The final publication is available at ACM via http://dx.doi.org/10.1145/3362789.3362823

GDPR Security and Confidentiality compliance in LMS' a problem analysis and engineering solution proposal.

Dani Amo Filvà†
Department Name
Institution/University Name
City State Country
email@email.com

Marc Alier, Maria Jose Casañ
ESSI
UPC
Barcelona
marc.alier@upc.edu ,
mjcasany@essi.upc.edu

Fonsi, Fran
Department Name
Institution/University Name
City State Country
email@email.com

ABSTRACT

The authors have studied the main Learning Management Systems (LMS) and found that all the personal information and activity and logs are stored unencrypted on the server filesystem and databases. This means that a user with access to such resources may have full access to all the personal information and meta information. Making the LMS installation very difficult to comply with the GDPR and very vulnerable to information leaks due to targeted hacker attacks.

In this paper, we analyze this problem from a technical and operational perspective for the open-source market leader LMS: Moodle, and we propose a solution and a prototype of implementation.

CCS CONCEPTS

• Security and privacy~Information-theoretic techniques • Computing methodologies~Self-organization

KEYWORDS

Learning analytics, GDPR, confidentiality, data privacy, digital identity, data security management, learning management systems

ACM Reference format:

FirstName Surname, FirstName Surname and FirstName Surname. 2018. Insert Your Title Here: Insert Subtitle Here. In *Proceedings of ACM Woodstock conference (WOODSTOCK'18). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

1 Introduction

In recent years some organizations that provide services in the education sector, have been caught red-handed using personal information of students and metadata for unjustified commercial purpose. This has contributed to raising concerns about the management of personal information and the data generated while using online educational apps, such as a Learning Management System (LMS). This is especially relevant in the case of Learning Analytics (LA) since it involves collecting, storing and analyzing student personal data, metadata of their behavior [3]. The participation of multiple actors leads to a situation where different entities can be using sensible information and their analysis for their own benefit.

We can find plenty of examples: Williamson points out in the use of apps such as ClassDojo to change social and emotional behavior in minors [14]. There are cases of unethical practices like inBloom Schools or AltSchool [1], in which students were treated like guinea pigs and raw material for the improvement of the software. And of course, China is using all kind of surveillance technologies on their schools [4–6] in practices that for western values would be unethical.

All these situations show that both the research and the applications of Learning Analytics need to improve in their consideration of ethics, confidentiality, and security. A recent study conducted by the authors detected that privacy, security, and confidentiality information were like the elephant in the room in the studies presented at the LAK conference in recent years.

The Learning Management System is the system that stores and manages most of the sensitive personal data and metadata. And a quick analysis of these systems reveals that the student's digital identity and personal data are also unprotected: Data is stored unencrypted in the database or the filesystem. Every user that has access to the server has full access to all information. If the system is hacked the bad guys hit the jackpot and can steal all data from students. According to The K-12 Cybersecurity Resource Center [10] one attack is successfully executed every 3 days since 2016. So, this is not a small matter.

Interoperability between platforms is a cause of breaking the chain of custody of information and a potential source of data leaks when using third-party untrusted or untrustworthy add-ons.

In addition, the transfer of data between countries converts the situation into one of more complex and political nuance. The GDPR [12] imposes legal regulations on data transfer processes outside the EU (GDPR, recital 101), but it is the recipient governments that end up having the management and control of access.

This complex situation of distrust in the use of personal data is approachable from two directions: a) the implementation of the legal regulations and ethical codes into the business processes. b) a sound technological implementation to support it. We propose a combined action so that the technological approach automates the business rules that ensure compliance with legality, agreements with users and the ethical code [2]. It is necessary a technological alignment with the current legislation, the confidentiality policies of the entities, the exercise of the rights of the users and the current LMS.

Automating the GDPR in the LMS could help solve the state of distrust if is communicated effectively. It implies solving a technological challenge: validating the user authorization in all data access transactions. Hence, we need guarantee the user has the proper authorizations every when the running code accesses data on its behalf. We consider it a double problem since it must be confirmed that the identity of the user is the same both in the execution of the LMS code and in the access to the stored data. We call the problem of "double authorization".

