Privacy Requirements for Learning Analytics – from Policies to Technical Solutions

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

Yong-Sang Cho
KERIS
Seoul
Republic of Korea
zzosang@keris.or.kr

ABSTRACT
This paper is a first exploration of how privacy requirements could influence the design of each of the processes in a learning analytics framework model under development. The international organization for standardization, ISO/IEC JTC 1/SC36, is working on a reference model specifying a learning analytics process loop of six processes: Learning Activity, Data Collection, Data Storing and Processing, Analyzing, Visualization, and Feedback Actions. All these processes are, according to the framework, influenced by privacy policies.

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, Data Sharing, Trust, Control of data, Privacy by Design, Interoperability.

1. INTRODUCTION
The subcommittee (SC) 36 of the joint technical committee (JTC) 1 under international organization for standardization ISO/IEC launched in 2015 a new Working Group 8 on learning analytics interoperability (LAI). The first work item of this group was to develop a multi-part standard for LAI. Part 1 of this standard will be a reference model describing the key concepts of a workflow for learning analytics (LA).

The aim of this paper is to explore how Privacy Protection and Privacy Policy affect the different processes in the draft LA workflow model of SC36/WG8. This paper, therefore, will use the workflow model in Figure 1 to structure a systematic exploration of interoperability requirements related to privacy.

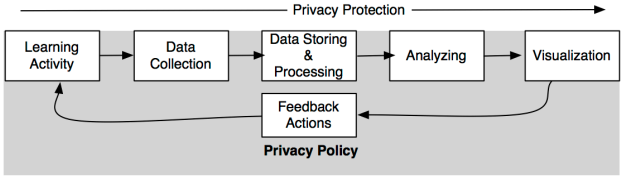

Figure 1: Workflow model of the learning analytics process

First, we do a review of research literature on learning analytics, looking for aspects related to data sharing, privacy, data protection, etc. Based on this review we analyze each sub process in order to glean requirements of technical, organizational or policy nature. The result is a table of requirements matched to LA processes. The paper concludes with an outlook of how this work could be brought forward and what contribution the international standards community could make.

This research fits in a Design Science tradition [15, 16] as it is positioned in the first Relevance Cycle (addressing requirements and field testing) of the three research cycles of Design Science [15]. The purpose of the research is to inform standards development by a structured and iterative solicitation and testing of requirements related to privacy issues matching the key concepts of a LA framework model.

2. BACKGROUND
Within the LA research community ethics and privacy have been identified as major issues to address in scaling up learning analytics in education and the workplace [11]. Also in the general discourse on big data it is recognized that "[b]ig data poses big privacy risks" [38].

A study from The European Union Agency for Network and Information Security (ENISA) [9] concluded, "privacy and data protection features are, on the whole, ignored by traditional engineering approaches when implementing the desired functionality". ENISA recommends that standardization bodies should include privacy considerations in the standardization process and provide standards for interoperability of privacy features.

This gap between requirements and design needs to be addressed. In this paper we are exploring how ethical principles and best practices could be turned into technical specifications for architectures and tools. However, in searching for 'hard requirements' we bear in mind that interoperability is a multi-layered concept. We will search for requirements for both the technical, organizational and political (related to practices) challenges the new LA technologies pose.

3. LITERATURE REVIEW
For each process the authors' definition of the process introduces the sub section.

3.1 Learning Activity
Learning Activity *is the starting point of learning analytics, and learning activities are the source of data for collection. In general, Learning Activity is performed within heterogeneous environments, using a mixture of 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.*

The centrality of activities and activity flows is highlighted in a recent report from the LACE project exploring requirements, specifications and adoption of LAI [14]. This report gives an in-depth discussion of specifications like xAPI, IMS Caliper and

W3C Activity Streams and their capabilities in binding activity statements for exchange between systems. Neither of these specifications has embedded privacy requirements, which has raised some debate in the standards community [14]. In order to understand why privacy has come up as an issue in this context we need to understand that these specifications are not only providing a language for talking about activity, but also a means for storing this conversation in LA systems.

