Mapping a Privacy Framework to a Reference Model of Learning Analytics

Yong-Sang Cho
KERIS
Dept. of Standard & Quality
Daegu
Republic of Korea
zzosang@keris.or.kr

Tore Hoel
Oslo and Akershus University
College of Applied Sciences
Oslo
Norway
Tore.Hoel@hioa.no

Weiqin Chen
Oslo and Akershus University
College of Applied Sciences
Oslo
Norway
Weiqin.Chen@hioa.no

ABSTRACT
This paper is a first exploration of how the privacy framework found in the ISO/IEC 29100 standard could be applied to learning analytics. In this case study a mapping is provided between the published ISO/IEC standard and the learning analytics framework under development as a reference model for learning analytics, ISO/IEC 20748. This mapping and the identified privacy requirements and principles will prove useful in designing learning analytics system as well as performing risk management to avoid privacy breaches.

CCS Concepts
I.6.4 [**Computing Methodologies**]: Model Validation and Analysis H.1.2 [**User/Machine Systems**]: Human Factors J.1 [**Administrative Data Processing**]: Education K.4.1 [**Public Policy Issues**]: Ethics, Privacy, Regulation

Keywords
Learning Analytics, Privacy Framework, Privacy Requirements, Data Sharing, Interoperability.

1. INTRODUCTION
In spite of growing interest for learning analytics, the application of these new technologies still faces the challenge of effective integration of various information and processes into a unified framework to support the development of an open and extensible learning analytics systems (Choi et al., 2014). Moreover, privacy and data protection are issues that engage policy makers globally resulting in new legislation and policies being introduced. [5, 6, 7]

The aim of this paper is to explore international ISO/IEC standard for privacy framework [1] in order to apply this standard to learning analytics. For the mapping from privacy framework to learning analytics, this paper will use the reference model for learning analytics, which is a project under development as ISO/IEC 20748 [2].

First, we do a case study on the privacy framework, known as ISO/IEC 29100 [1], looking for basic elements and principles relates to privacy. Also we briefly identify the processes and features of learning analytics. Based on this basis information the paper derives a mapping table from the privacy framework to learning analytics and assigns privacy principles to learning analytics processes.

The purpose of this research is to inform standards development by identifying privacy requirements and principles prior to system development for learning analytics.

2. CASE STUDY PRIVACY FRAMEWORK
The following sub-section summarizes the main features of ISO/IEC 29100 Privacy Framework.

2.1 Overview of ISO/IEC 29100
ISO/IEC 29100 is an international standard providing a high-level framework for the protection of Personally Identifiable Information (PII) within information and communication technology (ICT) systems. This standard describes organizational, technical, and procedural aspects in overall privacy framework. The privacy framework might be useful for organizations to define their privacy safeguarding requirements related to PII within an ICT environment by:

- specifying a common privacy terminology;
- defining the actors and their roles in processing PII;
- describing privacy safeguarding requirements; and
- referencing known privacy principles.

In addition, this standard might be used as a basis for additional privacy standardization initiatives, such as for:

- a technical reference architecture;
- the implementation and use of specific privacy technologies and overall privacy management;
- privacy controls for outsourced data processes;
- privacy risk assessments; or
- specific engineering specifications.

2.2 Basic Elements of Privacy Framework
The following components relate to privacy and the processing of PII in ICT systems and make up the privacy framework described in the ISO/IEC 29100 standard.

- actors and roles;
- interactions;
- recognizing PII;
- privacy safeguarding requirements;
- privacy policies; and
- privacy controls.

2.2.1 Actors and roles
According to ISO/IEC 29100 there are four types of actors who can be involved in the processing of PII, PII principals, PII controllers, PII processors and third parties.

PII principals provide their PII for processing to PII controllers and PII processors and, when it is not otherwise provided by applicable law, they give consent and determine their privacy preferences for how their PII should be processed. PII principals can include, for example, the learner listed in the school information system or learning management system.

A **PII controller** determines why (purpose) and how (means) the processing of PII takes place. The PII controller should ensure adherence to the privacy principles in this framework during the

processing of PII under its control (e.g., by implementing the necessary privacy controls).

A **PII processor** carries out the processing of PII on behalf of a PII controller, acts on behalf of, or in accordance with the instructions of the PII controller, observes the stipulated privacy requirements and implements the corresponding privacy controls.

Third party can receive PII from a PII controller or a PII processor. A third party does not process PII on behalf of the PII controller. Generally, the third party will become a PII controller in its own right once it has received the PII in question.

