Workshop paper accepted for presentation at The International Workshop on Learning Analytics and Educational Data Mining (LAEDM 2016) in conjunction with the International Conference on Collaboration Technologies (CollabTech 2016), Kanazawa, Japan - September 14-16, 2016

Implications of the European Data Protection Regulations for Learning Analytics Design

Tore Hoel & Weiqin Chen
Oslo and Akershus University College of Applied Sciences,
{tore.hoel,weiqin.chen@hioa.no}

Abstract

By May 2018 the new European Union's data protection reform will become enforced in most European countries. The reform claims to ensure that the citizen receives clear and understandable information when his or her personal data is processed. The new rules also strengthen the individual's rights to be forgotten. Learning analytics (LA) systems must operate within the confines of the law. However, there is not much evidence in the LA research literature that data protection and privacy constraints have played an important role in systems designs. This paper outlines the new data protection requirements put in place in Europe and investigates the implications for the design of learning analytics systems.

1. Introduction

Is privacy a show-stopper for learning analytics (LA)? This question is the title of a recent review report from the European LACE project [1]. The report, however, does not give a definite answer to the question given the complexity of the panorama of issues presented by the new and emerging LA technologies and practices. The technical environments are increasingly complex, sometimes putting the users in control of managing their data, sometimes keeping them completely out for the loop. The range of data sources involved makes learning analytics enmeshed with multiple personal and societal issues that are not yet fully analysed. What makes the situation even more convoluted is the fact that the rise of learning analytics "is not presented simply as a more effective way to carry out educational activities, but also as a means to transform the context in which the new methods are embedded" [1]. In this situation educational authorities may be tempted to ask for a time-out in order to sort out legal and other issues related to large-scale implementation of LA.

Maybe it was the emergency break that was pulled when a school agency, The Norwegian Centre for ICT in Education, concluded that most probably the application of LA would be against the law unless a number of principles were not adhered to [2]. In the guidelines from the Centre, the principles of data protection were summarised under the headlines of "lawfulness, purpose limitation, data minimisation, data quality, storing and deletion, right to know what information is stored, and information safety". These principles are derived from the Norwegian *Personal Data Act* of April 2000 [3], which in turn builds on the *European Data Protection Directive* (Directive 95/46/EC). The text of the Act describes these principles succinctly: The data controller shall ensure that personal data are processed only if the data subject has consented; "b) are used only for explicitly stated purposes that are objectively justified by the activities of the controller, c) are not used subsequently for purposes that are incompatible with the original purpose of the collection, without the consent of the data subject, d) are adequate, relevant and not excessive in relation to the purpose of the processing, and e) are accurate and up-to-date, and are not stored longer than is necessary for the purpose of the processing" [3].

What is interesting to note in the guidelines from the Norwegian school agency is the interpretation of the principles in the *Personal Data Act*, which lead to the assumption that "the school owner is not able to maintain the most important principle of data protection: The data subject should be in control of and agree to how their own data are used" [2]. The agency asks "how will the school owner make sure that information only are used for learning and not for other purposes, for example to control pupils and teachers? (..) What is the boundaries between information that are relevant for learning and information that are not relevant, but nevertheless are of interest for registration and analysis" [2]. It is evident to anyone in the field of LA, that there is no clear answer to these questions.

What makes it even more necessary to explore how the emergent LA practices are grounded in current laws is the new European data protection reform that will become law by May 2018. According to a factsheet from the European Commission (EC), the "new General Data Protection Regulation (GDPR) will ensure that you receive clear and understandable information when your personal data is processed. Whenever your consent is required, it will have to be given by means of a clear affirmative action before a company can process your personal data. The new rules will also strengthen individuals' right to be forgotten, which means that if you

no longer want your personal data to be processed, and there is no legitimate reason for a company to keep it, the data shall be deleted" [4].

This paper raises the question of how GDPR will influence the design of LA systems and practices. Based on a review of LA research literature, the new data protection regulations, and current design efforts the authors develop requirements for a development of LA based on the principles of "data protection by design" and "data protection by default".