Language or speech is perhaps a good metaphor for further analysis because, in the tradition of discourse analysis, the analysis needs to reflect on active tools and actors. One is prompted to ask about actor strategies; how they influence the interaction, genres of discourse, the relationship between the text and the context, power relations etc. This sidestep in reasoning in this paper brings the learner back into the frame; LA is not only about data exhaust of learning, it is about the learners' strategies to achieve life goals. Therefore, it will be a big mistake to exclude from the analysis issues like embodiment of learning [30], learner agency and vulnerability [32], and contextual differences of formal and informal learning [39].

Rambusch and Ziemke [30] found that "activity often appears to be conceived of as a cognitive process which is somehow detached from the body, making the body just an additional tool for the subject and not an integral part of human cognition". Introduction of LA might strengthen the separation of activity (or mind) from the body, a position that Rambusch and Ziemke now find rejected by findings in the field of cognitive neuroscience. Embodiment of learning brings student vulnerability and learner agency to the fore when discussing how we collect and use student data. Vulnerability and agency are the lens Slade and Prinsloo [32] use to explore the ethics of LA. They conclude, in "the context of the fiduciary duty of HEIs and the asymmetrical power relationship, higher education cannot afford a simple paternalistic approach to the use of student data. Such an approach should not be considered appropriate given the complexities within the nexus of privacy, consent, vulnerability and agency" [32].

In maintaining the primacy of learner and her agency in sharing information on her learning activities we still have to situate learning in a more concrete sense. Thomas [39] argues that we have failed to recognize the primacy of 'physical situatedness' to our conceptions of learning itself, and that our difficulty in understanding and articulating the nature of learning is partly brought about by our inability to articulate where learning takes place. When taking in that the real and virtual places, as well as the bodies we inhabit, are essential to understand the nature of learning in our age, the concept of context is emerging as a key analytical construct. Awareness of the different contexts of learning would, in the view of the authors of this paper, lead to seeing informal vs. formal learning more as shades of grey than as black and white.

Context is key to understand what is meant by privacy from an individual's perspective. "When we find people reacting with surprise, annoyance, indignation, and protest that their privacy has been compromised, we will find that informational norms have been contravened, that contextual integrity has been violated" [27, p.25]. A context integrity view on privacy implies that issues of privacy (and data) protection do not have fixed solutions. "Where, for example, anonymizing data, adopting pseudonyms, or granting or withholding consent makes no difference to outcomes for an individual, we had better be sure that the outcomes in question can be defended as morally and politically legitimate. When anonymity and consent do make a difference, we learn from the domain of scientific integrity that simply because someone is anonymous or pseudonymous or has consented does not by itself legitimate the action in question" [5].

3.2 Data Collection

Data Collection *is the process of gathering and measuring information on variables of interest in 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 flip side of Learning Activity is Data Collection where the individual meets the tools, authorities and agents that pull the data together in preparation for LA. Who should be in control of this process? Does the learner need to let go and leave it to authorities to do their thing? This tension between the individual and the system will prevail throughout all processes of the LA workflow; and, depending what stakeholder role we play, the locus of control will shift between the individual and the system. Where one chooses to put the primacy depends on a host of political, cultural, social and pedagogical factors. However, at the end of the day it is also a question of the efficacy of the system. If only an amiable approach to the learner will release data, then more learner control may be a good direction to go. Or the same approach might be wise if the abundance of data is so big that a paternalistic approach is bound to fail. There are moral arguments why users should be in control of their learning activity data; however, there are also economic arguments, e.g., that it serves an open information market. "The simplest approach to defining what it means to “own your own data” is to go back to Old English Common Law for the three basic tenets of ownership, which are the rights of possession, use, and disposal" [28].

The above definition of ownership leads to a need for an ownership management system, which is close to what we describe as identity management systems in education. These systems are also under development in order to serve wider needs for the individual to maintain her digital identity. Mertens and Rosemann [25] have presented ideas for Digital Identity 3.0, "a platform that puts customers in charge of their data, integrates that data into an evolving and comprehensive representation of their digital and physical world, allows people to share that data selectively in order for the platform to pull proactive services towards them, and unlock new experiences as they connect to their world in new and exciting ways".

Digital Identity 3.0 systems are not a reality yet, and nobody knows if the dream of putting the customers in charge of their data will come true. In the meantime, a balanced view of the legitimate interests of the different actor roles in data exchange is advisable. In the process of revising the European data protection laws a number of "opinions" are published by 'Article 29 Data Working Party', an independent European advisory body on data protection and privacy. This process gives useful insights into this new landscape. In Opinion 06/2014 on the notion of legitimate interests of the data controller it is stated: "It also required data controllers to put in place mechanisms to demonstrate consent (within a general accountability obligation) and requested the legislator to add an explicit requirement regarding the quality and accessibility of the information forming the basis for consent" [2].

