

A Teacher Survey on Educational Data Management Practices: Tracking and Storage of Activity Traces

Juan Carlos Farah¹, Andrii Vozniuk¹, María Jesús Rodríguez-Triana^{1,2}, Denis Gillet¹

¹ École Polytechnique Fédérale de Lausanne, Switzerland

{juancarlos.farah, andrii.vozniuk, maria.rodriigueztriana, denis.gillet}@epfl.ch

² Tallinn University, Estonia

mjrt@tlu.ee

Abstract When students participate in computer-mediated learning, their activities are often recorded for learning analytics and educational data mining purposes. While handling student data has associated privacy concerns and is often subject to legal regulations, there is a limited understanding of how educational data is currently managed in practice. To clarify this question, we conducted a survey of over 100 teachers, examining their experience with storing and analyzing student data. The survey identified the wide variety of platforms used to track and store student activity traces, highlighting the necessity to develop common data exchange standards and enforce data management policies that are consistent across platforms. In addition, the responses also revealed a mismatch between platform affordances regarding data tracking and storage, and teacher awareness of such affordances. This disconnect could be mitigated by reinforcing the transparency policies and usability of educational platforms, as well as by improving teacher data management literacy.

Keywords: student data, activity traces, learning analytics, educational data mining, data management, privacy, teacher survey

1 Introduction

As educational institutions continue to adopt digital education solutions, the volume of educational data recorded on digital platforms continues to rise [11]. The availability of this data, along with improvements in computational capacity, have led to an increase in the resources available to perform learning analytics (LA), and educational data mining (EDM), also motivating frameworks that address how student data should be collected, stored and later accessed [14]. The importance of addressing ethical and privacy issues within the context of digital education, LA and EDM is illustrated by projects such as Sheila³, LEA [15], and LACE [6].

Several studies have targeted the educational data privacy issue from multiple angles. Proper data management is one of the key dimensions of LA associated with corresponding privacy and ethics risks, as highlighted in [7]. These risks also create barriers for adoption of digital technologies in education. To promote trust in LA solutions, Drachler and Greller [3] proposed a list of requirements for LA implementations,

³ Sheila Project: <http://sheilaproject.eu/>

including educational data management as one of the aspects to be taken into consideration. Moreover, within a more general framework, the European General Data Protection Regulation (EU GDPR 2016/679) [5] defines a number of requirements regarding the collection and processing of personal data to be met starting May 2018.

The role of the teacher when dealing with ethical and privacy matters has been highlighted in several studies [4,12], and approaches towards putting the teacher in control of student privacy have been proposed [17]. Teachers are of particular importance when considering student privacy, as they are often the ones selecting appropriate technology to support their practice and play the leading role in conducting technology-mediated learning activities [12]. Furthermore, teachers are often the primary target audience for LA tools [4,13]. Hence, understanding how teachers use digital technologies for educational purposes is of particular importance.

Despite the existing regulations and the research interest in ethics and privacy at the teacher and classroom levels, to the best of our knowledge, there is limited information regarding the data management approaches followed in the field. Addressing this gap, we conducted a survey of 106 teachers aimed at shedding light on *the current state of educational data management in practice*. More concretely, the survey explores which digital platforms are used, what data is stored on these platforms, where and for what reasons it is stored, and how aware teachers are of how these platforms handle student data. Our hypothesis is that the teachers have limited control over the data retention policy or storage location employed by educational platforms, hence bringing forth a discussion regarding the adequacy of current data management architectures.

2 Methodology

To collect responses from teachers, we conducted an online survey⁴ in English. The responses to the survey were anonymous unless the respondents explicitly provided their email, which was optional. The survey was designed to be completed in under five minutes to increase the number of potential respondents. The first two parts of the survey focused on data collection and storage. Respondents were asked questions such as whether or not their students conducted online or computer-based activities and if so, how and where data generated from these activities was tracked, collected and stored. A third part addressed the use of this data for LA, covering questions such as whether or not respondents could perform analyses on student-generated data as well as the motivation for the nature of the metrics reported, and who had access to the results. We distributed the survey between May and June 2017 to teachers at focus groups within the framework of the Next-Lab project⁵ and through online message boards, social networks and mailing lists targeted at teachers in Europe, North and South America.