Therefore, we are faced with a technical problem with a complex solution that encompasses different levels of abstraction, starting by getting a clear understanding of the legal regulations and its implications down to the realm of software engineering and systems operations.

In this paper, we analyze the problem of double authorization in LMS. First, we introduce the problem. Then we continue with the legal approaches and continue with the technological ones, both in relation to the problem. Finally, we propose a neutral technological solution that could be potentially implemented in most LMS with a reference implementation prototype for Moodle.

2 Privacy and Confidentiality: the context marks its use

The concepts of privacy and confidentiality have often been confused and used incorrectly. We consider appropriate a clarification in the concepts of privacy and confidentiality:

The word "privacy" refers to protecting physical and intimate issues of a person, such as protecting her from physical contact. The word "confidentiality" refers to safeguarding the transmission of information, such as sharing academic data or health records. Privacy is invaded and confidentiality is breached [7, 11].

The word "privacy" comes from the Anglo-Saxon privacy and its translation into different languages of the European Union like "privacidad" in Spaniard refers to intimacy. However, data protection laws refer to the safeguarding of personal information collected. Therefore, in the GDPR "confidentiality" is used in the exercise of data protection instead of "privacy". However, in other legal contexts such as in the USA, the word "privacy" is closely linked to the protection of medical records and, therefore, to intimate matters. It is thus attributed to a meaning of protection of personal information, which is why its use is extended in contexts in which reference is made to data protection.

Therefore, when dealing with intimacy protection we can establish an order of importance where "privacy" is placed above "confidentiality" because it implies a superset of restrictions. However, in data protection laws the right word used is "confidentiality" since the legal context refers to the management of information of a person's data.

In the present work, we use confidentiality instead of privacy due to the context of information processing (personal data) in Learning Analytics processes. In the first place, because researchers are part of a country in the European territory where privacy refers to physical intimacy. Secondly, because the word confidentiality defines an adequate treatment of a person's information. And thirdly because only confidentiality appears in the legal texts of the GDPR, including in the English texts.

3 The double-authorization problem in the LMS

The interactions a student has online (access to resources and learning activities) create and manages a lot of personal data that LMS stores on the database and filesystem. The right to access and modify this information has very sensitive implications in terms of confidentiality and security.

The following questions need to be considered in terms of access rights:

- Who can access a particular set of information?
- Who can use it and why?
- Given a user and its role in the platform: which sets information can access and with what kind of access rights (read, write, read/write)?

And more questions arise related to security:

- Where is the information stored?
- How is the information stored? Is it encrypted? How safe is this encryption? Who has the keys?

The legal experts we have consulted pointed to us that a careful interpretation of GDPR requires that the identity of the user and the associated permissions be validated every time that confidential data is accessed: the user's authorization status needs to be checked.

But in the LMS confidential data is accessed from the source code of the LMS all the time. Actions like reading a class forum where we can see the classmate's names and pictures and links to their profile pages and accessing the list of course participants or the history tab on a wiki page, all require accessing confidential information about peer students. Obviously, a user with the role of teacher or instructor will be constantly accessing, modifying and adding confidential information like grades. In all these instances a proper use and access to confidential information are only guaranteed by the LMS source code. This means tens of thousands of lines of source code.

Can we guarantee that the code of your LMS of choice checks and respects carefully the permissions of every user and role before accessing confidential information? Off course not.

Can we guarantee that the code is bug-free? Neither. And even if we could do it at a level of the software vendor or open-source project, we still have to cope with all the custom code that most of the institutions run on their installations for purposes of integration with the .

To make matters worse modifying the code of an LMS is relatively easy because most of the programming languages for web development are interpreted, so they are visible and editable from a simple text editor. The source code can be modified by an employee or a hacker attack, especially in those LMS that run on interpreted languages like PHP or Python. Any modification at the code level can open yet another breach in the confidentiality or security of personal data and break with the legal requirements.