To give consent to share one's data, the data owners (i.e., the learners) will be more comforted if they know that the data controller have done everything they can to protect their personally identifiable information (PII). Anonymization is one

technique, which is well described by Article 29 in Opinion 05/2014:

"The Opinion concludes that anonymisation techniques can provide privacy guarantees and may be used to generate efficient anonymisation processes, but only if their application is engineered appropriately – which means that the prerequisites (context) and the objective(s) of the anonymisation process must be clearly set out in order to achieve the targeted anonymisation while producing some useful data. The optimal solution should be decided on a case-by-case basis, possibly by using a combination of different techniques, while taking into account the practical recommendations developed in this Opinion." [3]

A pragmatic approach to anonymization suggests that we are able to see the big picture of privacy protection. According to Rubenstein and Hartzog [31] the first law of privacy policy is there is no silver bullet.

"Neither technologists nor policymakers alone can protect us. But we have been missing the big picture. We should think of reidentification as a data release problem. Sound data release policy requires a careful equilibrium on multiple fronts: law and technology, data treatment and data controls, privacy and utility." [31]

A balanced approach to de-identification has implications for design, however, on that subject the research literature analyzed for this paper is quite silent.

3.3 Data Storing & 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.*

While devices, sensors and networks create large volumes and new types of data, and the cost of data storage is becoming negligible, there is a growing public interest in and demand for the re-use of these data [3].

It is the potential for re-use and the aggregation of datasets from different sources that make the Data Storing & Processing process so volatile when it comes to privacy. This is a new and unprecedented situation, which is exacerbated by the new use of social media in education. "As most work on privacy considers each record of data as corresponding to a unique individual, and records as independent, no mechanism is available to protect individuals' private information leaked through social links with other users" [21].

The fragility in the storage process is not reflected in the current development of architectures for LA, e.g., ADL Experience API (xAPI) [1] and IMS Caliper Analytics [19].

xAPI introduces a concept of a Learning Record Store (LRS). However, the LRS is only at the receiving end of activities streams, accessed by defined application program interfaces (APIs). IMS includes an EventStore as a development and demo environment rather than a component of a production Caliper system [14]. Neither xAPI nor IMS Caliper deal with privacy issues related to their storage concepts.

The discussion of storage and privacy is still on a conceptual level. In the new online (learning) culture with social data generated and accumulated online, "information privacy is dependent on not one single user, but on a web of users to whom this individual is connected and on the information that they disclose" [41]. Xu proposes the notion of privacy 2.0, acknowledging that information disclosure is co-constructed by users and their social connections, and that privacy responsibilities are distributed through such interpersonal relationships.

If not proposed for large-scale deployment in learning analytics, alternative storage models are being tested. de Montjoye, Shmueli, Wang and Pentland [26] are developing an Open Personal Data Storage solution that protects privacy of metadata through a solution they have termed SafeAnswers. "The goal of SafeAnswers is to turn an algorithmically hard anonymization and application-specific problem into a more tractable security one by answering questions instead of releasing copies of anonymized metadata" [26]. Kitto, Cross, Waters and Lupton [23] are developing a Connected Learning Analytics toolkit including a personalized LRS.

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 LA system.*

The Analyzing Process is where numbers meet the critical eye, where the abstraction potentially is fed back to the human agency that initiated the flow of data in the first place. We say potentially, because the agency could be delegated to a software agent, an algorithm. The ethical dilemmas pertaining to Analyzing rest in the polysemy of the concept of an analyzing agent, human or machine - or both. A human can be spoken to; in some sense, so is the case with a machine or an algorithm, however, the more abstruse the digital layers between the learner and the Analyzing is, the further from human appeal the conversation gets, and the more privacy could be in jeopardy.

Kitchin [22] has synthesized critical thinking about algorithms and details how they might be variously understood as "'black boxed'; heterogeneous, contingent on hundreds of other algorithms, and are embedded in complex socio-technical assemblages; ontogenetic and performative; and 'out of control' in their work".

