant is not met, the transfer of personal data may be carried out if the controller proves that:

(a) the data transfer is carried out with the consent of, or on the basis of an instruction by the data subject;

(b) in a third country, where personal data are to be processed, has been created sufficient specific guarantees for personal data protection, e.g. by other legal or professional regulations and security measures. Such guarantees may be specified in particular by a contract concluded between the controller and the recipient, if this contract ensures application of these requirements, or if the contract contains contractual clauses for personal data transfer to third countries published in the Official Journal of the Office

By the reception personal data from abroad (EU an nonEU) respect personal data processor during processing personal data Czech law. Meanwhile there must be the contraction between personal data processor and personal data controller where is explicitly stated extend, purpose and time of personal data processing and there must be a warranty from processor about technical and administrative personal data protection.

3.1.2 Regulations in Turkey

There is not yet a specific law in Turkey for privacy of personal data. Data protection in Turkish law is governed by the Constitution and a variety of general and sectorial laws such as the Civil Code, Criminal Code, Labor Law. There is a “Draft Law” on Data Privacy, which was announced on 9/11/2005 by the Ministry of Justice, however it was never enacted. The draft law was prepared as part of the process for becoming a member of the European Union, and it provisioned the establishment of an independent supervisory authority for data protection like in other EU member states. It also allowed the transfer of personal data only to countries with and adequate level of data protection. In November 2014, Turkish prime minister announced that the draft law was to be taken to the evaluation of grand national assembly again soon.

The latest legal development on data privacy was through the referendum of 12/09/2010, where personal data collected online has been taken under the protection of Article 20 of Turkish Constitution, which is entitled as “Privacy and Protection of Private Life.” It declares that, “Everyone has the right to demand respect for his private and family life. Privacy of individual and family cannot be violated.” Such right also includes the right to be informed about their personal data, and to access the data and request correction or destruction of such data. The article states that personal data may only be processed with the explicit consent of the relevant person or as envisaged by law. It is further stipulated that the principles and procedures regarding the use of personal data are to be regulated by legislation to that effect. As mentioned above, such a law has yet to be enacted in Turkey.

In addition to the Turkish Constitution, Articles 23 and 24 of the Turkish Civil Code and Articles 134-140 of the Turkish Criminal Code also relate to data privacy. Turkish Civil Code which provide for the "Protection of Personality" implicitly covers the protection of personal data. Turkish Criminal Code regulate the protection of privacy and make it an offense (punishable by imprisonment) to violate the confidentiality of private life.

On the other hand, the term "personal data" is not defined under Turkish laws. In legal practice, it is based on the precedents of the International Court of Human Rights and is understood to cover all information regarding an identifiable person, such as age, gender, place of birth, medical records, criminal records, educational status and other private information.

Henceforth, anonymised data i.e.data that does no longer contain any identifying elements is beyond the scope of all the available legal background. As with the common practice globally, Web based products in Turkey as well contain privacy statements, where personal identifying information is promised to be held private, yet data produced through interacting with the product is owned by the provider and can be used for the benefit of the user or for marketing purposes. Such activity is regulated by the Information and Communications Technologies Authority of Turkey (ICTA) where all data collecting services need to register.

3.1.3 Regulations in Austria

In Austria the compliance with the Datenschutzgesetz 2000 (DSG 2000)[5] is a requirement when processing personal data. The DSG 2000 applies to data from natural and legal persons and includes in §1 the fundamental right to data protection: "Jedermann hat, insbesondere auch im Hinblick auf die Achtung seines Privat- und Familienlebens, Anspruch auf Geheimhaltung der ihn betreffenden personenbezogenen Daten, soweit ein schutzwürdiges Interesse daran besteht." Everyone is entitled to non-disclosure of his/her personal data, particularly where such data refer to his/her private or family life. This right to data protection applies to personal data, in general, irrespective of the medium (e.g. paper, internet) used and the way of processing (e.g. recording, storing, transmitting).

The fundamental right to data protection also includes

- the right to information – i.e. information about which data is being processed, the origin of the data, and how they are used, in particular also if they are transmitted,
- the right to correction of inaccurate data and cancellation of data that is processed unlawfully, in accordance with the relevant statutory provisions.

The data collected may only be used for specific, clearly stated and lawful purposes (§6).

Data applications need to be notified with the data protection authority for the purpose of registration in the data protection register. Data applications that correspond to a standard application as defined by the Federal Chancellor in an ordinance[6], or that solely contain published data or indirectly personal data are not subject to notification (§17). Indirectly personal data is data relating to the subject in such a manner that the controller, processor or recipient of a transmission cannot determine the identify of the data subject by legal means.

The use of data for scientific and statistical purposes is considered as use for special purposes and separately regulated in the DSG 2000. For scientific purposes and statistical investigations with no

[5] Bundesgesetz über den Schutz personenbezogener Daten (Datenschutzgesetz 2000 – DSG 2000): https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=bundesnormen&Gesetzesnummer=10001597

[6] Verordnung des Bundeskanzlers über Standard- und Musteranwendungen nach dem Datenschutzgesetz 2000 https://www.dsb.gv.at/DocView.axd?CobId=30704

individual-related results all data may be used that is publicly available, that has been collected lawfully for other purposes or in the context of other investigations, and that is only indirectly personal (§46). For other data consent of the persons concerned is needed (or, alternatively, specific legal provisions or approval from the data protection authority). Consent is a voluntary and clear declaration of intent given by the person concerned, that he/she agrees – understanding the circumstances – with the use of his/her data.

Measures to ensure data security need to be taken (§14), this means data needs to be safeguarded against unlawful accessibility, destruction, or loss. Measures taken need to take into consideration the type of data, scope of use, as well as technical possibilities and economic feasibility.