2. Related Work

A 2015 survey of European citizens' attitudes to data protection [5] concluded that only 15% felt they had complete control over the information they provided online; one in three people (31%) thought they had no control over it at all. Nine out of ten Europeans expressed concern about mobile apps collecting their data without their consent, and seven out of ten worried about the potential use that companies may make of the information disclosed. Given this massive concern about data protection it should be noted that data protection did not appear in the proceedings of the main conference of the LA research community[1] in 2014 and 2015, and only one time in 2016 [6]. This might imply that the principles of Data protection by design and by default inscribed in the GDPR would have a way to go in order to influence LA design and practices. However, other research and community exchange have put the issues of ethics, privacy and data protection on the international agenda.

In two papers Mason, Chen and Hoel [7, 8] have found that issues related to ethics and privacy are on the top of the list of concerns that researchers and practitioners in the emergent field of LA want to be addressed. “Examples of some of the major questions are related to the ownership and protection of personal data, data sharing and access, ethical use of data, and ethical implications of the use of learning analytics in education”, observes the editors of Journal of Learning Analytics, which featured a special issue on ethics and privacy in 2016 [9]. In a guest editorial [10] Ferguson et al. examined the learning analytics challenges with ethical dimension identified within this special issue. They found 21 challenges, of which six related to the *duty to act*; one addressed *informed consent*; three concerned *safeguarding individuals' interests and rights*; two were about *equal access to education* and a *just society*; seven dealt with *data protection*; and the last two were related to the *privacy* as socio-cultural concept.

A review of recent literature on ethics and privacy confirms the identified gap between concerns and challenges and proposals for design to address these issues.

In a number of papers Hoel and Chen [11, 12, 13] have explored what technical solutions a privacy-driven design of LA might lead to. In [14] Hoel, Cho and Chen researched how privacy and data protection requirements would affect all processes of the LA cycle.

In analysing the design implications of the GDPR we will use the same process model (Figure 1) as in [14]. The research question is to do a first exploration of how GDPR will influence design of the different processes of LA.

3. GDPR – new regulations for the digital age

European Union legislation on data protection has been in place since 1995. The objectives and principles of this directive (95/46/EC) remain sound, according to the new GDPR [15]. The existing laws were, however, drafted before the advent of cloud computing, social networking sites, location-based services and smartphones, so there was a need to update the laws to "make sure people's right to personal data protection (…) remains effective in the digital age". The aim was "to give people more control over their personal data and make it easier to access it" [16].

These are key changes of the new regulations [15], as the European Commission explains it[2] from the point of view of the citizens:

Consent for processing data: When your consent is required, you must be asked to give it by means of a clear affirmative action. More transparency about how your data is handled, with easy-to-understand information, especially for children.

Easy access to your own data: Free and easy access to your personal data, making it easier for you to see what personal information about you is held by companies and public authorities, and making it easier for you to transfer your personal data between service providers (data portability).

Data breaches: Without undue delay you have the right to know when your data has been unrightfully accessed or hacked.

Right to be forgotten: If you no longer want your personal data to be processed, and there is no legitimate reason for an organisation to keep it, it must be removed from their system. Data controllers must prove that they need to keep the data rather than you having to prove that collecting your data is not necessary.

Data protection by design and Data protection by default: Data protection safeguards should be built into products and services from the earliest stage of

[1] The Learning Analytics & Knowledge conferences organised by the Society for Learning Analytics Research

[2] Factsheets, directives, and regulation of EU reform on Data Protection is found at http://ec.europa.eu/justice/data-protection/reform/index_en.htm

development; the default settings should be those that provide the most privacy. Companies will be obliged to inform you as clearly, understandably and transparently as possible about how your personal data will be used, so that you are in the best position to decide what data you share. This information may be provided in combination with easy to understand standardised icons.

The GDPR does not concern processing of anonymous data. However, real anonymisation (removing personally identifiable information where it is not needed) is hard to achieve, therefore techniques of pseudonymisation (replacing personally identifiable material with artificial identifiers), and encryption (encoding messages so only those authorised can read it) are often applied, and encouraged in the GDPR, to protect personal data. According to the EC this will encourage the use of "big data" analytics, which can done using anonymised or pseudonymised data [15]. However, pseudonymisation or encryption do not exclude any data processor to conform to the GDPR.