In trying to get a handle on what algorithms do with the learner's data, "it may make little sense to interrogate any algorithm singularly, but rather to unpack complex 'algorithmic systems' that are both products of social practices and productive of ways in which other social practices are enacted" [40]. However, this 'unpacking' will prove too much to ask from the individual learner. "[A]lgorithms are difficult to deconstruct, they often have a multiplicity of creators, more than one rationality and can be programmed to evolve over time and 'learn' independently of their creators" [36]. Algorithms derive their social power from their complexity, and the complexity might be found to be too much for an individual to break.

As relegations of decisions about an individual's learning trajectory to automated processes based on algorithms and artificial intelligence "raises concerns about discrimination, self-determination, and the narrowing of choice" [38], countermeasures need to be taken to make the Analyzing process more 'privacy proof'.

"It is imperative that individuals have insight into the decisional criteria of organizations lest they face a Kafkaesque machinery that manipulates lives based on opaque justifications. While we recognize the practical difficulties of mandating disclosure without compromising organizations' "secret sauce," we trust that a distinction can be drawn between proprietary algorithms,

which would remain secret, and decisional criteria, which would be disclosed." [38]

The questions initiating learning analytics should be disclosed. However, the layer of 'secret sauce' needs also to be as "thin" as possible. The educational community calls for transparency around algorithms and metrics [33, 34]. There are interoperability standards that make exchange of models and methods [7]. Openness of process, algorithms, and technologies are important for innovation and for meeting the varying contexts of implementation [35, 17]. If analytics technologies are built upon openness the academic community could through research and critique provide some checks and balances that could work against "secret processes and opaque and unaccountable algorithms [that] can hide arbitrary or unfair decision making" [24].

3.5 Visualization

Visualization *is the process of interpreting and presenting the analysis result of LA data in a (mainly) visual form that contributes to the understanding of the meaning of the data.*

Current information visualization techniques often assume unrestricted access to data, as if confidentiality and privacy were not an issue. However, in this LA Visualization process communication of meaning is the key task. The concept of 'meaning' implies a perspective and an actor role. What is meaningful for the teacher is not necessarily meaningful for the learner. Ideally, a visualization targeted a teacher should be rendered quite differently than the ones targeted a learner.

There is some research on how privacy conditions should be visualized, e.g., Ghazinour et al. [13] delivered a model for privacy policy visualization. This model presupposes that a privacy policy model is defined with properties like purpose, visibility, granularity, retention and constraints. A further development of this work would be to develop user profiles that could be matched with the privacy policy.

Differentiated access to visualizations is of course explored in other fields as health and security, and there are techniques for transformation of data to remove identifiers and sanitize data. For example, Dasgupta and Kosara [10] have proposed a an adaptive technique that based on knowledge about the sensitivity of the data, computes a clustered representation on the fly, which through the use of screen-space privacy metrics, provides the users with visualizations adapted to their screen parameters and interaction.

3.6 Feedback Actions

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.*

The feedback process of LA is concerned with mediation of the knowledge gleaned from the data. Not only the learners themselves are on the receiving end; course designers, developers of learning materials, teachers and administrators, etc. are also party to learning analytics.

However, none of these actors are guaranteed that learning analytics results actually are fed back to them. Prinsloo and Slade [29] have been concerned about the ethical implications of knowing, not knowing, and knowing more. The solution is not necessarily in knowing more, but ensuring that once we know, we respond in ethical, caring, disciplined and context-appropriate ways. In education, opting out of learning analytics is not an option, at least not an option given by many providers [29]. The more mandatory a system is, the more checks and balances are needed outside the system itself, tied to the policies of implementation. However, policies and codes of practice [33] are not enough, privacy concerns should also be built into the systems themselves as hard, technical requirements. The alternative to not embarking on such a design work may have grave consequences. It is observed that once the ramifications of big data analytics sink in, people will likely become much more conscious of the ways they are being tracked. The chilling effects on all sorts of behaviors, also learning, could become considerable [38].

4. PRIVACY REQUIREMENTS FOR LAI – SOME SUGGESTIONS FOR DESIGN

Interoperability requirements come in different 'measures', depending on stakeholder perspective and on what interoperability challenge having priority. Cooper and Hoel [7] found that the European Interoperability Framework (EIF) [12] provides a unified view on interoperability and a useful structure to consider issues of data sharing. In this paper three interoperability levels (policy, organizational and technical) are chosen to describe requirements for design of privacy solutions.

