Privacy of Crowdsourcing Educational Platforms in the Light of New EU Regulations

Katerina Zdravkova
University Ss. Cyril and Methodius, Faculty of Computer Science and Engineering
Rudjer Boshkovikj 16, 1000 Skopje, Macedonia
katerina.zdravkova@finki.ukim.mk

Abstract

Many crowdsourcing systems enable an anonymous access and opportunity to namelessly contribute self-generated content without providing any personal data. However, Internet browsers collect metadata on a large scale, including learning management systems (LMS), which collect and store many identity and contact data. System administrators and the teachers responsible for the courses can access them at any time. Interactive activities embedded in the LMS can reveal sensitive data, such as religious beliefs, political views, health, sexual orientation, race, or membership to organizations. They are visible to all the enrolled students. Educational organizations who are hosting LMS, also collect a lot of data that is usually transferred to third countries, but also transmitted to third parties, including university researchers or outside companies, often even governments. This paper examines the challenges of a prospective crowdsourcing platform intended for education, which must be taken into consideration by design. It presents examples of violated privacy in education, the student protection regulations, and the privacy concerns of learning management systems. The compliance of the most popular LMSs, MOOCs and crowdsourcing systems with GDPR are examined and compared. The paper concludes with the privacy policy guidelines of the prospective crowdsourcing educational platform in the light of GDPR.

Keywords: Crowdsourcing, GDPR, LMS, rights of data subject

1. Introduction

Many crowdsourcing systems enable an anonymous access and opportunity to namelessly contribute self-generated content without providing any personal data (Halder, 2014). However, Internet browsers, which support the functioning of crowdsourcing platforms, collect metadata on a large scale (Soltani and Seno, 2014). Digital traces include: users' IP address, their exact location, time zone and language, the type of the used device (PC, laptop, tablet, mobile), hardware features (CPU, graphics cards, RAM specifications), the operating system, the screen resolution, the battery level, the moment and the duration of accessing the browser, as well as the installed browser plugins. These facts generate a browser fingerprint, which is a very accurate method to identify unique browsers and track online activities (Eckersley, 2010). Moreover, servers send HTTP cookies to user's browser, such as the authentication ones, user preferences and settings, which are stored on the user's computer. Since data collection and cookie depositing are almost unavoidable, and permitted according to most privacy protection laws, crowdsourcing can be considered privacy safeguarded per se.

New learning management systems collect and store a lot of identity and contact data, such as: student ID, name, e-mail, picture, in addition to a list of server logs, all activities undertaken, their duration, grades of learning assignments, and the browser type and language (Flanagan and Ogata, 2017). System administrators and all the teachers responsible for the course can access them at any time.

Educational organizations who are hosting LMSs collect additional identifiable data. Student records are sometimes extensive and completely incompatible to modern laws, which tend to minimize the amount of personally identifiable information. Moreover, the collected data are usually transmitted to third parties via government agencies, mainly to education researchers (Joiner, 2018).

Interactive activities embedded in the LMS, such as the wikis, discussion forums and blogs are always associated with the name and the picture of the content provider, which can be either a teacher or another student enrolled into the same course (Poore, 2015). When students equate their performance within the interactive educational system with their behaviour in the social media, they can accidentally reveal some sensitive data, like their religious and political views, health status, sexual orientation, race, and membership to organizations, or intentionally impose their dogmas, enforced decisions, or beliefs (Zdravkova, 2016). Once posted, this information could remain visible to all the course participants. These issues are a further privacy threat that is usually not protected at all so far (Drummond and Fischhoff, 2017).

In 2016, EU approved the General Data Protection Regulation (GDPR), which was enforced in May 2018 (European Commission, 2018). It enhances the regulation responsible for personally identifiable information, processing and free movement. GDPR's main purpose is "to enhance data protection rights of individuals and to improve business opportunities by facilitating the free flow of personal data in the digital single market". It harmonized the protection of "fundamental rights and freedoms", in the context of technological developments, globalization, increasing scale of data collection and sharing, regarding the necessity of free flow of personal data, not only within EU, but also towards third countries.

Educational crowdsourcing systems are a symbiosis of both. For educational purposes, most of the previously mentioned data and metadata should inevitably be collected. Responsible platforms should enable their processing, accessing, sharing and transfer to third parties and countries obeying precisely the privacy protection principles. New EU regulations affect the creation of privacy policies of educational crowdsourcing.