3 Results

In this section we present our results, focusing on the data collection and storage part of our survey. Following our methodology, we obtained 106 responses: 55 female, 47

⁴ Online Survey: <https://goo.gl/R5Z3LL>

⁵ Next-Lab Project: <http://project.golabz.eu/>

male, and 4 N/A. Teaching experience ranged from 1 to 43 years (median 12). The age of participants ranged from 23 to 64 years (median 39). Three participants indicated teaching only at the preschool level, 12 only at the primary school level, 33 only at the secondary level, 38 only at the higher education level, 11 at more than one of these levels, and 9 N/A.

Use of online or computer-based activities. A total of 95 out of the 106 respondents reported that their students performed online or computer-based learning activities. These activities reportedly took place on 63 different platforms, with single respondents specifying up to 8 platforms (median 2). The most cited platform was Google Search (56 respondents), followed by Moodle (27 respondents), Google Classroom (26 respondents) and Wikipedia (13 respondents). Out of the 4 respondents who were not sure whether or not their students partook in these learning activities, 3 reported using platforms such as Google Search, Wikipedia, and Moodle, and 2 actually stated both that these learning activities were tracked and that student data was stored after the activities took place.

Tracking and storing activity traces. As depicted in Figure 1, while 73 respondents noted that data related to learning activities was tracked, 74 indicated that student data was stored after the activities took place. Eight respondents were not sure if student data was tracked, while 11 were not sure if student data was stored after the activity was over. We identified (*Group A*) 20 respondents who either reported that student activity was not tracked (14 respondents) or were not sure if student activity was tracked (6 respondents), but at the same time indicated use of platforms such as Google Search (10 respondents), Moodle (5 respondents), Google Classroom (4 respondents) and Wikipedia (3 respondents). Similarly, we identified (*Group B*) 21 respondents who either indicated that student data was not stored (11 respondents) or were not sure if student data was stored (10 respondents), but at the same time reported use of platforms such as Google Search (9 respondents), Wikipedia (6 respondents), Google Classroom (5 respondents) and Moodle (5 respondents). Twelve respondents were both in *Group A* and *Group B*.

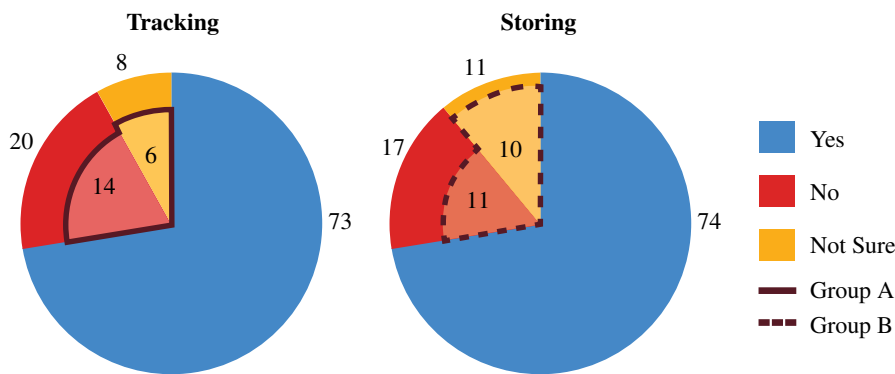


Figure 1: Reported tracking (left) and storing (right) of student activity traces. *Group A* and *Group B* represent teachers who evidenced possible contradictions in their responses. Teachers in those groups reported no tracking/storage or not being sure of tracking/storage, but asserted using platforms with their students that normally do track and/or store data.

Reasons for storing student data. Data was reportedly stored mainly because it was done automatically by the platform (62 respondents), with 17 respondents indicating this as the only reason they store data and the other 45 pairing it with one or more other reasons. Other reasons highlighted included storing because the teachers themselves conduct analysis on the data (52 respondents), because their institution asks for reports on the data (16 respondents) or because their institution performs analysis on the data (13 respondents).