Table 1 is the result of a first design cycle based on the literature review in Section 3 and use cases and design efforts, first in the ad hoc group leading up to the establishment of the SC36 WG8, and later in WG8. (It is important to note that members of WG8 are not responsible for this proposal, and that this standards group will not be committed by the results of this analysis.)

Table 1 identifies components of different nature related to each LA process. However, these components may be parts of one or more systems, and the systems may serve other ends than just learning analytics. The components, as they are defined in this first version, may also be partly overlapping in functions.

Table 1: Privacy requirements for processes in a learning analytics process cycle

Type of Process	***Policy requirements***	***Organizational requirements***	***Technical requirements***
Learning Activity	LA Process Ownership Policy Data Release Policy	Context Management Data Release Management	Consent to share Activity Data Service Exchange Monitor
Data Collection	Data Governance Policy	Metrics Management Identity Management (incl. de-identification)	Contextualizing LA Service Identity Management & Privacy Protection Service
Data Storing & Processing	Data Governance Policy	Cloud Storage Policies and Solutions	Personal Learning Record Store
Analyzing	Algorithms & Predictive Models Stewardship Policy	Participation in LA development networks (Open LA	Decision criteria browser

		Architectures)	
Visualization	LA Visualization and Feedback Policy	Code of Practice regarding access to Visualization from LA	Privacy enhanced visualizations
Feedback Actions	LA Visualization and Feedback Policy	Code of Practice regarding Feedback from LA	Accountability systems for feedback from LA systems

In the following the components identified in Table 1 are described more in detail, e.g., in terms of role, function, stakeholder interest, and relation to privacy concerns.

4.1 Policy requirements

LA Process Ownership Policy: This is a high level policy giving the base rules of why a LA system is put in place, who are using it, why are they using it, and how. For example, the policy defines if the learner has primacy in the process, and when conflicts of interests occur, whose interest should have priority. Stakeholder roles are clarified in this policy, which will position the LA system on a continuum between a radical learner centered system and a fully institutionally controlled system (refer Academic Analytics [4, 6].

Data Release Policy: When learners are asked to share data from their informal learning activities and institutions are informing about how data is released from LMS and library systems, the Data Release Policy is invoked. The policy should give the data subject enough information to be able to build her 'trust model' of the system in use.

Data Governance Policy: This policy outlines the responsibilities of the data controller, i.e., the institution or the vendor. The policy should explain how metrics are developed and used, and how the data flows between systems are managed. Furthermore, the policy should give high-level principles for choice of data storage and processing technologies, e.g., whether a centralized data warehousing strategy is followed, or whether a more decentralized strategy is chosen. Principles for privacy protection measures for exchange of data between systems should also be defined in this policy.

Algorithms & Predictive Models Stewardship Policy: Whether LA systems are developed in-house or provided by a vendor, algorithms and predictive models are subject to continuous development with feedback from a diverse set of users. Whether this is taking place as a black box activity or is led by a policy of openness and transparency is up to the institution. However, the stewardship policy should be described and known to the users of the system.

LA Visualization and Feedback Policy: This policy delineates how the LA system responds to its target groups, e.g., learners, teachers, course designer, etc. Principles for design and use of LA dashboards are described, and the policy defines how LA process loops should be managed.

4.2 Organizational requirements

Context Management: This system defines different contexts for use of LA systems and allows users to define rules of system behavior based on these context definitions.

Data Release Management: This system runs the rules defined in the Data Release Policy. A Data Release Management system could be shared by different institutions or school owners.

Metrics Management: Metrics defines what aspects of the released data will be collected, based on data structures developed according to the Data Governance Policy. Metrics Management controls the calibration and further development of metrics and makes sure that stakeholders are able to exercise the necessary control of the perspectives taken on the collected data.

Identity Management (incl. de-identification): All parts of the LA workflow depends on unique identification of data elements. However, all data grains do not need to be linked to personally identifiable information. An extended identity management system used across institutions and platforms should broker access to data and systems, and reveal enough, —but not more— information on the data subject to do analysis on the questions defined.

