Jens Doveren, Birte Heinemann and Ulrik Schroeder

Towards Guidelines for Data Protection and Privacy in Learning Analytics Implementation

Abstract

While leveraging data collected from learners to improve teaching and learning outcomes has an inherently desirable end goal, Learning Analytics researchers have to be aware of data protection policies and the justified desire for privacy while learning when rolling out such data collection efforts. Successful implementation requires knowing legal frameworks, coordinating with the personnel responsible at the individual institution, and clearly and openly communicating the extent and goal of the data collection effort to the learners and teachers.

In this paper, we present existing community guidelines and our own experiences from a rollout of Learning Analytics in the DigiLab4U project.

Keywords

Learning Analytics, Data Protection, Privacy

1 Introduction

The field of Learning Analytics seeks to leverage quantitative data about learning processes to improve teaching efficacy and learning outcomes. These improvements can derive from data directly, e.g., by presenting learners with insights into their own behaviour, or more indirectly, e.g., by informing decisions about how to improve future iterations of a course.

While such improvements to teaching are inherently desirable for students, gathering data about learning processes while they happen requires a certain degree of interference with the learner's privacy. Depending on the specific learning environment, the kind of data collection, as well as the mode of a collection, can vary widely, but examples include clickstream analysis in learning management systems or gaze and movement analysis in virtual reality applications. Some users might perceive such data collection as surveillance, which might in turn have an adverse effect on their willing-

ness to engage with the learning environment or might even lead to them refusing to engage at all.

Another complicating factor in data collection for Learning Analytics purposes, especially when working in an environment involving institutions from different jurisdictions, are different data protection policies and legal requirements. The intricacies of different data protection policies require communication and individual clearing with every institution that is involved.

Regardless of individual regulations, learners must be able to make an informed decision about whether they consent to the collection of their data. From this requirement follow two implications: one technical and one communicational. On the technical side, systems must be designed in a way that respects users' consent or the lack thereof, i.e., they must provide baseline functionality for users that have opted out of data collection, and they must be able to delete user data should consent to be withdrawn. On the communications side, challenges include finding ways to explain what pieces of data are collected and processed to potentially not particularly tech-savvy learners, as well as clearly communicating the potential benefits learners might reap from participation.

We aim to develop guidelines for researchers that intend to implement Learning Analytics data collection in real-world scenarios. These guidelines will be informed by previous work in that field, our own experience in rolling out Learning Analytics in various institutions of a multi-national research consortium, as well as a series of interviews with researchers and students about their expectations towards Learning Analytics, attitudes towards privacy and how they value the trade-off between the two.

2 Background

2.1 Data Protection Regulation

The European Union and its member states have long enacted regulations regarding the processing of personal data. As per (REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 2016), all handling of personal data must guarantee the following:

- Lawful processing
- Specified, explicit, and legitimate use
- Protection from secondary use
- Protection from inadequate and irrelevant processing

- Protection from the use of outdated information
- Protection from unnecessarily long data retention

In addition, the General Data Protection Regulation (GDPR) explicitly codifies the following people's privacy rights (What Is GDPR, the EU's New Data Protection Law?, 2020):

1. The right to be informed
2. The right of access
3. The right to rectification
4. The right to erasure
5. The right to restrict processing
6. The right to data portability
7. The right to object
8. Rights in relation to automated decision-making and profiling

The intricacies of how these principles influence the implementation of Learning Analytics are beyond the scope of this paper but are summarized in (Lukarov, 2019), as well as (Hoel et al., 2017). Underneath the overarching European regulation, there is an entire stack of more specific regulations and ordinances from a national down to an institutional level. Considering these highly heterogeneous regulations and institutional practices, the only general advice one could give to the aspiring Learning Analytics implementer is to communicate with the data protection officer responsible.

2.2 Community Experience

We are far from the first group to implement Learning Analytics in a real-world context, and the issue of privacy and data protection looms over all these efforts. After some rather unfortunate learning experiences, such as the failure of the Gates-funded inBloom (Singer, Natasha, 2014), where overzealous and ill-communicated collection of learning data from sensitive subjects led to a very public backlash, the Learning Analytics community has developed guidelines and checklists for effective communication and stakeholder involvement, such as the DELICATE checklist in (Drachsler & Greller, 2016) and (Schumacher & Ifenthaler, 2018).