2.2.2 Interactions

The actors identified in ISO/IEC 29100 can interact with each other in a variety of ways. As far as the possible flows of PII among the PII principal, the PII controller and the PII processor are concerned, the following scenarios can be identified:

a) the PII principal provides PII to a PII controller (e.g., when registering for a service provided by the PII controller);
b) the PII controller provides PII to a PII processor which processes that PII on behalf of the PII controller (e.g., as part of an outsourcing agreement);
c) the PII principal provides PII to a PII processor which processes that PII on behalf of the PII controller;
d) the PII controller provides the PII principal with PII which is related to the PII principal (e.g., pursuant to a request made by the PII principal);
e) the PII processor provides PII to the PII principal (e.g., as directed by the PII controller); and
f) the PII processor provides PII to the PII controller (e.g., after having performed the service for which it was appointed).

As far as the possible flows of PII among the PII controllers and PII processors on the on hand, and third parties on the other hand are concerned the following scenarios can be identified:

g) the PII controller provides PII to a third party (e.g., in the context of a business agreement); and
h) the PII processor provides PII to a third party (e.g., as directed by the PII controller).

2.2.3 Recognizing PII

According to ISO/IEC 29100, several factors need to be taken into account to determine whether or not a natural person should be considered identifiable. In particular, account should be taken of all the means, which can reasonably be used by the privacy stakeholder holding the data, or by any other party, to identify that natural person. ICT systems should support mechanisms that will make the PII principal aware of such PII and provide the natural person with appropriate controls over the sharing of that information.

Identifier is a very clear way for the PII principal to recognize PII. Information can be considered to be PII if it contains or is associated with an identifier, which refers to natural person (e.g., social security number), or which can be related to a natural person, e.g., a passport number, or an account number), or which can be used to communicate with an identified natural person (e.g., a precise geographical location, or a telephone number).

Other distinguishing characteristics can be used to identify PII. If information contains or is associated with a characteristic, which distinguishes a natural person from other natural persons, then it needs to be considered PII. In addition, if a combination of several attributes taken together distinguishes this natural person from other natural persons, then this case is also considered for PII. Any attribute which takes on a value, which uniquely identifies a PII principal is to be considered as a distinguishing characteristic.

Any information linked to a PII principal, such as medical records, financial profiles, or the personal interests derived from tracking use of Internet websites, also might be considered as PII. If the relationship with an identifiable natural person can be established, such information must also be treated as PII.

Pseudonymous data is one kind of ways to recognize PII. In order to restrict the ability of PII controllers and processors to identify the PII principal, identity information can be replaced by aliases. This replacement is usually performed by a PII provider before transmitting the PII to a PII recipient. The substitution is considered pseudonymization provided:

a) the remaining attributes linked to the alias do not suffice to identify the PII principal to whom they relate; and
b) the alias assignment is such that it cannot be reversed by reasonable efforts of the privacy stakeholders other than those that performed them.

Pseudonymization contrasts with anonymization. Anonymization processes also fulfill properties (a) and (b) above, but destroy linkability. During anonymization, identity information is either erased or substituted by aliases for which the assignment function or table is destroyed. Thus, anonymized data is no longer PII.

Metadata is also a way that it is not readily visible to the system user (i.e. to the PII principal). For instance, if PII principal's name stored as metadata in the properties of a document, and comments or tracked changes stored as metadata in a word processing document, then PII principal can decide whether the PII should not be processed in such a way or be shared publicly.

Unsolicited PII may be stored in an ICT system by PII controller or PII processor. The risk of collecting unsolicited PII can be reduced by considering privacy safeguarding measures at the time of the design of the system (also referred to as the concept of "privacy by design").

Sensitive PII contains healthcare information, such as medical prescriptions. PII must be treated as sensitive PII where such inference and knowledge of the identity of the PII principal is reasonably possible.

2.2.4 Privacy safeguarding requirements

ISO/IEC 29100 described that privacy safeguarding requirements can relate to many different aspects of PII processing, e.g., the collection and retention of PII, the transfer of PII to third parties, the contractual relationship among PII controllers and PII processors, the international transfer of PII, etc. Privacy safeguarding requirements can also vary in specificity. They might be very general in nature, e.g., consisting of an enumeration of high-level privacy principles, which an organization is expected to take into account when processing PII. However, privacy safeguarding requirements can also involve very specific restrictions on the processing of certain types of PII, or mandate the implementation of specific privacy controls.