Therefore, we are looking for a solution that can withstand bugs in the code and malicious code modifications.

4 Legal approach

The GDPR exposes that "rapid technological developments and globalisation have brought new challenges for the protection of personal data. The scale of the collection and sharing of personal data has increased significantly. Technology allows both private companies and public authorities to make use of personal data on an unprecedented scale in order to pursue their activities. Natural persons increasingly make personal information available publicly and globally." (GDPR, recital 6). Therefore, the European Union promulgates a new legal framework (GDPR) to reduce legal uncertainty and generate confidence in the treatment of personal data of its citizens.

4.1 Granularity in the GDPR

In point 2 of Article 4 of the RGPD, treatment is defined as "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;". Therefore, we consider that the right to object is critical in the resolution of the problem, as well as its contextualized application. In an LMS context, we consider that the LMS itself, describes the global one and the courses describe its subcontexts.

Each user of the LMS has the right to have guaranteed confidentiality and security of their personal data in a way that:

1. Can object to the processing of personal data by exercising the right to object. The right to object focuses on past, present and future data.
2. Can indicate when to eliminate the personal data by exercising the right to be forgotten. The right to be forgotten (deletion) focuses on past data.

The figure of the Data Provider Officer allows the resolving of requests regarding the erasure of personal data collected. Technically this is already possible. However, the LMS is not ready to resolve the requests regarding the right to object at a given time, and in a specific course.

4.2 Legal agreements and conditions

With the legal agreements, the institutions regulate the uses of the services provided to the students, such as the LMS. The relationship between the educational entity and the student is contractual. Students pay to obtain a service. This implies that:

1. The students of an LMS must accept the conditions expressed by the educational institution.
2. The conditions must be informed and made available in legal agreements, privacy policies or terms of use.
3. The students cannot discuss the clauses.
4. The students can use the LMS once they have accepted the conditions.

Moreover, the GDPR grant to students a series of rights that allow them to evade accepted agreements in exceptional situations. These rights are (GDPR, recital 156): to rectification, to erasure, to be forgotten, to restriction of processing, to data portability, and to object when processing personal data for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. Hence, the educational institutions have the obligation to set use clauses that respect the students' rights.

4.3 Right to object

The student has the right to object to the processing of the personal data, once the student accepts the conditions. The GDPR expressly recognizes the right of the interested party to object at any time, for reasons related to their particular situation (recital 50 y article 21), that concerning personal data are subject to a treatment based on the provisions of the article 6 of the chapter II "Principles", point 1, letter e) or f) of the GDPR. The person in charge of data processing will have to prove that their interests prevail over those of the interested party [13]. Moreover, the right to object must be granular in legal agreements.

The courses in an LMS, are usually legally covered by a unique legal agreement. However, due to personal causes, the students can exercise their right to object in all of the courses or in a specific course. Therefore, the LMS must technically consider this possibility, especially due to causes of gender violence or the preservation of identity for policies.

In no case, the student will be able to set conditions in the treatment or in the use of her data. The student cannot choose to hide or show her profile picture to specific classmates. When a student accepts the conditions, she only has some rights to exercise, such as the right to object. However, the student can exercise her rights in a context or another, being these contexts the departments or courses of the LMS of the same educational institution.

5 Technological approach

Since the introduction of GDPR, most LMS have been updated to facilitate the process of dissemination of terms of service and tracking the acceptance of terms of service by the users and blocking the access to the users who do not agree to terms. However, the granularity in the right of opposition is not supported in any of the LMS analyzed, including Moodle. The cases where a user could exercise her right to reject certain conditions when in theory it would not impede the access to the service are not supported. Is all or nothing.

This granular level of control, although considered an exception, must be available to avoid problems of social connotations. It is the case of gender violence victims or civil servants like police who need certain level of confidentiality. For example, a student could refuse to display her name to classmates or display an alias and fake picture.

Some may say (and said on interviews the authors performed) that these cases constitute exceptions and the design of the system depend on them. But the authors consider that these exceptions should indeed be integrated into the very design of the system, like accessibility ramps into public buildings.