Data may only be transmitted if they originate from a lawful application and the recipient has demonstrated appropriate legal competence or authority and the interests in secrecy of the data subject deserving protection are not infringed by the purpose and content of the transmission. Transmission of data to other EU member states is allowed and not subject to any restrictions. Furthermore, no authorisation for data transmission is required for data exchange with recipients in third countries that demonstrate an adequate level of data protection. These countries are enumerated in an ordinance by the Federal Chancellor[7]. In addition, transborder data exchange does not require authorisation if the data has been published legitimately in Austria, data that is only indirectly personal are transmitted, the data subject has clearly given his consent to the data transmission. Basic precondition for transborder data transmission is the legality of the data application in Austria. (§13). If a case of transborder data exchange is not exempted from authorisation, an application for a permit by the data protection authority has to be done.

3.1.4 Regulations in the UK

According to the Data Protection Act of 1998[8] Part I.1, "Personal Data" is data that refers to a living individual and allows for identification via the data itself or from a combination of the data and other information that is held by or can be easily obtained by the Data Controller. The "Data Controller" is defined as a person, either alone, jointly, or in common with other persons, determines the purposes in which the Personal Data will be processed. In Part I.5, the Act states that the Act only applies if the Data Controller or the Personal Data is being processed by equipment within the United Kingdom.

Referring to Part II.7, any individual is entitled to be informed by the Data Controller if any Personal Data is being processed by or on behalf of the Data Controller. This is to be done in the form of a brief description of the Personal Data is being collected, the reasons to why the Personal Data will be processed and whom the Personal Data will be disclosed to. A Data Controller is exempt from providing the information unless a request is made in writing and a prescribed fee has been paid or if in disclosing the information another individual's Personal Data would also be disclosed.

[7] Verordnung des Bundeskanzlers über den angemessenen Datenschutz in Drittstaaten (Daten-schutz-angemessenheits-Verordnung - DSAV) - https://www.dsb.gv.at/DocView.axd?CobId=30701

[8] http://www.legislation.gov.uk/ukpga/1998/29/contents

Part II.10 states that any individual is entitled, in writing and within a reasonable time frame, to request for their Personal Data to be no longer processed by the Data Controller if the individual feels that the processing of the data would cause unwarranted damage or distress to them or another individual. The Data Controller needs to, in writing, reply within 21 days stating that they have complied or intend to comply with the notice or state that the request to no longer process the data is unjustified and to explain to what extend the Data Controller intends to comply or has complied with the notice.

In addition to the Data Protection Act of 1998, the University of Birmingham has its own Data Protection Policy[9]. This policy was created to allow for easy and concise understanding of security measures to be implemented in accordance with the Data Protection Act. Within the policy additional requirements must be met where personal data is concerned. In summary these additions state that personal data shall:

- Not be kept longer than is necessary for the purposes of holding the data.
- Be kept safe from unauthorised access, accidental loss or destruction.
- Not be transferred to a country outside the European Economic Area, unless that country has equivalent levels of protection for personal data.

The Data Protection Policy also states that all staff working for the University of Birmingham are responsible for ensuring that:

- Any personal data, which they hold, is kept securely, for example:
    - if it is computerised, be password protected; or
    - kept only on disk, which is itself kept securely.
- Personal information is not disclosed either orally or in writing or accidentally or otherwise to any unauthorised third party.

3.2. EU Regulations

The transfer of personal data between countries in the EU is necessary in day-to-day business of companies and public authorities. Since conflicting data protection regulations of different countries might complicate international data exchanges, the EU has established common rules for data protection[10]. The application of this European legislation is monitored by national supervisory authorities.

The European data protection legislation considers the protection of personal data as a fundamental right. Current EU law is the 1995 Data Protection Directive[11], which applies to countries of the

[9] http://www.birmingham.ac.uk/documents/university/legal/data-prot-policy.pdf

[10] http://ec.europa.eu/justice/data-protection/index_en.htm

[11] Directive 95/46/EC of the European Parliament and of the Council: http://eur-lex.europa.eu/legal-content/en/ALL/;ELX_SESSIONID=Rx47J8MdGNsy1mb0TBHhKJ0fYQDJ0zB4yQd49c44nx1j92TlvbBf!-557265322?uri=CELEX:31995L0046

European Economic Area (EEA; i.e. all EU countries plus Iceland, Liechtenstein and Norway). The directive seeks to keep a balance between a high level of protection of individual privacy and the movement of personal data within the European Union. It applies to data that is collected and processed automatically (e.g. computer database) and in non-automated ways (traditional paper files). This directive refers to the national law applicable and indicates that each Member State shall apply the national provisions it adopts pursuant to this Directive to the processing of personal data.

According to the EU directive, personal data may only be collected legally under strict conditions and for a explicit and legitimate purposes. Personal data must be adequate and kept up to date. Identification of the data subjects shall be possible no longer than necessary for the purposes for which the data has been collected or further processed (Article 6). The use of data for scientific research is explicitly acknowledged as a meaningful reason for storing data longer than they are needed. In addition, data collected legitimately for any purpose may be further used for statistical research purposes. In these cases usually anonymisation will be required (FRA, 2014). Data must be appropriately protected from misuse, accidental or unlawful destruction, loss, alteration, and disclosure (Article 17).

Data controllers are the persons or entities collecting, managing, and processing personal data. Data collection and processing is legitimate (Section II, Article 7) when the individual concerned (data subject) has unambiguously given consent after being adequately informed about the circumstances in which the data are collected. At least the following information needs to be given (Section IV):

- identity of the data controller or his representative
- purpose of intended data processing
- further information, like whether replies to questions are obligatory or voluntary, possible consequences of failure to reply, the right of access to and to rectify data

Furthermore, data collection is legitimate (Article 7) if the data controller or a third party has a legitimate interest, as long as this interest does not affect the interests of the data subject, or infringe on his or her fundamental rights, in particular the right to privacy. This provision establishes the need to strike a reasonable balance between the data controllers' business interests and the privacy of data subjects.

Data controllers must respect the privacy and data protection rights of the data owners, i.e. those whose personal data is entrusted to them. Data subjects have the right of being informed when personal data is collected about them (right to obtain information), the right to get a copy of this data (right of access) and to ask for deletion, blocking, or erasing of the data (right to object) (Article 12).