The design of any ICT system that involves the processing of PII should be preceded by an identification of relevant privacy safeguarding requirements. The ICT systems involving the processing of PII should be resolved before those ICT systems are implemented. Organizations routinely perform broad risk management activities and develop risk profiles related to their

ICT systems. The privacy risk management process comprises the following processes:

- establishing the context, by understanding the organization (e.g., PII processing, responsibilities), the technical environment and the factors influencing privacy risk management (i.e. legal and regulatory factors, contractual factors, business factors and other factors);
- risk assessment, by identifying, analyzing and evaluating risks to PII principals (risks that they can be adversely affected);
- risk treatment, by defining privacy safeguarding requirements, identifying and implementing privacy controls to avoid or reduce the risks to PII principals;
- communication and consultation, by getting information from interested parties, obtaining consensus on each risk management process, and informing PII principals and communicating about risks and controls; and
- monitoring and review, by following up risks and controls, and improving the process.

2.2.5 Privacy policies

The organization should document its privacy policy in writing. Privacy policy should:

be appropriate to the purpose of the organization;

- provide the framework for setting objectives;
- include a commitment to satisfy applicable privacy safeguarding requirements;
- include a commitment to continual improvement;
- be communicated within the organization; and
- be available to interested parties, as appropriate.

2.2.6 Privacy controls

Organizations should identify and implement privacy controls to meet the privacy safeguarding requirements identified by the privacy risk assessment and treatment process. In addition, the identified and implemented privacy controls should be documented as part of the organization's privacy risk assessment.

As far as information security controls are concerned, it is important to note that not all PII processing requires the same level or type of protection. Organizations should distinguish among PII processing operations according to the specific risks they present to help determine which information security controls are appropriate in which instance.

2.3 Privacy Principles

The privacy principles described in ISO/IEC 29100 were derived from existing principles developed by a number of states, countries and international organizations. The standard focuses on the implementation of the privacy principles in ICT systems and the development of privacy management systems to be implemented within the organization's ICT systems. The following privacy principles form the basis for the basis for the standard.

1. Consent and choice
2. Purpose legitimacy and specification
3. Collection limitation
4. Data minimization
5. Use, retention and disclosure limitation
6. Accuracy and quality
7. Openness, transparency and notice
8. Individual participation and access
9. Accountability
10. Information security
11. Privacy compliance

2.3.1 Consent and choice

According to ISO/IEC 29100, adhering to the consent principle means:

- presenting to the PII principal the choice whether or not to allow the processing of their PII;
- obtaining the opt-in consent of the PII principal for collecting or otherwise processing sensitive PII;
- informing PII principals, before obtaining consent, about their rights under the individual participation and access principle;
- providing PII principals, before obtaining consent, with the information indicated by the openness, transparency and notice principle; and
- explaining to PII principals the implications of granting or withholding consent.

Provisions should be made to provide PII principals with the opportunity to choose how their PII is handled and to allow a PII principal to withdraw consent easily and free of charge.

2.3.2 Purpose legitimacy and specification

According to ISO/IEC 29100, adhering to the purpose legitimacy and specification principle means:

- ensuring that the purpose(s) complies with applicable law and relies on a permissible legal basis; and
- communicating the purpose(s) to the PII principal before the time the information is collected or used for the first time for a new purpose.

2.3.3 Collection limitation

According to ISO/IEC 29100, adhering to the collection limitation principle means:

- limiting the collection of PII to that which is within the bounds of applicable law and strictly necessary for the specified purpose(s).

2.3.4 Data minimization

According to ISO/IEC 29100, Data minimization is closely linked to the principle of "collection limitation" but goes further than that. Whereas "collection limitation" refers to limited data being collected in relation to the specified purpose, "data minimization" strictly minimizes the processing of PII.

Adhering to the data minimization principle means designing and implementing data processing procedures and ICT systems in such a way as to:

- minimize the PII which is processed and the number of privacy stakeholders;
- ensure adoption of a "need-to-know" principle;
- use or offer as default options, wherever possible, interactions and transactions which do not involve the identification of PII principals, reduce the observability of their behaviour and limit the linkability of the PII collected; and
- delete and dispose of PII whenever the purpose for PII processing has expired.

2.3.5 Use, retention and disclosure limitation

According to ISO/IEC 29100, adhering to the use, retention and disclosure limitation principle means:

- limiting the use, retention and disclosure (including transfer) of PII;
- limiting the use of PII to the purposes specified by the PII controller prior to collection;
- retaining PII only as long as necessary to fulfill the stated purposes, and thereafter securely destroying or anonymizing it; and
- locking (i.e. archiving, securing and exempting the PII from further processing) any PII when and for as long as the stated purposes have expired.