In any case, a technological approach is required to deal with the exceptions derived from the application of the GPDR.

The GDPR imposes a series of conditions on technological solutions to ensure adequate security and confidentiality of personal data. The GDPR provides that:

1. Guaranteeing the security and confidentiality of personal data requires a neutral technological approach and should not depend on the techniques used (GDPR, recital 15). Defining a solution based on a particular technology breaks with the recitals and articles of the GDPR. However, a paradox occurs since any technical solution requires applying concrete and appropriate approach to the context.
2. In order for the identity of the student to be preserved throughout the cycle of personal data processing, the technological solution must consider safeguarding processes for effective protection of data from the design and by default of the products and services [9].

To address the technological paradox and solve the problem from the design and by default, we propose a high-level solution and two actions. This double solution does not define the technology to be used, it helps to make the technology transparent to the user and guarantee the security and confidentiality of personal data:

All LMS must generate a matrix of granular access to data (access matrix) with identities, contexts and access permissions according to the legal agreements and the opposition rights already exercised and in force.

The storage system must double-check the identity by synchronizing with the LMS.

5.1 Access matrix

A student must accept the legal conditions of the LMS before using it. This process generates a record of those students who have accepted and those who are pending to accept. This register works as a granular data access matrix since it indicates in it which students the data can be seen and which ones are not yet. Therefore, this matrix is very important, since the LMS should use it to show or hide personal data of students.

In addition, the student can exercise the right of opposition to the processing of personal data. It must be reflected in the data access matrix both those students who have accepted or not the legal conditions and those who have exercised the right to oppose the processing of data.

By default, when a student accepts the conditions, the LMS allows their classmates to see certain personal data of theirs, such as name, surname or profile picture. The teacher always has access to your data, even if the student has exercised the right to object to the processing of personal data. In addition, the right of opposition must be granular per each course. Therefore, each casuistry must be distinguished for an LMS to apply the GDPR well:

- Legal conditions: When a student accepts the legal conditions allows any student enrolled in the same subject to see their personal data.
- Right of opposition to data processing: When a student exercises the right of opposition to the data processing in an LMS is indicating
  1. that no interaction is registered in the system
  2. that in those subjects in which this has been applied right students cannot see your data or know of its existence. While the student has accepted the conditions and is registered in a subject, the teachers of such subjects must be able to access their personal data and identify them for the proper functioning of the evaluation.

A data access matrix must have the following internal structure:

Table 1. Data access matrix

	Legal cond. 1	**Legal cond. 2**	**Legal cond. 3**	**Course 1**	**Course 2**
				Right to object	**Right to object**
Student 1	✓	✓	✓	X	X
Student 2	✓	✓	✓	X	✓

In the above data access matrix, we show "Student 1" who has accepted the legal conditions and has not exercised any opposition to data processing. On the other hand, "Student 2" has accepted the legal conditions, in "Course 1" it does not oppose the processing of data, but in "Course 2" it does oppose. The data access matrix must contemplate a higher granularity since the

law so indicates. For example, separating the right not to be seen by the students, but to store their interactions in their logs.

6 Hands on enforcing confidentiality

6.1 Add granularity and responsibility to the access matrix

All LMS implements a form of access matrix to determine what authorization each user has.

To ensure the level of confidentiality and security proposed by the GDPR, we must add complexity levels to the access matrix and add additional checks.

The access matrix must consider that on the one hand (1) the user logged in has accepted the legal conditions (and responsibilities), and on the other hand (2) guarantees the possible rights to object exercised by their peers. Thus, we ensure that:

- A user of an LMS can only see the data of other users according to the legal conditions accepted and their exercise of the right to object.
- Allow the person in the organization that performs the functions of Data Privacy Officer (DPO) to manage who or what roles, which by hierarchy are above the students, can freely access their data, as is the case of the technical administrator.

The condition (1) "the logged-in user has accepted the legal conditions (and responsibilities)" has been incorporated in the latest versions of systems such as Moodle, which manages this in its mid-2018 update to support the GDPR.