The privacy implications of different technology stacks and processing methods are the subject of ongoing research in the Learning Analytics community, having led to the use of such elaborate methods as differential privacy (Steil et al., 2019). A survey of the available literature can be found in (Ciordas-Hertel et al., 2019).

Apart from careful technology choice, open and honest communication is the most important ingredient to a successful rollout of Learning Analytics. That entails explaining what is stored, for how long, and how it is being used, but also making sure learners see a tangible benefit from having their data analysed.

3 Own Contributions

3.1 Choice of Data Warehousing Solution

When gathering data from learners, the technical implementation of how that data is stored is of particular importance when considering data ownership and privacy. In order to be able to make any guarantees with regard to retention policies, deletion of data upon user request, and such, a thorough understanding of the selected data warehousing solution is required.

As it would be dishonest to promise users properties that we ourselves have no way of ensuring, we only considered self-hosted solutions that are free and open source. All hosting and maintenance of the data warehousing solution in the DigiLab4U project were done at RWTH Aachen University.

Although that might seem counterintuitive to those who have never implemented a software system with data persistence, architecting such a system in a way that ensures data integrity, prevents accidental loss of data, and enables arbitrary user data to be deleted at will is a surprisingly tricky endeavour. Hence, many implementations of xAPI learning record stores do not allow the true deletion of data, which we did not deem satisfactory.

These considerations and a thorough survey of the options available led to the choice of *Learning Locker* as the learning record store in the DigiLab4U project. A more detailed description of the decision-making process, as well as an overview of the privacy implications of many ready-to-use Learning Analytics data warehousing solutions on the market can be found in (Lukarov et al., 2020).

3.2 Stakeholder Survey

One of the lab environments enhanced with Learning Analytics as part of the DigiLab4U project is the RFID measuring chamber at HFT Stuttgart. In late 2021, the students of the bachelor program in Information Logistics were asked to take a survey containing questions of interest to various research groups in the DigiLab4U project. We were particularly interested in

students' pre-existing experience with Learning Analytics, the value they see in it, as well as their attitude toward sharing their data to enable Learning Analytics not only for them personally but also for fellow students.

Of the 41 participants, 34 filled in the survey completely. The results are hence to be treated more as anecdotal data but can nevertheless give us pointers on how to improve the rollout of Learning Analytics in future courses.

The two aspects relevant to the acceptance of data collection for Learning Analytics purposes in this context are the perceived value for the learners and whether it is great enough to overcome an inherent reluctance towards data sharing. In order to judge the degree to which attitudes toward Learning Analytics results might be tainted by a general lack of statistical literacy, we asked participants to specify their pre-existing knowledge in statistics in general and data visualization interpretation in particular.

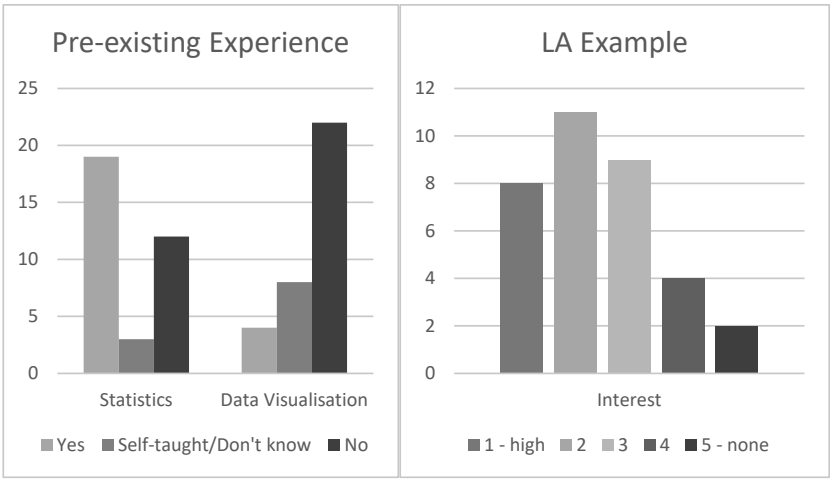

Figure 1 – Student Survey Results

As regards experience with statistics, of the 34 participants, 19 reported having taken a statistics course, 3 reported to be self-taught, and 12 reported no experience. In that same cluster of questions, 4 out of 34 students explicitly reported experience with interpreting data visualizations, 8 did not know, and 23 reported no experience.