2.3.6 Accuracy and quality

According to ISO/IEC 29100, adhering to the accuracy and quality principle means:

- ensuring that the PII processed is accurate, complete, up-to-date, adequate and relevant for the purpose of use;
- ensuring the reliability of PII collected from a source other than from the PII principal before it is processed;
- verifying, through appropriate means, the validity and correctness of the claims made by the PII principal prior to making any changes to the PII;
- establishing PII collection procedures to help ensure accuracy and quality; and
- establishing control mechanisms to periodically check the accuracy and quality of collected and stored PII.

2.3.7 Openness, transparency and notice

According to ISO/IEC 29100, adhering to the openness, transparency and notice principle means:

- providing PII principals with clear and easily accessible information about the PII controller's policies, procedures and practices with respect to the processing of PII;
- including in notices the fact that PII is being processed, the purpose for which this is done, the types of privacy stakeholders to whom the PII might be disclosed, and the identity of the PII controller including information on how to contact the PII controller;
- disclosing the choices and means offered by the PII controller to PII principals for the purposes of limiting the processing of, and for accessing, correcting and removing their information; and
- giving notice to the PII principals when major changes in the PII handling procedures occur.

2.3.8 Individual participation and access

According to ISO/IEC 29100, adhering to the individual participation and access principle means:

- giving PII principals the ability to access and review their PII, provided their identity is first authenticated with an appropriate level of assurance;
- allowing PII principals to challenge the accuracy and completeness of the PII and have it amended, corrected or removed as appropriate and possible in the specific context;
- providing any amendment, correction or removal to PII processors and third parties to whom personal data had been disclosed, where they are known; and
- establishing procedures to enable PII principals to exercise these rights in a simple, fast and efficient way, which does not entail undue delay or cost.

2.3.9 Accountability

According to ISO/IEC 29100, adhering to the accountability principle means:

- documenting and communicating as appropriate all privacy-related policies, procedures and practices;
- setting up efficient internal complaint handling and redress procedures for use by PII principals;
- informing PII principals about privacy breaches that can lead to substantial damage to them as well as the measures taken for resolution;
- notifying all relevant privacy stakeholders about privacy breaches as required in some jurisdictions and depending on the level of risk;
- allowing an aggrieved PII principal access to appropriate and effective sanctions and/or remedies, such as rectification, expungement or restitution if a privacy breach has occurred; and
- considering procedures for compensation for situations in which it will be difficult or impossible to bring the natural person's privacy status back to a position as if nothing had occurred.

2.3.10 Information security

According to ISO/IEC 29100, adhering to the information security principle means:

- protecting PII under its authority with appropriate controls at the operational, functional and strategic level to ensure the integrity, confidentiality and availability of the PII, and protect it against risks such as unauthorized access, destruction, use, modification, disclosure or loss throughout the whole of its life cycle.

2.3.11 Privacy compliance

According to ISO/IEC 29100, adhering to the privacy compliance principle means:

- verifying and demonstrating that the processing meets data protection and privacy safeguarding requirements by periodically conducting audits using internal auditors or trusted third-party auditors;
- having appropriate internal controls and independent supervision mechanisms; and
- developing and maintaining privacy risk assessments in order to evaluate whether program and service delivery initiatives involving PII processing comply with data protection and privacy requirements.

3. REFERENCE MODEL FOR LEARNING ANALYTICS

The goal of learning analytics is to understand and improve learning and its environment, entailing the tasks of measurement, collection, analysis and reporting of data about learners and their contexts, while preserving confidential user information and protecting the identities of the users at the required level as needed [4]. These abstract steps of learning analytics under the protection of privacy policy can be depicted as an abstract workflow as shown in Figure 1.