Personal data that is particularly sensitive in relation to fundamental rights or privacy (racial or ethnic data, political opinions, religious or philosophical beliefs, and health-related data (European Commission, 2011), including genetic data) deserve specific protection and should not be processed (unless specifically authorized by a law, necessary to protect the vital interest of the person concerned or another person, or data is processed that has been manifestly made public by the person concerned).

The EU Data Protection Directive also defines specific rules for international transfers of personal data to countries outside the EU/EEA. Data transfer outside the EU/EEA may only be done under the precondition that an adequate level of protection is guaranteed (Article 26 (2)). Standard contractual clauses have been defined for transfers to data controllers and processors outside the EU/EEA.

The European Commission is currently in process of establishing a reform of the data protection legislation, to enforce protection of personal data, in particular by updating and modernising data protection rules and principles and bringing them to the digital age.

The European Data Protection Directive has been extended by a specific directive for data communication in the electronic communication sector[12] (ePrivacy directive) to address the specific requirements with respect to privacy and data protection in the context of information and communication technologies, especially the Internet and electronic messaging services. This directive shall help to 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.

Electronic communications service providers must protect the security of their services. They need to ensure that personal data is accessed by authorised persons only, protect personal data from being destroyed, lost or accidentally altered, and ensure the implementation of a security policy on the processing of personal data.

The Directive echoes national legislations of the Member States, to ensure the confidentiality of communications made via public electronic communications. The user must be informed about the data processing and its purpose and must be given the option to withdraw his/her consent. Thereby, an opt-in’ approach is taken, i.e. users must have given their prior consent before such communications are addressed to them.

The required informed consent for storage of or access to information stored on a users’ device also applies to cookies[13], i.e. users need to be asked if they agree to cookies (or similar technologies), before a site starts to use them. For consent to be valid, it must be informed (clear and comprehensive information about the purpose of the storage or access), specific, freely given and must constitute a real indication of the individual's wishes.

[12] Directive 2002/58/EC of the European Parliament and of the Council: http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32002L0058

[13] http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm

4. Report from the meeting with the Ethics Advisor

An external Ethics Advisor has been appointed in LEA's BOX to consult for and oversee ethical aspects and questions involved in the project's research and development. Prof. Dr. Heinrich Römer has been invited and agreed to take over this advisor role. Prof. Römer has extensive expertise in ethics and is the second chairman of the Ethics Commission of the University of Graz, Austria, and representative for natural sciences in this commission.

A first meeting with the external Ethics advisor has taken place on 27 November 2014 in Graz. In this meeting first an introduction and overview of the LEA's BOX project, the project objectives and research has been given. This was followed by a fruitful discussion about relevant ethical issues; a summary of relevant discussion points is given below.

Ethical Use of Data

Regarding the use of data, the ethics advisor underlined that there are no ethical concerns as long as the data is not directly identifiable; i.e. as long as the data does not contain information that may serve to re-identify the persons concerned. Using learning data collected by project partners in the past in a different context for the purpose of LEA's BOX may be considered as a kind of a scientific meta-analysis; the mental or physical integrity of the data subjects as individuals or as a group is not affected.

Consent

The ethics advisor highlighted the importance of gathering data subjects' consent, that they agree to the use of data relating to them. This declaration of intention of the data subjects need to be given after having been informed appropriately about the circumstances of data collection and use. Consent is crucial especially when dealing with data of vulnerable groups. In case of children as data subjects the parental consent needs to be gathered.

While in the literature approaches of passive consent (i.e. sending information letters to parents and asking for a response only if they do not wish their child to participate in the research – e.g. Fisher, 2013) have been discussed or used as viable or possible alternative to active consent methods (e.g. Ellickson & Hawes, 1989; Range, Embry, MacLeod, 2001), Prof. Römer strictly advised against a passive consent process for parental permission. Despite the obvious advantages of passive consent (e.g. higher participation/response rates, ease of administration), he referred to the ethical dilemma inherent in passive consent procedures and underlined that institutional review boards or ethics committees will likely not accept this kind of approach.

In the context of parental permission the issue of using tools developed and provided by the LEA's

BOX project as part of educational practice. If a teacher decides to use the tool in his/her lessons no consent from students or parents is needed, as long as no data collected/recorded with the tools is used in the project. If data collected shall be processed and analysed in the project, consent is necessary (also in case of having teachers collecting/inserting data about their students, as it is the case in the MyClass tool). In order not to withhold the educational tools from students, or their participation in instruction, the ethics advisor suggests to use a consent method that differentiates between the agreement that the child may participate in a certain instructional set-up (use of the tool) and the permission to the use of the data collected.

Transparency

Persons concerned should always be informed appropriately about the data collection and application. In this way, the signature on the consent form shall clearly indicate that the data subject has understood the prevalent circumstances and agrees with it. In case of learning analytics methods applied that provide a measure of learning performance or achievement, the information for data subjects should describe the type of analysis in an appropriate manner (without the need of going into detail on the complex algorithms used). Such an explanation may state, for example, that the analysis procedures used try to mimic the rationale of a teacher in performance evaluation. Transparency in terms of an appropriately detailed description of the analytic procedures and instruments (without the need of going into full detail on the theoretical background and algorithms) also needs to be provided when gathering ethical clearance for an evaluation/pilot study.

Learning Analytics as Moral Practice

The ethics advisor explained that from his perspective the use of collected learning data for learning analytics research is ethical as long as the result of the analysis does not have any direct impact on the students concerned. When researching new learning analytics approaches, in a first step the new methods and algorithms need to be tested and evaluated and should not directly affect data subjects. Only in a second step, after the methods could be validated, the implementation of consequences or interventions on the basis of the analytics results should be approached. A possible approach to validation would be the comparison of learning analytics results with teachers' grading.