This paper examines the challenges of a crowdsourcing platform intended for education, which should be taken into consideration prior to its launching. It continues with examples of violated privacy in education, privacy concerns of learning management systems, and student protection regulations. In section 3, the compliance of the most popular LMSs, MOOCs and crowdsourcing systems with GDPR, is examined and compared. Section 4 is dedicated to enetCollect's affiliated organisations EURAC and ILIAS. The paper concludes with the privacy policy guidelines of a prospective crowdsourcing platform.

2. Privacy in education

One of the major imperatives of European higher education area (EHEA) is student-centred learning, which promotes supportive and inspiring learning environment based on innovative teaching methods, pedagogical innovation and digital technologies (Bergan and Deca, 2018). The effectiveness of digitally supported education highly depends on the well-established privacy protection (Zeide and Nissenbaum, 2018). Privacy concerns additionally grow due to the emergence of the MOOCs over the existing online learning management systems (Sandeen, 2013). They enable universal access, which amplifies their disruptive nature (Jones and Regner, 2016). The involvement of many non-educational institutions in the MOOCs additionally aggravates the intention to establish strict privacy policy regulations. The following subsections observe three aspects: examples of violated privacy, general privacy concerns of learning management systems, and the privacy protection regulations applied to education.

2.1 Violated privacy in education

Suzanne Widup's (2010) exhaustive report revealed that from 2005 to 2009, more than 2 800 data breach incidents occurred, 549 of them in educational organizations. The amount of breached records exceeded 10 million (Widup, 2010). According to this report, one of the crucial reasons for such a high occurrence of data violations in education was the absence of monitoring systems that might prevent the malicious use of student data. Another report has recently proved that larger universities, universities with more financial resources, and universities with weak privacy policies were more susceptible to data breaches (Mello, 2018).

DLA Piper study reports almost 60 000 data breaches in Europe after the introduction of GDPR, more than one sixth in UK (DLA Piper, 2019). Most notifications were spotted among private and public organisations from the Netherlands, Germany and UK. Even though the report doesn't highlight the type of the organisation, it is very realistic that at least 10 000 belong to educational establishments.

2.2 Privacy concerns of learning environments

Academic analytics became an inevitable and a very reliable tool for assessment and auditing of education (Campbell, DeBlois and Oblinger, 2007). It is usually combined with educational data mining "providing useful insights into student behavior online" (Baepler and Murdoch, 2010). The process of gathering, analysing, and presenting student data is usually performed within learning management systems. Student data have nowadays expanded to big data (Chen, Mao and Liu, 2014; Godwin-Jones, 2017). Their huge volume makes them a fruitful arena for rich data analysis, which increases the possibility of uncontrolled data mining and significantly reduces privacy (Johnson, 2014).

An additional problem is the redirection of the traditional eLearning methods towards cloud services, where privacy and security issues are a real challenge (Sen, 2015). However, the greatest privacy challenge for the learners and their teachers is the opportunity to generate interactive content, where all the uploaded information is visible to all other participants of the course, and the authorship is associated to its creator. Even when the content is erased, the traces of its existence remain permanent.

2.3 Student protection regulations

Most LMSs, MOOCs and crowdsourcing projects are hosted in the US, and are used massively outside of them, which led to the necessity to establish a reasonable framework, in order to avoid some prospective international conflicts. In spite of many regulations, such as: FERPA, PPRA, IDEA and COPPA there is not a single comprehensive federal U.S. law regulating the collection and use of personal data (https://www.usa.gov/privacy)[1]. To handle the problem, mutual EU-US and Swiss-US privacy agreements have been established. They regulate data privacy, safety and security, as well as cross-border data transfers. The two frameworks are standardised for all other European National Privacy regulations, so if one organization is compliant with GDPR, it is very probable that it also fulfils the national regulations.

3. Compliance with GDPR

The new EU privacy protecting regulations contain 99 articles divided into 11 chapters (EC, 2018). For the prospective crowd-oriented learning system, it is essential to study the "rights of data subject", where "data subject" is any "identified or identifiable natural person" (chapter 3), and the "transfers of personal data to third countries or international organisations" (chapter 5). Article 85, which deals with the "processing and freedom of expression and information", might also be decisive for enetCollect. If the rights of data subject, and the cross border data transfers are not carefully established, all the "remedies, liability and penalties" from chapter 8 will be implemented. They can be gigantic, like the fine of 50 million EUR, which was imposed on Google by French data protection watchdog (DLA Paper, 2019).