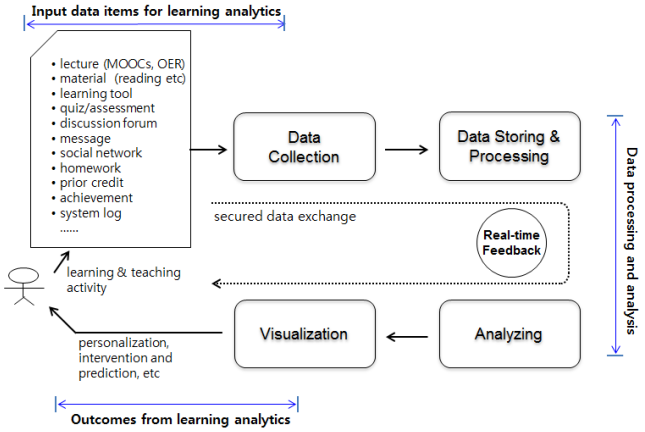

Figure 1. Reference model for learning analytics

The reference model for learning analytics has been developed by International Organization for Standardization (ISO), in particular subcommittee 36 for Information Technology for Learning, Education and Training (ITLET). The following sub sections describe the main features of the reference model for learning analytics, known as the ISO/IEC 20748 project.

3.1 Learning activity

Learning activity *means starting point of learning analytics, and learning activities are the source of the data for collection. In general learning activity is performed within diverse environments and tools. This process regulates both data release as well as data modeling or profiling to be able to generate learning activity data that could be used for analytics.*

As described in Figure 1, learning activities involve using learning content and tools, participating in collaborations or forums, etc. This process assumes service license agreement (SLA) between content/service provider and learner. However, to the knowledge of these authors, a there are no SLA in the market that describes learning analytics services with privacy requirements.

3.2 Data collection

Data collection *is the process of gathering and measuring information on variables of interest in the learning and teaching activities. In this process some features, such as authority and control of data source, interoperability of data, and efficiency of flow and exchange, are required for a system to work.*

The data produced by activities must be controlled by the privacy policy and managed and authorized accordingly to the users' consent. Therefore, prior to or during the data collection process, the direction for data treatment should be defined, such as anonymization or pseudonymization. This decision for the data treatment will influence the whole process of the learning analytics.

Data collection processes may be subject to conformance testing prior to storing data in a Temporary Data Store, such as the Event Store in IMS Caliper or Learning Record Store in xAPI, to be used in later processing.

3.3 Data storing and processing

Data storing and processing *is the process of preparing and storing data from heterogeneous sources for transport to data analysis, utilizing a standardized data model and representation fit for analysis.*

The learning activity data stored in temporary data store are processed by the data translator and/or filter. A general purpose of data filtering may be applied to the translation process driven by the filtering conditions to clean and transform the data. The results from translating or filtering are stored into a data store for analysis. In this process direction for retaining or preservation of the data should be adopted.

3.4 Analyzing

Analyzing *is the process of systematic examination of learning data in order to extract descriptive and possibly predictive knowledge about the learners and their contexts based on questions and models defined by the learning analytics system.*

Various analysis algorithms, such as predictive analytics, adaptive analytics, discourse analytics, and other assessment using ICT are applied via the analysis interface.

Privacy concerns exist wherever personally identifiable information or other sensitive information is collected and stored. Even if identity is pseudomized or anonymized, other distinguishing characteristics derived from learning activities may be managed through privacy risk management.

3.5 Visualization

Visualization *is the process of creating visual representation of abstract data including text and geographic information to allow users to see, explore, interact, and understand large amounts of information in analyzing and reasoning about data and evidence.*

A primary goal of visualization is to communicate information clearly and efficiently to users via the statistical graphics, plots, information graphics, tables, and charts selected, and thus making complex data more accessible, understandable and usable.

The dashboard information may show comparisons or progresses, recommendations, and real-time assessments, topic-based assessment, social-network graph, and so on.

Data access control issues are also reflected in this visualization process. Also data retainment issues need also to be taken into account.

3.6 Feedback

Feedback *actions serve the results of a cycle of learning analysis back to the learners and their contexts so that corrective actions can be taken.*

One of the feedback actions is recommendation for learning pathway based on analysis results. In this case, a lot of data are required and used for inference related to competency level, favouring of media types, etc.

4. ADOPTION OF PRIVACY FRAMEWORK TO THE LEARNING ANALYTICS

As introduced in section 2, ISO/IEC 29100 Privacy Framework is general in nature and places organizational, technical, and procedural aspects in an overall privacy framework pertaining to ITLET area. To adopt the standard privacy framework for learning analytics, basic elements and principles need to be interpreted to learning, education and training (LET) domain.

4.1 Identifying Privacy Elements

Table 1 is a draft sketch of a mapping from the basic elements of privacy framework to learning analytics based on the reference model.