Location where student data is stored. Student data was primarily reported to be stored on the platform hosting the activity (67 respondents). A total of 14 respondents confirmed that they were able to specify in the settings of at least one of the platforms they used, where exactly they wanted student data to be stored, while 38 replied that this was not possible and 32 were not sure. Of the 14 respondents who indicated they could configure the storage location, 9 confirmed that they had actually used this functionality.

4 Discussion

As stated in the introduction, our hypothesis was that teachers had limited control over the retention policy or storage location of data generated by students when performing online or computer-based activities. To explore our hypothesis, we analyze (1) the nature of the **infrastructure** on which these activities take place, and (2) the **contradictions** that arise in the responses. Based on our analysis we infer a possible disconnect between the respondents' *understanding* of how the platforms they use handle student data, and how these platforms *actually* handle student data. In this section, we present how this inference emerges from the survey results.

Infrastructure. Our results demonstrate that students of 95 of the 106 teachers surveyed perform computer-mediated activities. Respondents indicated a plethora of platforms where student activities take place. Given that several of these platforms are not learning management systems (LMSs), but open websites such as Google Search and Wikipedia, it is clear that we are in the *beyond the LMS* era [9]. A vast majority of respondents also indicated that student data generated during these activities is tracked and stored predominantly in the platform on which it took place, suggesting that student data could mostly end up siloed in disconnected repositories. This situation reinforces the need to develop common data exchange standards and data management policies consistent across platforms in order to achieve interoperability for LA [1,2,8,9,15].

Additionally, the fact that the majority of the respondents reported that data was stored because it was done automatically by the platform puts forth a number of questions related to the effect of automatic collection and storage. Namely, (1) how do the default settings affect educational data retention, (2) whether those respondents who cited automatic collection as a reason for storing data would still decide to store it if their explicit consent was required, and (3) whether the 17 respondents who *only* cited automatic collection as a reason for storing student data had the option to opt-out.

Contradictions. Results also show some areas where responses suggest confusion or are outright contradictory. A total of 29 respondents were included in *Group A* or *Group B* as defined in Section 3, meaning they evidenced confusion or contradiction in their answers regarding the storage and tracking of student activity traces. These re-

spondents reported no tracking/storage or not being sure of tracking/storage, but cited use of platforms that normally do track and/or store data, such as Google Classroom and Moodle. These discrepancies highlight a possible lack of transparency with regards to tracking and data storage policies from the part of the platforms cited. The aforementioned EU GDPR 2016/679 addresses the need for transparency when processing personal data, stating in Article 5 that “personal data shall be (a) processed lawfully, fairly and in a transparent manner” [5]. The discrepancies that emerged in some of the responses to our survey could suggest that this transparency is not apparent in some of the platforms used by respondents. Whether this is a design limitation or a possible lack of awareness from the part of those surveyed, it is clear that this is an issue that would need to be addressed to correctly implement EU GDPR 2016/679.

Furthermore, it is important to note that interoperability at the technical level needs to be accompanied by measures to ensure transparent privacy policies *across* platforms. This would allow teachers to benefit from cross-platform aggregation for comprehensive analyses of learning activities without risking data privacy violations due to transfer of data *between* tools. The need for such aggregation is well-justified by our results, given respondents reported using on average two platforms for performing computer-based activities with their students, and 32 teachers reported using 3 or more platforms.

5 Conclusion

In this study, we presented the results of a survey of 106 teachers, focusing on their experience with collecting and storing student activity traces. The results show the wide variety of platforms used for these purposes. To enable data integration at the technical level, employment of common educational data exchange standards for platform interoperability is required. At the same time, analysis of such data requires that collection and retention policies employed by these platforms be in agreement and in compliance with respective regulations. Such alignment lacking, there is a possibility that differing data management and retention policies, as well as adherence to conflicting legal frameworks could complicate or prevent integration.