6.2 Encrypt and add double authorization

In most or all of the main LMS of the market, when a student accesses an LMS, he generally uses authentication based on the user name or email and password. At the time of authentication, a temporary session is created to validate any student action within the LMS. That is the standard authentication process, and it guarantees us (to some extent) that the user is who he is and therefore the LMS code can make use of his level of authorization when executing code in his session. For that, as we have seen in 6.1, the access matrix that the LMS implements internally is used for each context of the LMS (course, group, activity), user, and function used.

Although we can expand the functionality of the access matrix to comply with what is explained in 6.2, we still have a significant problem: confidential data is not encrypted. With what confidential information can be leaked in cases like:

- Errors in the source code. The majority of LMS in the market are projects with more than 10 years of life [8], backward compatibility problems of plugins and content, and presumably with thousands of lines of "croft" code that nobody dares to touch just in case. This makes it very difficult to ensure that (1) there are no bugs, (2) that the access matrix is strictly respected and (3) that the modifications in the suggested access matrix will not introduce more bugs.
- Employees with the professional ethics of an expired peanuts snack pack who decide to sell confidential information to the highest bidder.
- Hackers attacks. Once a hacker has access to the system, he has access to everything. There are no internal layers of security.

Therefore, we propose that confidential data in tables and files (such as log files that feed LA systems) should be encrypted. Encrypting log files does not imply any technical problem since the system normally only writes to them, and only external learning analytics plugins or systems use them.

Encryption of data in tables, such as the table containing the user data, is more complex because it is accessed from all parts of the system. In systems, with strict object-oriented programming and where we are sure that the programmers do not take "shortcuts", it could be done easily creating subclasses that will manage the access to encrypted information. Of these systems, we do not know the existence of any.

The solution we propose is to create a database driver that encapsulates the real driver of the database used by the LMS. This driver, called AuthChecker (see Figure 1), analyzes the queries emitted by the LMS code. In the case of access requests to confidential data, the AuthChecker's mission is to consult the access matrix to verify that the user has authorization or not, before providing the decrypted data.

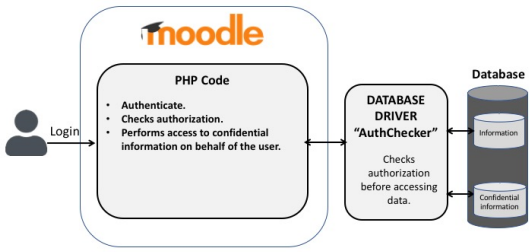

Figure 1: The solution we propose, called AuthChecker.

The AuthChecker should have two parts: one as a database driver compatible with the current system, which the LMS code can invoke transparently. And another part located in a safe place and only accessible to the DPO where the encryption and decryption of data are done.

The AuthChecker should have two additional components: a monitoring one that reports the unauthorized attempts to detect hackers and anomalies, and a debugging one in case the wrong queries are caused by system bugs.

7 Moodle

We have developed the solution to the problem of double authorization (see Figure 1). The development consisted in:

- Update and create a set of tables to store and encrypt the access matrix.
- Create an API to query the access matrix from the database, called "Matrix Access API".
- Create an API to log the unauthorized queries, called the "AuthChecker Monitoring API".
- Create a database driver (AuthChecker) to make requests to the Matrix Access API and enforce authorization to data queries.

Figure 2 shows the organization of the different developed components and the execution flow that enable the solution in Moodle.

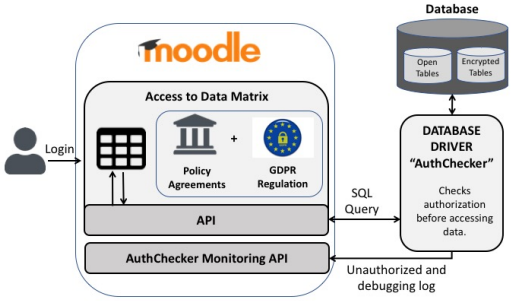

Figure 2: Components and execution flow of the developed solution.