This perspective of validating learning analytics is highly interesting and reasonable. It stands somewhat in contrast to discussions on the ethical value of an 'obligation to act', as presented and discussed in the literature (e.g. Campbell et al., 2007; Kay et al., 2012; Willis et al., 2013) – i.e. the idea of an ethical duty to act on the information gained from learning analytics, like information about students at risk of dropping out. Prof. Römer emphasized that at the stage of researching and testing new analytics and assessment procedures, learning analytics researchers should not have the moral claim of a responsibility to use the information or insight gained to act on behalf of the student. Rather, immediate consequences should not be implemented at this stage, in order to avoid any potential negative impact, e.g. grounding an intervention on an erroneous result. Thus, taking this position of validating a new learning analytics approach is directly related with 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 may be due to flaws in analytics algorithms or due to the fact that the data does not sufficiently reflect or cover the learning process.

In this sense, the investigation of a new learning analytics method may be divided into two stages: the validation of the data collection and analysis, and the implementation and evaluation of decision making and action based on the results from collection and analysis. Schwartz (2011) claims in his elaboration on ethical issues and analytics that decision making in the analytic process needs to be grounded on reasonably accurate analytic output. Considering the suggested two-step approach, this may be paralleled with a layered evaluation approach in adaptive system (e.g. Brusilovsky, Karagiannidis, & Sampson, 2004), i.e. evaluating the assessment phase and the adaptation decision making separately.

When gathering students'/parents' consent, it may be explicitly stated that a negative analytics result or conclusions will not be communicated (unless the data subject explicitly wishes to be informed) – similar to methods for consent and communication of research results usually applied in medical studies (e.g. Shalowitz & Miller, 2008).

The first meeting with the external ethics advisor has been an interesting discussion and reflection on the work in LEA's BOX and resulted in helpful suggestions on how to deal with ethical and privacy issues in the project. Further meetings to elaborate more and discuss specific ethical questions will be done in due course throughout the project lifetime.

5. REQUIREMENTS FOR LEA'S BOX

To find a balance between learning analytics research and beneficial uses of data, on the one hand, and individual privacy, on the other hand, LEA's BOX needs to appropriately address privacy and data protection principles and comply with relevant legal regulations.

The aim of this section is, as illustrated in Figure 1, to translate the frameworks and guidelines proposed in the literature to deal with ethical and privacy issues (Section 2), the different relevant jurisdictions (privacy and data protection laws; Section 3), and the advice and suggestions from the external ethics advisor (Section 4) into a coherent set of requirements for the LEA's BOX project. These requirements should go beyond outlining philosophical ideals, but should actually be applied as ethical principles and to fed into the design and development of the project's technologies project (see Figure 1). In line with Schwartz (2011), the requirements shall represent an accountable approach reflecting the specific ethical and data protection issues relevant for the project. They shall provide an appropriate frame for researching and exploring the educational possibilities to benefit from learning analytics without sacrificing privacy (Bomas, 2014).

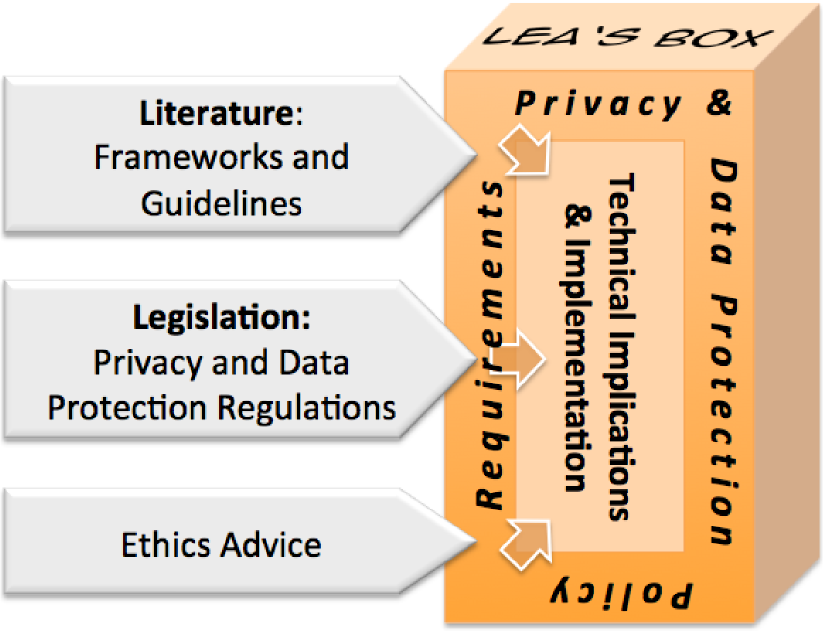
LEA 'S BOX
Privacy & Data Protection Policy
Literature: Frameworks and Guidelines
Legislation: Privacy and Data Protection Regulations
Ethics Advice
Requirements
Technical Implications & Implementation

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

Concretely, a set of principles relating to privacy, data protection, and ethics has been identified, which form the requirements for LEA's BOX. These principles have been derived from an integration and harmonization of general guidelines for fair information practice relating to personal data, ethical frameworks proposed for big data and learning analytics, complemented by the discussion points of the ethics advice, and in alignment with the aspects of data protection and privacy covered by national and European regulations. Table 3 presents an overview and mapping between the ethical and privacy principles from these different resources. The mapping has been done based on the consideration of the individual principles; principles have been mapped to a common topic if a reasonable overlap and pragmatic matching in the idea behind could be identified. It can be seen that the three different types of sources nicely overlap and cover very similar aspects. The principles and requirements derived for LEA's BOX and indicated in Table 3 were formulated based on this integration of privacy and data protection resources and the identification of the relevant aspect covered under each topic. While the way, how these principles are actually applied and implemented may take different forms and may change during project lifetime, compliance with the current laws and regulations shall be ensured at any stage of the project as a main requirement of privacy and data protection.

Table 3: Overview and integration of privacy and data protection principles.