Presented with a fictional example of data visualizations for learning feedback and asked whether they would find such feedback interesting, 8 respondents gave a 1 (very interested) on a five-step Likert scale, 11 gave a 2, 9 gave a 3, 4 gave a 4, and the remaining 2 gave a 5. When asked whether they would be willing to share their learning data to help in the generation of such feedback, only one participant gave an unconditionally positive response, 20 specified that they would require their data to be anonymized, and 13 did not give any response.

4 Closing Thoughts

The results of the learner survey suggest a certain degree of hesitation towards sharing their data. One possible cause may be a failure to see how they might personally benefit from feedback generated using Learning Analytics. Over 40% of participants reported only middling or no interest in the examples that were provided, which might explain the lack of enthusiasm.

These experiences once again underline the importance of early communication with learners as stakeholders. Only they can articulate their needs and the value that they attach to any given form of feedback, which in turn must be evaluated by teachers with respect to its didactical value.

Another factor that was beyond the scope of this paper is the issue of scaling up Learning Analytics infrastructures—moving from a smaller, more experimental rollout to a larger, institution-wide one often requires the use of different, industry-grade big data processing frameworks, which again come with their own privacy considerations. The bigger such a project grows, the higher the incentive for standardization, which on one hand facilitates collaboration and exchange of knowledge, but on the other hand, to a certain degree limits technological choices.

Acknowledgements

This work has been funded by the German Federal Ministry of Education and Research (BMBF) through the project DigiLab4U (No. 16DH2114).

References

Ciordas-Hertel, G.-P., Schneider, J., Ternier, S., & Drachsler, H. (2019). Adopting Trust in Learning Analytics Infrastructure: A Structured Literature Review. Journal of Universal Computer Science, 25(13), 1668–1686. https://doi.org/10.3217/jucs-025-13-1668

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge — LAK '16*, 89–98. https://doi.org/10.1145/2883851.2883893

Hoel, T., Griffiths, D., & Chen, W. (2017). The Influence of Data Protection and Privacy Frameworks on the Design of Learning Analytics Systems. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 243–252. https://doi.org/10.1145/3027385.3027414

Lukarov, V. (2019). *Scaling up learning analytics in blended learning scenarios*.

Lukarov, V., Ehlenz, M., & Schroeder, U. (2020). Building a data warehouse for multimodal learning analytics research projects. *Companion Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 25–28.

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2016:119:FULL&from=EN

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, *78*, 397–407. https://doi.org/10.1016/j.chb.2017.06.030

Singer, Natasha. (2014, April 21). InBloom Student Data Repository to Close. *The New York Times*. https://bits.blogs.nytimes.com/2014/04/21/inbloom-student-data-repository-to-close/?_r=0

Steil, J., Hagestedt, I., Huang, M. X., & Bulling, A. (2019). Privacy-aware eye tracking using differential privacy. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, 1–9. https://doi.org/10.1145/3314111.3319915

What is GDPR, the EU's new data protection law? (2020). https://gdpr.eu/what-is-gdpr/

Authors

Jens Doveren
RWTH Aachen University
Learning Technologies Research Group
Ahornstr. 55
52074 Aachen, Germany
https://elearn.rwth-aachen.de/doveren
doveren@cs.rwth-aachen.de

Birte Heinemann
RWTH Aachen University
Learning Technologies Research Group
Ahornstr. 55, 52074 Aachen, Germany
https://elearn.rwth-aachen.de/heinemann
heinemann@cs.rwth-aachen.de

Prof. Dr.-Ing Ulrik Schroeder
RWTH Aachen University
Learning Technologies Research Group
Ahornstr. 55
52074 Aachen, Germany
https://elearn.rwth-aachen.de/schroeder
schroeder@cs.rwth-aachen.de