Table 1. Mapping privacy elements from privacy framework to learning analytics

Types of elements	Basic elements of privacy framework	Interpretation to learning analytics
Actors and roles	PII principal	Learner, Parent or Teacher
	PII controller	Chief Information Officers of institutions
	PII processor	Administrator for the system
	Third party	Third party learning tool or service
Interactions	The scenarios described in subsection 2.2.2 can be applied without changes.	
Recognizing PII	Identifiers	Student ID, account for the system, or phone number on the data collection
	Other distinguishing characteristics	Sex, major, age, grade, or attended school, preference profile, etc., on the analyzing process
	Information linked to a PII principal	Any tracking data relates to learning activities on the analyzing process
	Pseudonymous data, metadata, unsolicited PII and sensitive PII described in subsection 2.2.3 can be applied without changes.	
Privacy safeguarding requirements	Described in general requirements	Need to define specifically for factors relates to learning analytics system
Privacy policies	Described in general ways	Need to define for specific purpose relates to learning analytics required by applicable law or regulation
Privacy control	Risk management is a central method in this process.	learning system as well as analytics system needs to apply organization's information security management framework.

Because ISO/IEC 29100 just provides a high-level framework, even if it can be applicable in a very wide range of domains or jurisdictional regions, it is required to refine the basic elements for learning analytics systems.

4.2 Adopting Privacy Principles to Workflows of Learning Analytics

Figure 2 is a draft sketch of assignment of privacy principles to the reference model for learning analytics. The number and name of privacy principles in Figure 2 is referenced from sub-section 2.3.

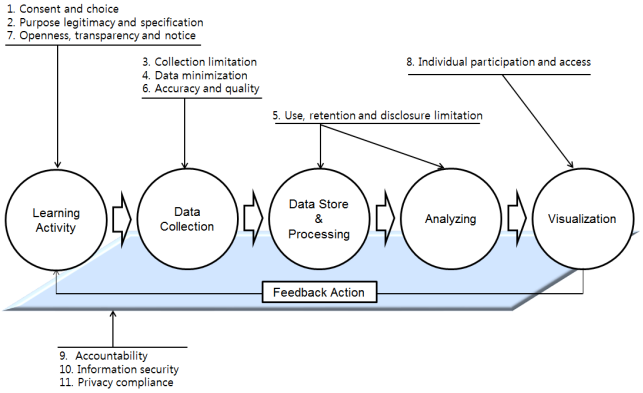

Figure 2. Privacy principles on learning analytics

Privacy principles for consent and choice, purpose legitimacy and specification, and openness, transparency and notice need to be applied in learning activity process prior to data collection.

Collection limitation, data minimization, and accuracy and quality principles need to be applied in data collection process rather than learning activities.

Use, retention and disclosure limitation principles influence the data storing and analyzing processes.

Individual participation and access principle is required to the visualization process.

The other principles - accountability, information security, and privacy compliance - are influencing all processes of learning analytics.

5. CONCLUSIONS

This paper has developed a first sketch of learning analytics applying privacy framework based on case study, in particular ISO/IEC 29100 those which provides high-level framework. Whereas privacy framework is referenced widely in ICT systems and jurisdictional regions, the field of learning analytics has just started to recognize the needs for protection and risk for privacy breaches. Even if this first sketch is very abstract, it might be used as basis to identify privacy requirements and principles for learning analytics.

In order to evaluate and improve the proposals more specific flows and description related to privacy issues on learning analytics need to be developed. This works should be combined with development of ISO/IEC 20748 Learning Analytics Interoperability projects in the standards community [3]. One of possible option may be new part of the ISO/IEC 20748, titled 'Application of privacy framework for learning analytics'.

6. References

[1] ISO/IEC (2011). ISO/IEC 29100 Information technology - Security techniques - Privacy framework.

[2] ISO/IEC (2015). ISO/IEC PDTR 20748-1 LAI - Part 1: Reference model.

[3] ISO/IEC (2015). ISO/IEC PDTR 20748-2 LAI - Part 2: System requirements.

[4] Bae, J.H., Cho, Y.S., & Lee, J.H. (2015). Designing a Reference Model for Learning Analytics Interoperability. Proceedings of the 23rd International Conference on Computers in Education.

[5] European Commission. 2012. On the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). COM(2012) 11 final

[6] European Commission. 2013. Data Protection Directive (Directive 95/46/EC)

[7] EU Commission. (2016). EU Commission and United States agree on new framework for transatlantic data flows: EU-US Privacy Shield. Press release Strasbourg, 2 February 2016. Retrieved from http://europa.eu/rapid/press-release_IP-16-216_en.htm