OECD basic privacy principles (2013)	**Federal Trade Commission on Fair Information Practice Principles**	**Schwartz' ethical principles for analytics (2011)**	**Kay et al.'s (2012) guiding principles for analytics in higher education**	**Slade and Prinsloo's (2013) principles for an ethical framework for learning analytics**	**Pardo and Siemens' (2014) ethical and privacy principles for learning analytics**	**Privacy and data protection legislation[1]**	***Ethics advice topics[2]***	**LEA's BOX Principles**
		Compliance with legal requirements				Fundamental right to privacy		Data privacy
		Compliance with cultural and social norms	Comfort and care	Learning analytics as moral practice;			*Ethical use of data*	
Purpose specification		Responsible limits on analytics in sensitive areas or with vulnerable groups		Education cannot afford not to use data		Data collection for specified, explicit and legitimate purpose	*Learning analytics as moral practice*	Purpose and data ownership
Use limitation	Choice / consent		Choice and consent			Right to object	*Consent*	Consent
Collection limitation								
Openness	Notice / awareness		Clarity	Transparency	Transparency	Information to be given to the data subject	*Transparency*	Transparency and trust
Individual participation	Access / participation			Students as agents	Student control over data	Right of access to data		Access and control

Accountability	Enforcement / redress	Accountable measures tailored to identified risks	Consequence and complaint		Accountability and assessment	Judicial remedy for breach of rights		Accountability and assessment
Data quality	Integrity / security	Quality of data collection, data, and analytics results		Student identity and performance are temporal dynamic constructs		Data quality; Transfer of personal data		Data quality
				Student success is a complex and multidimensional phenomenon				
Security safeguards		Appropriate safeguards to protect the security of data			Right to access	Confidentiality and security of processing		Data management and security

Notes: 1. Key areas as represented in privacy and data protection legislation are indicated.
2. The principles listed are not meant to represent a self-contained privacy and data protection framework, but solely indicate the topics discussed with the ethics advisor and summarised in Section 4.

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 need to be fair and provide appropriate protection of privacy. Information on privacy and data protection practices should be available and easily understandable.

Users having the feeling their privacy is endangered may show resistance (Greller & Drachsler, 2012). To give them the feeling 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 by legislation and by additional institutional privacy regulations (Campbell et al., 2007), as represented by the privacy principles at hand.

Users' desire for privacy stands somehow in contrast with initiatives in learning analytics research towards greater openness of educational datasets. Both perspectives are comprehensible – private users may not want to disclose their personal data, at least, wish to ensure controlled disclosure of their data, while learning analytics researchers are aiming at getting access to appropriately large data sets to test and refine their methods. The anonymisation and processing of data according to legal requirements is a key factor mediating and harmonising between these positions. In general, there is strong legal protection of personal data (see also Section 3); sometimes even competing with other legal frameworks, like the Freedom of Information Act in United Kingdom (Greller & Drachsler, 2012).

Key implications for the realisation of LEA's BOX include:

- Privacy and data protection policy should be available for inspection.
- Reasonable measures should be taken to ensure data privacy, including the use of encryption (where appropriate), password protection, minimal and anonymised data exchange through APIs. Insofar as possible the user should not be able to be identified from data that is stored about them, although information will need to persist in the database about the user's credentials in order for the system to render information about learners (e.g. in a non-anonymous form, to teachers, where appropriate).
- Data for analysis should be stored anonymously, associated only with a 'key' (where appropriate). Data should be stored (electronically or otherwise) anonymously and anonymised at the earliest opportunity, where this is not automated.
- Data use must be in alignment with each of the above policies where data is maintained or passes through the country to which the policy pertains.

PURPOSE AND DATA OWNERSHIP

The purpose and boundaries of a learning analytics application should be clearly defined and available before processing is started. "Processing personal data for undefined and/or unlimited purposes is unlawful" (FRA, 2014, p. 68). In essence, considering learning analytics as a moral practice, learning analytics should aim at supporting learners (e.g. Slade & Prinsloo, 2013; The Open University, 2014). When researching new learning analytics methods, though, focus may be on studying new methods of assessment and on establishing a better understanding of learning processes, in a first instance, without implementing any direct consequences or interventions based on analytics outputs. In this case, establishing and ensuring reasonable accuracy of analytics results (i.e. creating truly actionable knowledge) represents the ethical standard to be addressed first (H. Römer, personal communication, 27 November 2014; Schwartz, 2011), before dealing with ethical questions on 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 in this regard there is a lack of legal clarity, when considering learning analytics applications (Greller & Drachlser, 2012).

Traditionally, the data collected about a person, i.e. before anonymisation, belongs to the owner of the data collection tool (data client). Meanwhile, there is a trend of considering users as the owners of the data collected about them and institutions are borrowing them 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 – and the question arises, who the owner of such kind of model is (Pardo, 2014). So, even if the raw personal data is owned by the user, what about the information derived from it? While for raw learning data there is no issue of copyright, copyright and 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 thinking of the integration of learning data from different sources, which may potentially mean different organisations/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.

It has been argued that, in fact, the concept and consideration of data ownership may not be most appropriate and helpful, but more relevant are the notions of data controller and data processor as used in data protection regulations (Sclater, 2014). 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).

An adequate specification and documentation of the purpose of data processing needs to be ensured in LEA's BOX at any stage, and must be made available – together with information about the data controller – for access by the data subject or supervisory authorities.

Some of the key implications for realisation of LEA's BOX include:

- Data collected, processed and used is for the purpose of the analytic algorithms and visual methods of LEA's BOX only.
- The accuracy and maturity of the analytics should be disclosed to those who use them.
- Where integration between multiple entities exists, the user should be informed about where the data has been in order for the analytic to be produced.

CONSENT

Informing users about the collection of their data and gathering their consent need to be realised as a basic ethical principle and procedure (Greller & Drachsler, 2012). It has been argued that in learning analytics there should be virtually no reasons to waive informing users about the use of their data and to set up a clear policy of informed consent (Slade & Prinsloo, 2014).

The principle of consent refers to giving data subjects the possibility to agree/disagree to a 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 and thus, may refer to "a broad definition of the range of potential uses to which a student's data may be put" (Slade & Prinsloo, 2014).

LEA's BOX has to apply appropriate techniques for gathering consent from students and parents, as a legal basis for processing personal data. Thereby, consent shall be collected before any information is collected. Consent need to be free, informed, specific and given unambiguously. Sufficient information needs to be provided to the data subject, to assure he/she is clearly informed about the object and consequences of consenting before taking the decision. Information needs to be precise and easy to understand. Consent given non-explicitly on the basis of inactivity (passive consent from parents) is usually not considered as unambiguous and should be avoided (FRA, 2014; H. Römer, personal communication, 27 November 2014). Although the European regulations do not explicitly mention a general right to withdraw consent at any time, it is widely presumed and accepted that such right exists (FRA, 2014).