4. Requirements for LA systems

Figure 1 depicts the main processes of a LA system [14] as been specified in an ISO/IEC standard (ISO/IEC 20748:2016). All processes are affected by the GDPR, which foresees that the principle of data protection by design and by default will be effective tools to create technological and organisational solutions [16]. In the following we will go through the LA processes one by one to solicit requirements from the new GDPR for system design.

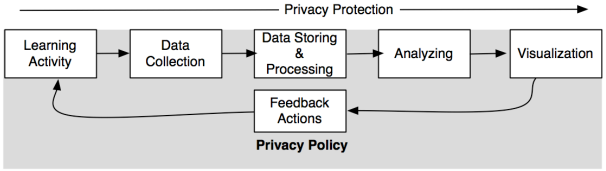

Figure 1. Learning Analytics processes [14].

Learning Activity: This is the process where the LA is set up, aims are decided, metrics for data collection defined, etc. GDPR requires that a dedicated conversation is set up between the LA processor and the student, parent, teacher or any other user of the system so that the data subject can "be informed of the existence of the processing operation and its purposes" [15, recital 60]. The information provided should take into account "the specific circumstances and context in which the personal data are processed. Furthermore, the data subject should be informed of the existence of profiling and the consequences of such profiling" [15, recital 60]. Given that context qualifies most things then pedagogical approach and the use of predictive models (profiling) should be part of this conversation. It is clear that this cannot be left to the machine to communicate alone; however, the outcome of such discussions must be documented by the system.

Data Collection: The process of gathering and measuring information on variables of interest is dependent upon consent by the learner. It is not enough to take the consent for granted when the learner enrolls in a course. GDPR requires "a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her. (..) Silence, pre-ticked boxes or inactivity should not (..) constitute consent [15, recital 32]. This requirement will lead to development of unprecedented software [12], as the consent should cover every purpose set out for the data collection.

Data Storing & Processing: This is the process of preparing and storing data from heterogeneous sources for transport and preparation to data analysis. In designing this process developers must develop mechanisms that allow the data subjects "to request and, if applicable, obtain, free of charge, in particular, access to and rectification or erasure of personal data" [15, recital 59]. In addition the end-users of the LA system have the right to object to information collected, and there need to be some provisioning of their exercise of this right. Again, there needs to be put in place log systems that will record actions taken for data protection audits. The GDPR enforces the principles of fair and transparent processing. The regulation introduces the idea of "standardised icons in order to give in an easily visible, intelligible and clearly legible manner, a meaningful overview of the intended processing. Where the icons are presented electronically, they should be machine-readable" [15, recital 60]. This provides requirements for a personal LA data monitoring tool, which gives the learner the opportunity to check at any time what data is gathered and stored about him or her, and to initiate actions. These actions could be rectification or erasure of personal data, or download their full personal dataset for storage in a personal data record store, or for transmission to another LA provider [15, recital 68].

The right to be forgotten gives the data subject the right to have his or her personal data erased and no longer processed "where the personal data are no longer necessary in relation to the purposes for which they are collected or otherwise processed, [or] where a data subject has withdrawn his or her consent or objects to the processing of personal data concerning him or her" [15, recital 65]. The system needs to communicate with other systems that have "links to, or copies or replications of those personal data" [15, recital 66]. However, this right is not absolute, and the system should be able to "handle conflicts", e.g., "temporarily moving the selected data to another processing system, making the selected personal

data unavailable to users, or temporarily removing published data from a website" [15, recital 67].

In this LA process steps should be taken towards pseudonymisation and/or encryption of personal data. The GDPR encourages use of measures of pseudonymisation [15, recital 29]. The solution needs to specify how the keys, the "additional information for attributing the personal data to a specific data subject" [15, recital 29], are managed either within the organisation running the LA or outside. There is also a need to set up procedures for risk evaluation to establish "whether data processing involve a risk or a high risk" [15, recital 76].

Analysing: The process of systematic examination of learning data in order to extract descriptive and possibly predictive knowledge about the learners and their contexts is the core of a LA system. This is a kind of profiling. GDPR defines profiling as "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements" [15, Article 4 (4)]. The models used for analysis are often impenetrable for the end-user; the GDPR, however, give clear incentives for the LA system developers to give "meaningful information about the logic involved" [15, Article 13, (f, g, h)]. It would be interesting to find out how far the learner's rights to information about algorithms and predictive models in LA system go with this new European law. The strong requirement gleaned from a design perspective is the need to plan for transparency and possibly some kind of open sharing of these technologies.