The basic rights of data subject of the most popular LMSs, MOOCs and crowdsourcing systems are presented in Table 1, which appear at the end of the paper. GDPR rights are clustered into five sections: transparency and modalities, information and access to personal data, rectification and erasure, right to object and automated individual decision-making, and restrictions (EC, 2018). The compliance of the educational systems with them is judged according to their privacy notes and terms of use. The defined criteria for each are presented in the following five paragraphs.

The compliance with the transparency and modalities legal items among other, means that the existence of an appointed controller; provided written or oral information related to data processing; provided information related to data transfers to a third country or to an international organisation; controller's duty to protect data processing; protection of data subject from any legal effects based solely on automated processing; and implementation of suitable measures to safeguard the data subject's rights and freedom, and legitimate interests.

[1] All the online resources, privacy policies and terms of use were last retrieved on 10th April 2019.

Information and access to personal data refer to: the purpose of data collection; contact details of the controller; the recipients of collected data; the period of storing the data; the right to access the data; the right to demand an erasure of personal data; the right to restrict processing; detailed information of accessing data; and direct access to collected data.

Rectification and erasure clauses imply that the data subject has the right to: demand a rectification of inaccurate personal data; right to erasure ('right to be forgotten'); right to restriction of processing; notification that any of the three later actions have been performed; and the right to receive the personal data.

The right to object and the automated individual decision-making, are comprised of the rights to object data processing at any time; and the rights to object data processing for direct marketing purposes.

Restriction refers to a limited scope of obligations in special circumstances related to the fundamental rights and freedoms; and safeguarding of democratic society.

Blackboard is one of the leading LMSs, and as said by them, #1 Global Education Software Provider. With more than 100 million users, Blackboard must guarantee the best conditions, including privacy. Blackboard has a very strict and detailed privacy, which is EU-U.S. Privacy Shield certified. The compliance with GDPR is presented in the 21 pages long GDPR White Paper.

Canvas is Instructure's LMS with more than 18 million users (instructure.com), intended for K-12 and university students. In parallel with the privacy policy, Canvas has extensions for the residents of the EU and Switzerland. Canvas is also dedicated to adapting their own privacy policy to GDPR. They are self-certified under the EU-U.S. Privacy Shield. Recently, there were complaints about data treatment and third parties (privacy.commonsense.org/ evaluation/canvas).

With more than 300 million users and "world's largest collection of language-learning data", Duolingo is the biggest educational community dedicated to language learning, which presents completely crowdsourced language courses ai.duolingo.com/). It has the most comprehensive privacy policy, which carefully covers all the privacy, safety and security rights of data subject, (duolingo.com/privacy). In spite of the declared readiness to protect users' data, the application is criticized for "third-party advertising or tracking services" (privacy.common sense.org/evaluation/duolingo).

Intended for K-12, Edmodo is another example of a learning management system with detailed privacy policy (go.edmodo.com/privacy-policy/) and terms of service. These regulations are not fully compatible with GDPR, but still offer significant rights to data subjects. In May 2017, Edmodo suffered a severe data breach, which affected 77 million users (EHL, 2017).

EdX is an open-source platform and MOOC provider with more than 130 partners and 18 million users. They claim: "edX is making a good faith effort to comply, given our global reach with learners and partners." The privacy policy proves it (edx.org/edx-privacy-policy).

FutureLearn is a digital educational platform "wholly owned by The Open University" (future learn.com/about-futurelearn). Highly experienced OU prepared a very concise and fully GDPR compliant privacy policy (about.futurelearn.com/terms/privacy-policy).

Khan Academy is a global multilingual classroom for millions of users. Their privacy policy is carefully prepared, and it includes special clauses for European users only (khanacademy.org/about/privacy-policy).

Mechanical Turk's privacy notice redirects towards Amazon, whose privacy has not been recently updated, (mturk.com/privacy-notice), thus it is hardly compliant with GDPR. It might be crucial for their unethical acting while harvesting Facebook profiles and manipulating people (EFF, 2018).

Moodle is the most popular open source LMS with almost 150 million registered users (moodle.net/stats/) who are striving for the highest ethical standards. MoodleDocs privacy rights are compatible with GDPR at all points. But, this January, Moodle experienced an outage (Greidanos, 2019). Unlike Edmodo, it suffered from lack of reliability.