In the online environment of LEA's BOX consent may be collected by clicking a box on the screen, providing the user the choice to agree/disagree with the collection and use of the data being collected from them. This may have to be complemented by paper-based consent gathering from parents or legal guardians. According to current privacy legislation the collection of consent also needs to be implemented for the use of Cookies. In case of gathering consent online, layered information notices have been suggested as a good solution to provide access to adequate information in concise and more extensive versions (FRA, 2014). The language used for information need to be understandable for the concerning group of individuals and consent needs to be collected in an explicit manner, providing the option of later withdrawal of consent.

For the LEA's BOX system, the following are important:

- Informed consent needs to be given using an opt-in policy before data collection takes place (whether this is electronic or otherwise).
- Users have the right to withdraw their permission for the data to be used at any time, therefore the mechanisms handling data collected electronically about students' and teachers' use of LEA's BOX must have facilities/processes to exclude specific users' data within analysis.
- Even if consent is not given for data to be collected, users should still have access to the analytic facilities of LEA's BOX.
- For parts of the system where Cookies are used, an explicit agreement from the user is required for proceeding (e.g. using a modal dialogue), if the Cookie will persist after the end of a session of use.

- All consent requests must be in the local language and contain plain wording.

TRANSPARENCY AND TRUST

Transparency is probably the issue that relates to most concerns in ethical considerations on 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 to provide 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 and 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 measures of data protection taken (Willis & Pistilli, 2014).

In a web-based environment, like LEA's BOX, notice and transparency may be created by posting the respective information unavoidable, understandable, and readily accessible at a prominent location on the website.

According to the Fair Information Practice Principles (Federal Trade Commission, 1998), notice of the following information is considered essential to consider data subjects as properly informed: the entity collecting the data, the uses to which the data will be put, potential recipients of data, the type of data collected and data collection method, consequences of refusal, and measures taken to ensure data quality and security. Frequently also information on consumer rights is included. In case of learning analytics, an appropriate and understandable description of the analytic models/procedures should be provided (H. Römer, personal communication, 27 November 2014). Data subjects should be enabled 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, but it is also key to foster trust in data subjects – 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, to raise users' awareness and understanding of the learning analytics approach and, potentially, involve them as active agents in the implementation of learning analytics.

For the LEA's BOX system this means:

- Information regarding what data is used and how it is gathered, recorded, processed etc.

should be easily accessible within the system, and should be easily understandable.

ACCESS AND CONTROL

In addition to gathering users' consent and providing transparency of when and how data is collected and analysed, data subjects should be given control of their own data. This means, users should be given access to the data collected from them, and the opportunity to correct them, 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 outdated, need to be provided to users.

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

Dashboards and open learner models are approaches of visualising 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 access to the data whenever and for how long they want and thus, offer transparency to data subjects on the data collected about the learning process (Pardo & Siemens, 2014). More recent approaches of negotiated user models reflects the idea of student control, since the open learner model is used to interactively negotiate and potentially update the content of the learner model. The work on open learner model communication and negotiation, as carried out in WP4 of LEA's BOX can therefore be considered a realisation and application of the ethical principle of access and participation.

Access and control over data need to be governed by technically implementing appropriate authentication mechanisms and the establishment of an access right structure. Simple and understandable procedures for indicating inaccurate data, for updates or corrections, and for verifying information need to be established and implemented in the management and maintenance of data files.

In the context of the realisation of LEA's BOX:

- All data held about users should be available to inspect.
- Facilities to manage underlying data (create, read, update, delete) need to be provided to

users, in alignment with the purposes of the tools. In the case of the open learner model, this is also partially addressed by a negotiation component.

- Users should be authenticated (i.e. their identity ascertained) before access to the outcomes of analytic processes may be observed.
- Different users may have different authorisation privileges (e.g. teacher may see their students data, students may wish to grant permission for their peers to see their data). Users will always be able to inspect information which is about them.

ACCOUNTABILITY AND ASSESSMENT

Principles of data protection can only work with 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 need to be identified. In the LEA's BOX a clear structure of 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 optimise and support learning and teaching. This can only be achieved when ensuring correctness of the data and analytics algorithms. In fact, when using learning analytics outputs as a basis for taking decisions, for educational interventions etc., the possibility of error should be taken into account and it should be ensured that these outputs are reasonably accurate (cf. Schwartz, 2011). This accuracy cannot be assumed from the outset of introducing and testing a learning analytics approach. As a result, assessment of a new learning analytics is needed; in the beginning ideally through an approach of pure validation and without any direct consequences for data subjects. This is to avoid any harm to users and softens the ethical claim of an obligation to act on the basis of the newly gained knowledge and the question for defining responsibilities for taking action, as frequently discussed in the literature (e.g. Campbell & Oblinger, 2007).

A constant reviewing and adjusting of analytics methods will increase the accuracy of results and suitability of the learning analytics process and maximise impact (Pardo, 2014; Van Harmelen & Workman, 2012). The importance of the review and revision stage in analytics is also highlighted by Schwartz (2011). Beside that, he also refers to the assessment of the impact of using analytics on the basis of stakeholders trust.

In LEA's BOX a continuous assessment, refinement and enrichment of learning analytics methods and tools is targeted as a basis for on-going 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 the pilot and evaluation studies.

Implications for the LEA's BOX system include:

- The limitations and maturity of the visual analytics and the underlying processes should be made clear in the system, or minimally to the end users.
- Regular reviews of analytic processes should take place. If changes are made, minimally end users need to be aware of this, and messages within the software should be considered. User expectations must be managed.

DATA QUALITY

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

Especially when gathering and combining data from multiple sources care needs to be taken to use reliable sources. It needs to be acknowledged that the data collected may provide an incomplete picture of the learning process and only represents a snapshot in time and context. Bias and stereotyping need to be prevented by constantly taking into account the incomplete and dynamic nature of individual learning and experience (Slade & Prinsloo, 2014).

