Learning Tracking Data Analysis
How Privacy Issues Affect Student Perception on e-Learning?

Madeth May¹, Sébastien Iksal¹ and Claus Alexander Usener²
¹LUNAM University of Maine, Avenue Olivier Messiaen, 72085 Le Mans, France
²University of Münster, Leonardo-Campus 3, 48149 Münster, Germany

Keywords: Privacy Issues in E-Learning, Privacy Threat, Tracking System, Tracking Data, Data Analysis.

Abstract: Research works from the past ten years have demonstrated that technologies could in many ways enhance learning experience. Meanwhile, technologies can also create obstacles to the latter. For instance, using tracking approaches on the majority of e-learning platforms to monitor learners’ activities raises many privacy questions. As for learners, knowing that their personal data are being used, even for educational purposes, they could radically change their perception on e-learning technologies. This paper presents a study on privacy issues in e-learning, based on both existing research findings and an experiment that we have conducted with the participation of students from three universities in France and one university in Germany. The study covers two main aspects. First, it outlines various tracking approaches in e-learning. Second, it analyzes how the participants perceive the use of their tracking data and the related privacy issues. The major contribution of this paper is the awareness-raising of privacy concerns, which are often overlooked by researchers and e-learning content providers.

1 INTRODUCTION

In 2011, when we first presented a study on security and privacy issues in e-learning (May and George, 2011a), we pointed out the lack of data protection measures from learning content providers. We also discussed how users relied on trust when accessing online learning platforms, and how technical solutions still had their limitation in terms of privacy protection. Since then, we continue to expand our research scope by focusing more on users and their perception on the use of their personal data in various educational settings. The study presented in this paper combines existing research findings with the empirical data acquired throughout the experiment that we have conducted with the participation of students from three universities in France and one university in Germany.

The purpose of this paper is to share scientific findings based on field studies and empirical data. Our research team has no intention to make any claim regarding how to definitely solve privacy matters that one might encounter accordingly to a variety of factors, including institution’s policies and regulations, learning contexts and cultural points of view. Nonetheless, we hope that the discussion made in this paper could raise awareness of the issues in question, which are often neglected in the research efforts that involve user tracking and data analysis.

This paper is structured as follows. The second section provides an overview of our research work that emphasizes on an explicit tracking approach to efficiently monitor users’ activities on e-learning platforms. A general discussion on user tracking approach is made in the third section. It is based on a number of related works, which help us gain a broader perspective on what causes privacy concerns. The fourth section is dedicated to a presentation of our experiment. Data analysis and commenting on results are made in the same section. In the last section, we draw a conclusion and highlight future work.

2 RESEARCH CONTEXT

2.1 User Tracking Approach to Enhance e-Learning Experiences

E-learning has been evolving rapidly, from purely Web-based to mobile and ubiquitous learning experiences, thus providing even more personalized
solutions that better suit each individual needs. For that reason, institutions, teachers and learners seemingly embrace e-learning (Popovici and Mironov, 2015) and consider it among the most innovative learning mediums (Gaur et al., 2015). In fact, within the past ten years, we witness a strong ongoing growth of research interests in e-learning (Scott and Vanoorbeek, 2007) and an emergence of technologies that better support user interaction (Gamage et al., 2014) and content sharing (Lau et al., 2013).

In order to make e-learning more efficient in terms of student monitoring, other research efforts like (Corbi, 2014) focus more on user tracking approach that consists of collecting data of users and of their activities throughout the learning environment. By doing so, the teachers are able to follow, for instance, the activities being undertaken by the students and to observe their behaviors on learning resource consumption (Gómez-Aguilar et al., 2015). As for the students, having records of their own activities allows them to keep track of their individual progress, their interactions among other students as well as their achievements during the learning session (Qahmash, 2013).