SAP SuccessFactors is a cloud provider with 120 million users, whose cloud security and data privacy are carefully designed and maintained, providing complete compliance with privacy and security standards worldwide (www.suc cessfactors.com/content/ssf-site/en/about/privacy.html).

In parallel with the rights of data subjects, the compliance with the Article 85 of all the studied platforms was also examined. After a very exhaustive examinations of their corresponding policies, it was noticed that none mentions the freedom of expression and information. An exception is Moodle, which contains a word censorship filter, intended to disable the submission of "obscene or other unwanted words in the text" within forums and wikis (https://docs.moodle.org/36/en/Word_censorship_filter). It can be misused to restrict the free expression, because the censor.php file can be tailored to disable some word strings.

Most observed educational and crowdsourcing systems have shown a very high social responsibility and a serious concern about privacy rights of their users. Unfortunately, the abuse of users' confidence has occurred in both observed crowdsourcing systems.

4. EnetCollect and new EU regulations

The major motivation of this study was to discover the deficiencies of the related educational platforms in order to avoid them carefully while creating the enetCollect's crowd-oriented language learning system. It was concluded that declaratively, all of them respect the rights of data subject and pay attention to information security. Well established policies and terms of use converge to some general rules and recommendations, which should be taken into consideration for the prospective platform.

It is very probable that the selection of the platform provider will be done among the two technically most engaged partners of the action: EURAC or ILIAS. Namely, the official presentation of enetCollect is hosted by EURAC (http://enetcollect.eurac.edu/), while the intranet website is available from ILIAS (https://enetcollect.net/). How much are they compliant to new EU regulations?

EURAC research has a privacy policy which has been recently adjusted according to EU Regulation 2016/679 (eurac.edu/en/aboutus/Pages/Privacy.aspx). However, it warns the users about the use of Google Analytics, without an immediate possibility to "decline the use of cookies". Furthermore, the website "may use the third-party cookies" including some social plugins. With these official announcements, EURAC research disclaims responsibility for any privacy violation.

Although ILIAS is a multi-language open-source LMS, their privacy policy, or more precisely, the terms of service are presented in German only (docu.ilias.de/ilias.php?cmd=showTermsOfService&cmdClass=ilstartupgui&cmdNode=k8&baseClass=ilStartUpGUI). The policy starts with the intellectual property rights under GPL, carries on with the limitations of inappropriate content, and continues with data protection. The compliance with GDPR is not explicitly highlighted, but all the rights of data subject are carefully examined. The possibility of using the LMS by people with blindness or visual impairments, which is guaranteed by the Marrakesh Treaty in not enabled (www.wipo.int/marrakesh_treaty/en/). This is the only system, which reveals the responsible authority for all the data protection issues (http://www.ldi.nrw.de).

5. Conclusion

EnetCollect's crowdsourcing framework for language learning can initially adopt EURAC's prudent privacy policy. Privacy notes should be accompanied with terms of use, and with a rational acceptable use policy. Furthermore, Marrakesh Treaty should also be taken into consideration, to enable access to learning resources to all the learners and teachers, without any disability discrimination. The corresponding regulation for US, which is not a member of the World Intellectual Property Organization is the Equality Act (equalityhumanrights.com/en/equality-act). All the pointed issues are primarily recommended for enetCollect's framework, but they are also applicable to all the existing or new educational platforms worldwide, including the crowd-oriented ones.

After alerting the prospective users about all these documents, a written consent about data privacy and intellectual property should be obtained from all of them. But first, the users should be properly introduced to the documents and advised to read them carefully. To do so, they should be as clear as possible, very concise and easily comprehensible.

To guarantee that all the sensitive student information are safeguarded, the regulations defined should be obeyed with no exclusions. Accountability measures should be strict. Otherwise, enetCollect's system will be one of those experiments, which impose "privacy concerns and the safety of student data as obstacles" (Johnson et al, 2016).

6. Bibliographical References

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *Int. journal for the scholarship of teaching and learning*: 4(2), 17.

Bergan, S., & Deca, L. (2018). Twenty years of Bologna and a decade of EHEA: what is next? *European higher education area: The impact of past and future policies* (pp. 295-319). Springer, Cham.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE review*, 42(4), 40.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, 19(2), 171-209.

DLA Piper (2019). DLA Piper GDPR data breach survey: https://www.dlapiper.com/en/uk/insights/publications/2019/01/gdpr-data-breach-survey/

Drummond, C., & Fischhoff, B. (2017). Individuals with greater science literacy and education have more polarized beliefs on controversial science topics. *Proceedings of the National Academy of Sciences*, 114(36): 9587-9592.