Beside an adequate quality of learning raw data, in LEA's BOX it needs to be ensured that data is used wisely for carrying out integration and analysis. Any interpretation, enhancement, or manipulation of data with the aim of extracting meaning should be grounded on a sound technique; the analytics models should be transparent and available for review and testing.

For the LEA's BOX system:

- New information and inferences added should have almost immediate effect in updating the analytics with relevant, accurate, up-to date information. This may mean that analytics require a clear datestamp stating time/date they were last amended.

DATA MANAGEMENT AND SECURITY

In general, personal data needs to be treated and managed in a sensitive and ethical way. 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 unauthorised access, loss, destruction, or misuse. This includes a clearly defined policy of who is authorised to access the data, to which parts of the data and the application, and which kind of data operations are allowed (Pardo & Siemens, 2014). Processes for redress need to be provided to users in case of any unauthorised access or use of personal data. Preservation and storage of data needs to be aligned with national and EU regulations.

Special attention needs to be paid to this principle, in particular when personally identifiable or even sensitive data is managed. Anonymisation is often used as a strategy to foster willingness to disclose data. Beside lower reluctance of user to share their data, data protection regulations are eased with this kind of data, which provides greater flexibility and possibilities in the data application.

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 a managerial and on a technical level (Federal Trade Commission, 1998; FRA, 2014). On the managerial level, internal organisational rules should be established that cover, for example regular information of 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 etc. Technical measures for data security relate to having the right equipment (hardware and software) in place, encryption in data transmission and storage, the use passwords to limit access, data storage on secure servers etc.

For the technical realisations of LEA's BOX:

- Data should be protected (e.g. through the use of regularly changed passwords to access databases and APIs, and encryption of sensitive information).
- Data should be regularly backed up.
- Account management facilities are needed (e.g. to change passwords)
- Data should be anonymised at the earliest opportunity.
- Appropriate security measures should be implemented on all information transfers between components.

6. Example: Privacy and data protection @ SEBIT

As an example of a data protection policy in action, we provide a commentary with regard to existing tools that will have a level of integration with LEA's BOX, based at SEBIT.

SEBIT as with its products such as Vitamin is registered to ICTA in Turkey, and with its products such as Adaptive Curriculum and Uzzingo is registered to the US federal authority.

It is imperative to discriminate between personally identifying (demographic) information and the data created by the user while interacting with the system (footprint). The footprint data is owned by SEBIT while demographic information is owned by the user. SEBIT is free to share the footprint data. At the “privacy statement” of Uzzingo it is denoted as “Information collected through Uzzingo may be supplied to affiliates of SEBIT, who must keep such information confidential.” At the “privacy statement” of Adaptive Curriculum it is denoted as “Information collected through this Site may be supplied to affiliates of SEBIT, LLC, and other companies and organizations who perform work for us under contract or sell products or services that complement our products and services.”

An example case is the collaboration between SEBIT LLC and Knewton Inc. in providing Knewton powered adaptive learning service to SEBIT customers in USA. For such service SEBIT is sharing the footprint information with Knewton under a joint project contract. Knewton’s business model relies on this model and all major publishers in the World cooperate with Knewton by sharing data to provide their customers an adaptive learning service.

The Knewton-SEBIT joint service case is also an example about the sharing of demographic data in an anonymized, personally “unidentifiable” way. Linked to the footprint data, certain demographic information such as grade or school type can be shared with affiliates as long as the information cannot be used to personally identify who the user is.

To that end the Adaptive Curriculum privacy statement denotes as “The information we learn from users helps us personalize and continually improve your experience at http://www.adaptivecurriculum.com/ in furtherance of serving the user's needs and our legitimate business purposes.”

The personally identifiable part of the demographic data is never to be abused or given to third parties unless (a) in response to a subpoena, court order or legal process, to the extent permitted and required by law; (b) to protect user security or the security of other persons, consistent with applicable laws; (c) in connection with a sale, joint venture or other transfer to some or all of the assets of SEBIT; or (d) in order to enforce the site's terms of use.

SEBIT exercises commercially reasonable care to not otherwise share or disclose the names of users or any personally identifiable information with third parties, except with the prior approval of the user. A user's personally-identifying information may be used to improve the website, and also for the provider’s own marketing and promotional purposes. Even so, the user must always have a chance to opt-out. Adaptive Curriculum privacy statement denotes that “Except as specifically stated herein, our policy is not to share this information with third parties, but we may share this information with other companies within SEBIT network. We will never sell this information to third-party marketing companies. You can choose not to provide certain information, but then you might not be able to take advantage of many of our features. We use the information that you provide for such purposes as responding to your requests, customizing features for you, and communicating with you.”

7. CONCLUDING REMARKS

To conclude this deliverable, we want to reference Stephanie Moore (2008) who 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 education software and in the context of making school studies, individual beliefs, values and preferences influence the scientific work. The reason of this deliverable is to avoid a situation where we know we should address the topic, however, push it aside because of not knowing how.

This deliverable provides the project’s foundations for a proper code of conduct; in particular we need to assure that technology and tools we are about to develop in the project and also those 3rd party technologies we may propagate through the project are in line with these foundations. Thus, we transfer the respective regulations into an approach of “ethics by design”.

Despite the ethical challenges of Learning Analytics in general, and in the context of a research project that is developing novel tools and algorithms, in particular – *education cannot afford not to use (big) data*, to say it in the words of Sharon Slade and Paul Prinsloo (2013).

In the context of this complex and sensitive field, this deliverable cannot claim to be complete; for example, critical further aspects concern tracking of IP addresses, the accessing of individual data such as done by many Smartphone apps (e.g., GPS location), the identifiability of users among each other, or the access to webcams or chat functions (a critical introduction in the context of online gaming is given for example in the *iX Developer* journal, volume 1/2015). Still, is provides our ‘personal’ code of conduct, strengthens our ‘personal’ awareness, and derives a number of concrete technically requirements.