Cloud Storage Policies and Solutions: Cloud storage is an alternative that will be chosen by many institutions and vendors for economical, administrative, technical and other reasons. A Policy will describe the legal state-of-art (e.g., International Safe Harbor Privacy Principles). In addition, the Policy will define positions regarding institutional approaches to data warehousing; personal storage of learning data and possibilities for data sharing; SafeAnswers solutions where LA systems have only indirect access to learner activity data; etc.

Participation in LA development networks (Open LA Architectures): Networks of universities, research institutions, vendors etc. plays an important role in LA development through sharing of data for training algorithms, testing visualization techniques, developing predictive models, etc. By defining policies on commitment to Open Learning Analytics Architectures institutions will increase transparency, trust building and privacy.

Code of Practice regarding access to Visualization from LA: This Code of Practice will describe how LA providers use visualizations to improve learning and its contexts for different target groups.

Code of Practice regarding Feedback from LA: This Code of Practice will describe how different actors within a LA system get access to, use and feed back on the results of LA process cycles.

4.3 Technical Requirements

Consent to share Activity Data Service: Via simple and innovative tools this service will allow the user to give the LA system access to self-declared data.

Exchange Monitor: This tool gives the user an intuitive overview of all data used for LA to support her learning.

Contextualising LA Service: In order to see if a metric is "fair" users of LA systems should be able to see how interventions supported by LA are sitting in an overall process of teaching and learning. The Contextualizing LA Service shows how data metrics are related to learning programs, learning goals, competency frameworks, etc. This service could be part of a more comprehensive TEL system, e.g., a system that bridges a LMS and a competency and activity profile system.

Identity Management & Privacy Protection Service: LA tools and services should use external Identity Management & Privacy Protection Service. By serving more than one system, such a service will allow data sharing without compromising privacy and data protection.

Personal Learning Record Store: A Personal Learning Record Store gives the learner more control over how data are shared and used. This tool could coexist with other storage technologies, or

could be part of an open participatory learning ecosystem, "which operates outside of the traditional LMS" [23].

Decision criteria browser: This tool lets the user explore the algorithms and predictive models used in the LA system in order to learn how they respond to different learning scenarios. The browser could make use of synthetic data to allow the user to test how the LA system would respond to hypothetical learning trajectories.

Privacy enhanced visualizations: This tool adapts the rendering of visualizations according to the privacy preferences defined for different user groups and by the users themselves.

Accountability systems for feedback from LA systems: This system is used for feedback actions and keeps record of the interaction between the learner and the system, teachers and others that take place in the learning process.

5. CONCLUSIONS AND OUTLOOK

This paper has developed a first list of requirements at a policy, organizational, and technical level for enhanced privacy and data protection in LA systems based on a literature review scaffolded by a workflow model of LA processes. Approaching the rapidly growing corpus of 'privacy and big data literature' using the ISO LA workflow model as a lens proved beneficial for soliciting requirements. However, the proposed design solutions are just a first step towards translating 'soft' privacy policies into 'hard' organizational and technical solutions.

In order to validate the proposals put forward in this paper the authors would advice a multi-pronged approach. First, there is a need for more use cases. These should be solicited following as a template the ISO/IEC framework model (Figure 1) that has been further developed in this paper. The use cases should be used to drive a new development cycle based on the design requirements identified in this paper.

Second, the proposals should also be tested against privacy frameworks developed in other contexts. ISO/IEC 29100 [20] defines a privacy framework for the protection of personally identifiable information ICT systems. This standard is general in nature and places organizational, technical, and procedural aspects in an overall structure. How will the actors and roles defined in our work fit within the broader framework? And how will the possible flows of PII described in this standard among actors (i.e. PII principal, PII controller, and PII processor) and a third party be handled in a LA context?

We would encourage the standards community to pursue the direction of development of privacy solutions put forward in this paper. There is a need to specify what privacy policies entails in a LA context, and we would like to suggest this as a new and separate part in the multi-part standard ISO/IEC 20748 under development.

6. References

[1] ADL, 2013. Experience API v 1.0.1. Retrieved from https://github.com/adlnet/xAPI- Spec/blob/master/xAPI.md [Accessed December 28, 2015].