Eckersley, P. (2010). How unique is your web browser? *International Symposium on Privacy Enhancing Technologies Symposium*, Springer: 1-18.

EFF (2018). Yet another lesson from the Cambridge Analytica fiasco: Remove the barriers to user privacy control, https://www.eff.org/deeplinks/2018/03/why-we-didnt-make-fix-my-facebook-privacy-settings-tool

EHL, Edmodo Help Center (2017). Important notice about your Edmodo account: https://support.edmodo.com/hc/en-us/articles/115007376848-Important-Notice-About-Your-Edmodo-Account

EC, European Commission (2018). Data protection: 2018 reform of EU data protection rules, https://eur-lex.europa.eu/eli/reg/2016/679/oj

Flanagan, B., & Ogata, H. (2017). Integration of learning analytics research and production systems while protecting privacy. *The 25th International Conference on Computers in Education*, New Zealand: 333-338.

Godwin-Jones, R. (2017). Scaling up and zooming in: Big data and personalization in language learning. Language Learning & Technology, 21(1), 4-15.

Greidanos, P. (2019). Moodle.org outage and data loss: https://moodle.org/news/#p1535490

Halder, B. (2014). Evolution of crowdsourcing: potential data protection, privacy and security concerns under the new media age. *Revista Democracia Digital e Governo Eletrônico*, 1(10): 377-393.

Johnson, J. A. (2014). The ethics of big data in higher education. *International Review of Information Ethics*, 21(21), 3-10.

Johnson, L., Becker, S. A., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). NMC horizon report: 2016 higher education edition (pp. 1-50). The New Media Consortium.

Joiner, M. C. (2018). To see or not to see: the constant conflict between promoting public access to information whilst maintaining confidentiality, *Student Records*.

Jones, M. L., & Regner, L. (2016). Users or students? Privacy in university MOOCS. *Science and engineering ethics*, 22(5): 1473-1496.

Mello, S. (2018). Data Breaches in Higher Education Institutions, *University of New Hampshire*

Poore, M. (2015). Using social media in the classroom: A best practice guide. Sage.

Sandeen, C. (2013). Assessment's place in the new MOOC world. *Research & practice in assessment*, 8, 5-12.

Sen, J. (2015). Security and privacy issues in cloud computing. *Cloud technology: concepts, methodologies, tools, and applications* (pp. 1585-1630). IGI Global.

Soltani, S., & Seno, S. A. H. (2017). A survey on digital evidence collection and analysis. *7th International Conference on Computer and Knowledge Engineering (ICCKE)*, IEEE: 247-253.

Widup, S. (2010). The leaking vault: Five years of data breaches. *Digital Forensics Association*, 1-42.

Zdravkova, K. (2016). Reinforcing social media based learning, knowledge acquisition and learning evaluation. *Procedia - Social and Behavioral Sciences*, 228: 16-23.

Zeide, E., & Nissenbaum, H. (2018). Learner privacy in MOOCs and virtual education. *Theory and Research in Education*, 16(3), 280-307.

Rights of data subject	Transparency and modalities	Access to personal data	Rectification and erasure	Right to object	Restrictions
Blackboard blackboard.com	Partial compliance	Complete compliance	Complete compliance	Complete compliance	Complete compliance
Canvas canvaslms.com	Partial compliance	Complete compliance	Complete compliance	Partial compliance	Partial compliance
Duolingo duolingo.com	Complete compliance	Complete compliance	Complete compliance	Complete compliance	Complete compliance
Edmodo edmodo.com	Partial compliance	Complete compliance	Not designated	Not designated	Complete compliance
EdX www.edx.org	Complete compliance	Complete compliance	Complete compliance	Complete compliance	Not designated
FutureLearn futurelearn.com	Complete compliance	Complete compliance	Complete compliance	Complete compliance	Complete compliance
Khan Academy khanacademy.org	Complete compliance	Complete compliance	Complete compliance	Complete compliance	Complete compliance
Mechanical Turk mturk.com	Partial compliance	Not designated	Not designated	Not designated	Not designated
Moodle: Moodle.org	Complete compliance	Complete compliance	Complete compliance	Complete compliance	Complete compliance
SAPSuccessFactors successfactors.com	Complete compliance	Complete compliance	Complete compliance	Complete compliance	Complete compliance

Table 1: Rights of data subjects in learning management systems and crowdsourcing platform