REFERENCES

Australian Government (1988). *Privacy Act 1988. No. 119, 1988 as amended*. Canberra: Office of Parliamentary Counsel

Australian Government (2013). *Privacy Regulation 2013*.

Barnes, S.B. (2010). A privacy paradox: Social networking in the United States. *First Monday, 11.* Retrieved December 10, 2014 from http://firstmonday.org/article/view/1394/1312

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

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

Brusilovsky P., Karagiannidis C. & Sampson D. (2004). Layered evaluation of adaptive learning systems. *International Journal of Continuing Engineering Education and Life-Long Learning, 14*, 402-421.

Bull, S. & Kay, J. (2010). Open Learner Models. In R. Nkambou, J. Bordeau and R. Miziguchi (Eds.), *Advances in Intelligent Tutoring Systems* (pp. 318-338). Berlin: Springer.

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

Cavoukian, A. (2011). *Privacy by design. The 7 foundational principles. Implementation and mapping of fair information practices*. Information and Privacy Comminssioner of Ontario. Retrieved December 12, 2014 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.

Dirndorfer Anderson, T. & Gardiner, G. (2014). What price privacy in a data-intensive world? In: *iConference 2014 Proceedings* (pp. 1227-1230).

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.

Ellicksoon, P.L. & Hawes, J.A. (1989). An assessment of active versus passive methods for obtaining parental consent. *Evaluation Review, 13,* 45-55.

Ess, C. & AoIR (2002). *Ehtical decision-making and Internet research: Recommendation from the AoIR Ethics Working Committee*. AoIR.

European Parliament (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.

European Commission (2011). *Advice paper on special categories of data („sensitive data“). Article 29 Data Protection Working Party*. Brussels: European Commission, Directorate General Justice,

Federal Trade Commission (1998). *Privacy Online: A report to Congress*. Federal Trade Commission. United States of America.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*, 304-31.

Fisher, C.B. (2013). *Decoding the ethics code. A practical guide for psychologists.* Thousand Oaks: SAGE Publications.

FRA (2014). *Handbook on European data protection law. European Union Agency for Fundamental Rights.* Council of Europe. Retrieved December 10, 2014 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.

Government of Canada (2004). *Personal Information Protection and Electronic Documents Act*. Canada: Minister of Justice.

Markham, A. & Buchanan, E. (2012). *Ethical decision-making and Internet research: Recommendations from the AoIR Ethics Working Committee (Version 2.0).* AoIR.

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

Kay, D., Korn, N., & Oppenheim, C. (2012). Legal, risk and ethical aspects of analytics in higher education. *JISC CETIS Analytics Series: Vol. 1 No. 6*. Retrieved October 22, 2014 from http://publications.cetis.ac.uk/c/analytics

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: Srpnger.

Levin, A. & Nicholson, M.J. (2005). Privacy law in the United States, the EU and Canada: The allure oft he 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.

Moore, S. L. (Ed.) (2008). Special Issue: Practical Approaches to Ethics for Colleges and Universities. *New Directions for Higher Education*, 2008(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 (2013). *The OECD Privacy Framework*. OECD Publishing.

OECD (2013). Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data". *OECD Digital Economy Papers No. 222*. OECD Publishing.

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.

Pardo, A. & Siemens, G. (2013). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*, 438-450.

PMCA (2013). *Big Data – Bedeutung, Chancen, Datenschutz* [Big Data – Meaning, Chances, Data Protection] [Press release]. Retrieved from http://www.pmca.at/pmca-impuls-16-9-2013-pressemeldung-big-data-bedeutung-chancen-datenschutz/

Range, L., Embry, T., & MacLeod, T. (2001). Active and passive consent: a comparison of actual research with children. *Ethical Human Sciences and Services, 3,* 23-31.

Rodotà, S. (2009). Data protection as a fundamental right. In Gutwirth, Y. Poullet, P. de Hert, C. de Terwangne,S. Nouwt (Eds.), Reinventing data protection? (pp. 77-82). Dortdrecht: Springer.

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.

Sclater, N. (2014, October 29). N*otes from Utrecht Workshop on Ethics and Privacy Issues in the Application of Learning Analytics* [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/

Shalowitz, D.I. & Miller, F.G. (2008). Communicating the results of clinical research to participants: Attitudes, practives, and future directions. *PLoS Med, 5*, 714-720. Retrieved December 11, 2014 from http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0050091

Slade, S. & Prinsloo, P. (2013). Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist, 57,* 1509-1528.

Spiekerman, S. & Cranor, L.F. (2009). Engineering privacy. *IEEE Transactions on Software Engineering, 35,* 67-82.

Stiles, R.J. (2012). Understanding and managing the risks of analytics in higher education: A guide. *EDUCAUSE*. Retrieved December 9, 2014 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. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2010)* (pp. 1553-15262). ACM: Atlanta.

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.

The Open University (2014, Septemer). *Policy on ethical use of student data for learning analytics.* Milton Keynes: The Open University. Retrieved December 2, 2014 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: The White House.

US Government (2004). *Code of Federal Regulations. Education. Family Educational Rights and Privacy. 34 CFR Part 99.* Washington: Department of Education.

Van Harmelen, M. & Workman, D. (2014). *JISC CETIS Analytics Series: Vol.1 No.3, Analytics for learning and teaching.* University of Bolton, 2012. Retrieved August 19, 2014 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.

Willis, J.E. (2014, August). Learning analytics and ethics: A framework beyond utilitarianism.

EDUCAUSE Review. Retrieved October 28, 2014 from http://www.educause.edu/ero/article/learning-analytics-and-ethics-framework-beyond-utilitarianism

Willis, J.E. & Pistilli, M.D. (2014). Ethical discourse: Guiding the future of learning analytics. *EDUCAUSE Review*. Retrieved December 1, 2014 from http://www.educause.edu/ero/article/ethical-discourse-guiding-future-learning-analytics

Willis, J.E., Campbell, J.P., & Pistilli, M.D. (2013, May). Ethics, big data, and analytics: A model for application. *EDUCAUSE Review*. Retreived October 28, 2014 from http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application