[2] Article 29. (2014). Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 844/14/EN WP 217. Retrieved from http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/index_en.htm

[3] Article 29. (2014). Opinion 05/2014 on Anonymisation Techniques. 0829/14/EN WP216. Retrieved form http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/index_en.htm

[4] Baepler, Paul and Murdoch, Cynthia James (2010) "Academic Analytics and Data Mining in Higher Education," International Journal for the Scholarship of Teaching and Learning: Vol. 4: No. 2, Article 17. Available at: http://digitalcommons.georgiasouthern.edu/ij-sotl/vol4/iss2/17

[5] Barocas, S. & Nissenbaum, H. (2015) Big Data's End Run around Anonymity and Consent. In J. Lane, V. Stodden, S. Bender & H. Nissenbaum (eds.), Privacy, big data, and the public good, (pp. 44-75). New York, NY: Cambridge University Press.

[6] Campbell, J. P., & Oblinger, D. G. (2007). Academic Analytics. Educause Article. Retrieved from https://net.educause.edu/ir/library/pdf/PUB6101.pdf

[7] Cooper, A. & Hoel, T. (2015). Data Sharing Requirements and Roadmap. Public Deliverable D7.2 from the LACE project. Retrieved from http://www.laceproject.eu/deliverables/d7-2-data-sharing-roadmap/

[8] Cooper, A. (2014). Specifications and Standards - Quick Reference Guide. Retrieved from http://www.laceproject.eu/dpc/learning-analytics-data-sharing-current-examples- 2014/

[9] Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Le Metayer, D., Tirtea, R., & Schiffner, S. (2015). Privacy and Data Protection by Design – from policy to engineering. The European Union Agency for Network and Information Security (ENISA). Retrieved from https://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/privacy-and-data-protection-by-design

[10] Dasgupta, A. & Kosara, R. (2011). Adaptive Privacy-Preserving Visualization Using Parallel Coordinates, in Visualization and Computer Graphics, IEEE Transactions on Visualization and Computer Graphics, vol.17, no.12, pp.2241-2248, Dec. 2011 doi: 10.1109/TVCG.2011.163

[11] Drachsler, H., Hoel, T., Greller, W., Kickmeier-Rust, M., Steiner, C. & Griffiths, D. (2016) Ethics and Privacy for Learning Analytics – a Review of Current Issues and their Solutions. LACE project Review. Retrieved from www.laceproject.eu

[12] European Commission. (2010). European Interoperability Framework (EIF) for European public services. Annex 2 to the Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions 'Towards interoperability for European public services'. Retrieved from http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

[13] Ghazinour, K., Majedi, M., & Barker, K. (2009). A Model for Privacy Policy Visualization (pp. 335–340). Presented at the 2009 33rd Annual IEEE International Computer Software and Applications Conference, IEEE. http://doi.org/10.1109/COMPSAC.2009.156

[14] Griffiths, D., Hoel, T. & Cooper, A. (2016) Learning Analytics Interoperability: Requirements, Specifications and Adoption. Public Deliverable D7.4 from the LACE project (European Commission Seventh Framework Programme, grant number 619424).

[15] Hevner, A. (2007). A Three Cycle View of Design Science Research. Scandinavian Journal of Information Systems, 2007, 19(2):87-92, 1–6.

[16] Hevner, A., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. Mis Quarterly, 28(1), 75–105.

[17] Hoel, T. & Chen, W. (2014). Learning Analytics Interoperability – Looking for Low-Hanging Fruits. Wu, Ying-Tien Supnithi, Thepchai Kojiri, Tomoko Liu, Chen-Chung Ogata, Hiroaki Kong, Siu Cheung Kashihara, Akihiro (Red.), Workshop Proceedings of the 22nd International Conference on Computers in Education. 39. s. 253-264. Asia-Pacific Society for Computers in Education.

[18] Hoel, T., & Chen, W. (2015). Privacy in Learning Analytics – Implications for System Architecture. In Seta, K. and Watanabe, T.(Eds.) Proceedings of the 11th International Conference on Knowledge Management.

[19] IMS (2015). IMS Global Learning Consortium Caliper Analytics, Retrived from http://www.imsglobal.org/activity/caliperram.

[20] ISO/IEC (2011). Information technology — Security techniques — Privacy framework. ISO/IEC 29100:2011(E)

[21] JIA, H., & Xu, H. (2015). Big Social Data: New Challenges to Information Privacy. In Seta, K. and Watanabe, T.(Eds.) Proceedings of the 11th International Conference on Knowledge Management.