The number of reported contradictions can indicate the lack of transparency regarding data management policies employed by educational platform providers. In such cases, better explanations and more user-friendly cues for awareness of these policies should be integrated into the platforms. Nevertheless, there is a possibility that contradictions were also in part due to limited teacher proficiency at the technological level, including a lack of understanding of data management approaches and best practices. To foster the proper adoption of data management policies while using digital technologies for educational purposes, there is a need to reinforce teacher awareness and literacy with regards to the relevant privacy issues, as suggested by [10,16]. Given the teacher’s role when dealing with ethics and privacy in digital education [12], being proficient in how platforms handle tracking and storage of student activity traces is paramount.

Acknowledgements

This research has been partially funded by the European Union in the context of Next-Lab and CEITER projects (Horizon 2020 Research and Innovation Programme, grant agreements nos. 731685 and 669074).

References

1. Berg, A., Scheffel, M., Drachler, H., Ternier, S., Specht, M.: Dutch cooking with xAPI recipes the good, the bad, and the consistent. *Proceedings - IEEE 16th International Conference on Advanced Learning Technologies, ICALT 2016* pp. 234–236 (2016)
2. Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Thüs, H.: A reference model for learning analytics. *International Journal of Technology Enhanced Learning* 4(5/6), 318 (2012)
3. Drachler, H., Greller, W.: Privacy and analytics: It's a delicate issue a checklist for trusted learning analytics. In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. pp. 89–98. LAK '16, ACM, New York, NY, USA (2016)
4. Dyckhoff, A.L., Zielke, D., Bültmann, M., Chatti, M.A., Schroeder, U.: Design and implementation of a learning analytics toolkit for teachers. *Educational Technology & Society* 15(3), 58–76 (2012)
5. European Union: Regulation 2016/679 of the European parliament and the Council of the European Union. *Official Journal of the European Communities* 2014(April), 1–88 (2016)
6. Ferguson, R., Hoel, T., Scheffel, M., Drachler, H.: Guest editorial: Ethics and privacy in learning analytics. *Journal of learning analytics* 3(1), 5–15 (2016)
7. Greller, W., Drachler, H.: Translating learning into numbers: A generic framework for learning analytics. *Educational technology & society* 15(3), 42–57 (2012)
8. Hoel, T., Griffiths, D., Chen, W.: The influence of data protection and privacy frameworks on the design of learning analytics systems. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17* pp. 243–252 (2017)
9. Kitto, K., Cross, S., Waters, Z., Lupton, M.: Learning analytics beyond the lms: The connected learning analytics toolkit. In: *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*. pp. 11–15. LAK '15, ACM, New York, NY, USA (2015)
10. Nathan, L.P., MacGougan, A., Shaffer, E.: If Not Us, Who? Social Media Policy and the Ischool Classroom. *Journal of Education for Library and Information Science* 55(2), 112–132 (2014)
11. Pardo, A., Siemens, G.: Ethical and privacy principles for learning analytics. *British Journal of Educational Technology* 45(3), 438–450 (2014)
12. Rodríguez-Triana, M.J., Martínez-Monés, A., Villagrà-Sobrino, S.: Learning analytics in small-scale teacher-led innovations: ethical and data privacy issues. *Journal of Learning Analytics* 3(1), 43–65 (2016)
13. Schwendimann, B., Rodríguez-Triana, M., Vozniuk, A., Prieto, L., Boroujeni, M., Holzer, A., Gillet, D., Dillenbourg, P.: Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies* 10(1), 30–41 (2017)
14. Slade, S., Prinsloo, P.: Learning Analytics: Ethical Issues and Dilemmas. *American Behavioral Scientist* 57(10), 1510–1529 (2013)
15. Steiner, C.M., Kickmeier-Rust, M.D., Albert, D.: Lea in private: a privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics* 3(1), 66–90 (2016)
16. Tsai, Y.S., Gasevic, D.: Learning analytics in higher education — challenges and policies. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17* pp. 233–242 (2017)
17. Vozniuk, A., Govaerts, S., Bollen, L., Manske, S., Hecking, T., Gillet, D.: AngeLA: Putting the Teacher in Control of Student Privacy in the Online Classroom. In: *Information Technology Based Higher Education and Training (ITHET)* (2014)