Visualisation: This 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. All the general requirements of the GDPR about transparent information, communication and modalities for the exercise of the rights of the data subject come into play here. However, we don't find that GDPR gives any LA specific requirement for this process.

Feedback Actions: Such 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 GDPR regulates that the learner or the teachers "should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing" [15, recital 71]. The automatic learning machine does not seem to be favoured by the GDPR; on the contrary, the regulations provide a strong incentive to design for human mediation in feeding back the results from LA. The design must give the learners a possibility to object to being targeted as a specific type of learner, if based predominantly on invisible statistical inferences. The principle of profile transparency gives the learners a right to be informed about how they are anticipated.

However, the GDPR goes further: If there is any automated individual decision-making, including profiling, the data subject have the right to be informed about "the significance and the envisaged consequences of such processing" [15, Article 14, g]. In order to establish some common ground for discussing envisaged consequences, the strict minimum period for which personal data could be stored [15, recital 39], specific circumstances and contexts in which personal data are processes [15, recital 60], etc. it is clear that the system design coming out of the GDPR requirements point in the direction of open systems with extensive conversational capabilities.

5. Discussion

Big data – the computational analysis of large dataset to reveal patterns, trends, and associations – can be used for many good purposes, and education is one of them. Big data entails a new way of looking at data, where data are assigned value in itself. The value of the data lies in its potential future uses. Big Data's business model is the antithesis of data minimisation and purpose limitation, which are key principles of privacy protection [16]. Previously, the data protections risks could be made to go away, almost by definition, by applying anonymisation of personal data. With big data new challenges arise with the risk of re-identification, which makes anonymisation less effective as a method for preventing the privacy disadvantages associated with profiling and other data analysis.

The GDPR is meant to strengthen data protecting while making it easier to run digital enterprises. For education it makes it necessary to go back and ask what the core values and ideas of teaching and learning as a business. It is not only about selling a course, but to build capabilities together with the students in a dialogue that is more complex than any consumer store transaction. This invites to more openness and transparency on what goes on with the use of learners' data for analytics. It also opens up for a dialogue with the learner that brings other principles of GDPR into play, e.g., the principle of proportionality, the principle of fair and transparent processing, etc. If data is collected for one purpose, it does not automatically prohibit processing for a different purpose or restrict raw data for use in analytics. "A key factor in deciding whether a new purpose is incompatible with the original purpose is whether it is fair. Fairness will consider factors such as; the effects on the privacy of individuals (e.g. specific and targeted decisions about identified persons) and whether an individual has a reasonable expectation that their personal data will be used in the new way" [17].

In the requirements this paper has solicited from GPDR (Section 4) a strong message comes through, and that is the need to involve the end-users of the systems in a dialogue that encompasses all the different processes of LA systems. This is not only a question of design of the LA system as such, but also concerning the execution of each individual process. It will not be possible to enter the discussion about the use of LA system only at enrollment and agree upon conditions of use by presented a list of check boxes. The negotiation about use of personal data and their analysis has to be a continuous process that will need tools that still have to be developed. If these tools are provided, there should not be limits to what could be achieved by LA. As the EC explains "companies are free to base processing on a contract, on a law or, on, in the absence of other bases, on a "balancing of interests". These 'formal requirements', such as consent, are set out in the rules to provide the necessary control by individuals over their personal data and to provide legal certainty for everyone" [17].

6. Conclusions and further work

This paper has explored what requirements for LA design come out of the new legislation passed in the European Union regarding protection of personal data and free movement of such data. Even before the GDPR was known there has been a considerable uncertainty to whether emergent LA practices were according to national and European data protection laws. GDPR maintain the principles from the 1995 directive, but clarifies a number of requirements that have to be built into LA architectures and systems now being designed. Openness, transparency and continuous negotiation between data subjects and data processors (i.e., school owners, universities, and vendors) are the principles the authors of this paper would highlight as the take away for further research and development.

The paper is a first exploration of how new European legislation will impact LA system design. Further research is now necessary to identify how other jurisdictions might deliver requirements for privacy and data protection in LA systems, and how the global market will be influenced by this European legislation.