This organization and operation ensure that the conditions defined in the access matrix are checked both in the Moodle code and at the query level of the database, thus checking who does it and if it has permission to access the information, or even if have the right to make the request.

8 Conclusions

In recent years some organizations that provide services in the education sector, have been caught red-handed using personal information of students and metadata for unjustified commercial purpose. This has contributed to raising concerns and distrust about the management of personal information in the Learning Management System (LMS). This is especially relevant in the case of Learning Analytics (LA) since it involves collecting, storing and analyzing student personal data, metadata of their behavior.

This complex situation of distrust in the use of personal data can be addressed by a) implementing legal regulations, such as GDPR, and ethical codes into the business processes b) and a sound technological implementation to support it. We propose a combined action so that the technological approach automates the business rules that ensure compliance with legality, agreements with users and the ethical code.

In most or all of the main LMS of the market, when a student accesses an LMS, he generally uses authentication based on the user name or email and password. At the time of authentication, a temporary session is created to validate any student action within the LMS. That is the standard authentication process, and it guarantees us that the user is who he is and therefore the LMS code can make use of his level of authorization when executing code in his session. All LMS implements a form of access matrix to determine what authorization each user has. However, this is not enough to ensure the levels of confidentiality and security established by the GDPR in regard to personal data. Therefore, we must add complexity levels to the access matrix and add additional checks.

We analyze this problem from a technical and operational perspective for the open-source market leader LMS: Moodle, and we propose a solution and a prototype of implementation.

The solution we propose is to create a database driver to enforce a second authorization check. The driver, called AuthChecker, ensures that the conditions defined in the access matrix are checked both in the LMS code and at the query level of the database.

REFERENCES

[1] AltSchool, funded by tech execs, is closing schools, losing students - Business Insider: 2017. *https://www.businessinsider.com/altschool-why-parents-leaving-2017-11?IR=T*. Accessed: 2019-04-07.

[2] Amo, D. et al. 2019. Personal Data Broker Instead of Blockchain for Students' Data Privacy Assurance. Springer, Cham. 371–380.

[3] Analytics - MoodleDocs: 2019. *https://docs.moodle.org/37/en/Analytics*. Accessed: 2019-04-07.

[4] Brainwave-tracking start-up BrainCo in controversy over tests on Chinese schoolchildren: 2019. *https://www.scmp.com/tech/start-ups/article/3005448/brainwave-tracking-start-china-schoolchildren-controversy-working*. Accessed: 2019-06-07.

[5] Camera Above the Classroom: 2019. *https://www.sixthtone.com/news/1003759/camera-above-the-classroom*. Accessed: 2019-04-07.

[6] Chinese schools use facial-recognition gates to monitor pupils: 2019. *https://www.dailymail.co.uk/news/article-7153981/Chinese-schools-use-facial-recognition-gates-monitor-pupils.html*. Accessed: 2019-06-07.

[7] Francis, L.P. Privacy and Confidentiality: The Importance of Context. *The Monist*. Oxford University Press.

[8] Hernandez, J.C.G. and Chavez, M.A.L. 2008. Moodle security vulnerabilities. *2008 5th International Conference on Electrical Engineering, Computing Science and Automatic Control* (Nov. 2008), 352–357.

[9] Hoel, T. and Chen, W. 2016. *Implications of the European Data Protection Regulations for Learning Analytics Design.*

[10] K-12 Cyber Incident Map: 2018. *https://k12cybersecure.com/map/*. Accessed: 2019-04-07.

[11] Privacy vs. Data Protection vs. Information Security: 2016. *https://blogs.upm.es/sse/2016/11/01/privacy-vs-data-protection-vs-information-security/*. Accessed: 2019-06-07.

[12] 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 Da: 2016. *https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679*. Accessed: 2019-06-27.

[13] Vollmer, N. 2018. Article 6 EU General Data Protection Regulation (EU-GDPR). (Sep. 2018).

[14] Williamson, B. Decoding ClassDojo: psycho-policy, social-emotional learning, and persuasive educational technologies. DOI:https://doi.org/10.1080/17439884.2017.1278020.