[22] Kitchin, R. (2014). Thinking critically about and researching algorithms. The Programmable City working paper 5. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2515786

[23] Kitto, K., Cross, S., Waters, Z., & Lupton, M. (2015). Learning analytics beyond the LMS (pp. 11–15). Presented at the the Fifth International Conference, New York, New York, USA: ACM Press. http://doi.org/10.1145/2723576.2723627

[24] MacCarthy, M. (2014). Student Privacy: Harm and Context. International Review of Information Ethics, 21, 11–24. Retrieved from: i-r-i-e.net/inhalt/021/IRIE-021-MacCarthy.pdf

[25] Mertens, W. & Rosemann, M. (2015). Digital Identity 3.0: The Platform for People. Working Paper published from Queensland Univeristy of Technology, Austrailia. Retreived from http://eprints.qut.edu.au/87075/

[26] de Montjoye Y-A, Shmueli E, Wang SS, Pentland AS (2014) openPDS: Protecting the Privacy of Metadata through SafeAnswers. PLoS ONE 9(7): e98790. doi:10.1371/journal.pone.0098790

[27] Nissenbaum, H. (2014). Respect for Context as a Benchmark for Privacy Online: What It Is and Isn't. In Dartiguepeyrou, C. (ed.) The Futures of Privacy. ISBN 978-2-915618-25-9. Fondation Télécom, Institut Mines-Télécom, February 2014 Edition: Uniqueness, pp. 19-30.

[28] Pentland, A. (2009). Reality Mining of Mobile Communications: Toward a New Deal on Data. The Global Information Technology Report

[29] Prinsloo, P., & Slade, S. (2015, March). Student privacy self-management: implications for learning analytics. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (pp. 83-92). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2723585 is

[30] Rambusch, J., & Ziemke, T. (2005). The role of embodiment in situated learning. In Proceedings of the 27th Annual Conference of the Cognitive Science Society : CogSci05. - Mahwah, NJ : Lawrence Erlbaum Associates. - 0-9768318-1-3 - 978-0-9768318-1-5 ; s. 1803-1808

[31] Rubinstein, I. S., & Hartzog, W. (2015). Anonymization and Risk. NEW YORK UNIVERSITY SCHOOL OF LAW PUBLIC LAW LEGAL THEORY RESEARCH PAPER SERIES (Vol. 15, pp. 1–55). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2646185

[32] Slade & Prinsloo 2015 (workshop) and 2016 (in press)

[33] Sclater, N. (2015). A taxonomy of ethical, legal and logistical issues of learning analytics v1.0, 1–12.

[34] Sclater, N. (2015). Developing the Code of Practice. Effective Learning Analytics. Jisc. Available at http://analytics.jiscinvolve.org/wp/2015/01/30/developing-the-code-of-practice/

[35] Siemens, G., Gašević, D., Haythornthwaite, C., BAKER, R. S. J. D., & Dawson, R. (2011). Open Learning Analytics: an integrated & modularized platform. Retrieved from http://www.solaresearch.org/OpenLearningAnalytics.pdf

[36] Souto-Otero, M., & Beneito-Montagut, R. (2016). From governing through data to governmentality through data: Artefacts, strategies and the digital turn. European Educational Research Journal, 15(1), 14–33. http://doi.org/10.1177/1474904115617768

[37] Steiner, C., Masci, D., Johnson, M., Türker, A., Drnek, M., & Kickmeier-Rust, M. 2014. Privacy and Data Protection Policy. Deliverable D2.3 from LEA's BOX project. Online at http://css-kmi.tugraz.at/mkrwww/leas-box/downloads/D2.3.pdf, accessed 2015-10-30

[38] 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, Vol. 11 (5).

[39] Thomas, H. (2010). Learning spaces, learning environments and the dis"placement" of learning. British Journal of Educational Technology, 41(3), 502–511. http://doi.org/10.1111/j.1467-8535.2009.00974.x

[40] Williamson, B. (2016). Digital education governance: An introduction. European Educational Research Journal, 15(1), 3–13. http://doi.org/10.1177/1474904115616630

[41] Xu, H. (2012). Reframing Privacy 2.0 in Online Social Networks. Journal of Constitutional Law, 1–26.