7. References

[1] Griffiths, D., Drachsler,H., Kickmeier-Rust, M., Steiner, C., Hoel, T., Greller, W. (2016). Is Privacy a Show-stopper for Learning Analytics? A Review of Current Issues and Solutions. Learning Analytics Review 6. Published by the LACE project. ISSN:2057-7494 http://www.laceproject.eu/learning-analytics-review/privacy-show-stopper

[2] Dahl, M. (no date). Læringsanalyse. Memorandum published at the Norwegian Centre for ICT in Education website http://iktsenteret.no/ressurser/notat-laeringsanalyse. Accessed 2016-06-01

[3] Datatilsynet (n/d) Personal Data Act. Act of 14 April 2000 No. 31 relating to the processing of personal data (Personal Data Act) Online at https://www.datatilsynet.no/English/Regulations/Personal-Data-Act-/ Accessed: 2016-06-01

[4] European Commission (2016). How does the data protection reform strengthen citizens' rights? Online http://ec.europa.eu/justice/data-protection/document/factsheets_2016/factsheet_dp_reform_citizens_rights_2016_en.pdf

[5] Special Eurobarometer 431, Data protection: http://ec.europa.eu/public_opinion/archives/eb_special_439_420_en.htm#431.

[6] Drachsler, H., & Greller, W. (2016). Privacy and Learning Analytics – it's a DELICATE issue. LAK '16, April 20-24, 2016, Edinburgh, UK, 1 ACM 978-1-4503-3417-4/15/03.

[7] Mason, J., Chen, W., & Hoel, T. (2016). Questions as data: illuminating the potential of learning analytics through questioning an emergent field. Research and Practice in Technology Enhanced Learning, 1–14. http://doi.org/10.1186/s41039-016-0037-1

[8] Hoel, T., Mason, J. & Chen, W. (2015) Data Sharing for Learning Analytics – Questioning the Risks and Benefits. In Ogata, H. et al. (Eds.) (2015). Proceedings of the 23rd International Conference on Computers in Education. China: Asia-Pacific Society for Computers in Education

[9] Gasevic, D., Dawson, S., & Jovanovic, J. (2016). Ethics and privacy as enablers of learning analytics. Journal of Learning Analytics, 3(1), 1–4. http://doi.org/10.18608/jla.2016.31.1

[10] Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Guest editorial: Ethics and privacy in learning analytics. Journal of Learning Analytics, 3(1), 5–15. http://doi.org/10.18608/jla.2016.31.2

[11] Hoel, T. & Chen, W. (2015). Privacy in Learning Analytics – Implications for System Architecture — In Watanabe, T. and Seta, K. (Eds.) (2015). Proceedings of the 11th International Conference on Knowledge Management. ISBN 978-4-9908620-0-8 Presented at ICKM 15 in Osaka, Japan, 4 - 6 November 2015 ickm.kis.osakafu-u.ac.jp

[12] Hoel, T., & Chen, W. (2016). Privacy-driven design of learning analytics applications: Exploring the design space of solutions for data sharing and interoperability. Journal of Learning Analytics, 139–158. http://doi.org/10.18608/jla.2016.31.9

[13] Hoel, T., & Chen, W. (2014). Learning Analytics Interoperability– looking for Low-Hanging Fruits. In Liu, C.-C. et al. (Eds.) Proceedings of the 22nd International Conference on Computers in Education. Japan Asia-Pacific Society for Computers in Education.

[14] Hoel, T., Chen, W., & Cho, Yong-Sang (2016) Privacy Requirements for Learning Analytics – from Policies to Technical Solutions. Paper presented at Workshop on Ethics and Privacy for Learning Analytics, Monday, April 25th 2016 at the 6th International Conference on Learning Analytics and Knowledge (LAK '16), Edinburgh, United Kingdom

[15] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

[16] Datatilsynet. (2013). Big Data - privacy principles under pressure. https://www.datatilsynet.no/globalassets/global/04_planer_rapporter/big-data-engelsk-web.pdf

[17] European Commission (2015). Questions and Answers - Data protection reform. Online at http://europa.eu/rapid/press-release_MEMO-15-6385_en.pdf. Accessed 2016-06-01