Acknowledging the contributing factor of “user tracking approach” to high quality teaching and learning guidance in e-learning as pointed out by (Jermann et al., 2001), researchers and learning content providers choose to integrate systematically a tracking system in their educational settings. To back up this claim, a more recent study of (Alowayr and Badii, 2014) reviews a variety of learning platforms that make use of students’ tracking data for different purposes, among which are student assessment and evaluation. Further evidence on how tracking approaches are broadly used to help enhance e-learning can be found in the research work of (Burdescu and Mihăescu, 2010; Jeske et al., 2014; Mehra, 2015).

Implementing a user tracking system in e-learning has been done in numerous ways as identified by (Corbi, 2014; Kim et al., 2008; Popescu and Cioua, 2012; Pozzi et al., 2007). With the progress being made in regard to the tracking mechanism, the technique of collecting data has become more sophisticated and powerful while being effortless to be deployed in an existing e-learning environment. In the meantime, such progress has increased privacy questions, which lead to a situation where privacy protection is becoming crucial for users, researchers and e-learning content providers.

### 2.2 Tracking Data Analysis

Our research team works on Technology Enhance Learning (TEL) and studies numerous research questions related to “Modeling the Observation of Usage Tracks and their Analysis” (Choquet and Iksal, 2007). A part of our research work involves user tracking on e-learning platforms where Computer-Mediated Communication tools (i.e., forums, chat, newsgroups, etc.) are widely used to foster learning activities and to compensate the lack of face-to-face interactions among the participants (Berge and Collins, 1995). We proposed an explicit tracking approach to assist the conceptual design and the implementation of a tracking system for a variety of Web-based communication tools (May et al., 2008). The proposed approach was built upon a tracking mechanism that simultaneously collects fine-grained data from both client and server sides. The technical aspects of the tracking approach in question can be found in (May et al., 2010a).

Another part of our research effort focuses on tracking data exploitation where research challenges were studied in (May et al., 2011). Our goal is to make use of the tracking data in order to help the participants in exploring their past and ongoing activities on a learning platform. With the technical support of UTL (Usage Tracking Language) (Iksal, 2007) in terms of data analysis and visualization, the participants could not only examine their activities, but also make an assessment of their effectiveness and achievements.

The need for data analysis and visualization tools can be briefly explained as follows. With the current support of e-learning that prioritizes content sharing and user communication (Gamage et al., 2014; Lau et al., 2013), the participants are compelled to neglect some fundamental facets of online learning, such as self-monitoring. Not to mention the online interaction between the participants that makes the supervision task very laborious for the teachers. Last but not least, while the students usually encounter difficulties in getting feedback on their own activities, the teachers, on the other hand, are often constrained by the lack of technical assistance to conduct a proper analysis on the students’ tracking data.

Having studied these issues, we addressed the importance of technical support in tracking data analysis and visualization in order to enhance e-learning experiences for both the teachers and the students. Therefore, we have designed and developed TrAVis (Tracking Data Analysis and Visualization tool) for the teachers who are in need
of supervising students’ activities (May et al., 2010b), and also for the students who seek to self-monitor during their learning sessions (May and George, 2011b).

To illustrate how tracking data analysis could contribute to e-learning enhancement in terms of student monitoring, we give an example of a “data indicator” representing a group activity on a discussion forum below. For the sake of comprehension, the “data indicator” refers to a piece of information, extracted from a set of tracking data. Generally computed in graphical representations, a data indicator features the process of the considered “cognitive system” learning activity, the characteristics or the quality of the interaction being performed in a learning environment.

![Figure 1](image)

Figure 1: An example of data indicator that analyzes the participation level of two groups of students.

Figure 1 gives an example of a visualization of two groups of students on three forums that have the same structure, dedicated to group collaboration. Each radar graph, filled in with a distinct color, provides a quick perception of the forum and its access frequency, number of active participants, threads, messages, and files, etc. Hence, the teacher can make use of these indicators to analyze the interactions among the students. On top of that, the given quantitative information allows the teachers to evaluate the collaboration level of each group. For instance, Figure 1 shows that group A has more intense interactions than group B in almost the three forums. Such indications can be used to (i) compare the participation rates of both groups during the collaborative task or (ii) to evaluate the productivity rates of one group in relation to another, according to the number of created threads and shared files. More data indicators and their analysis can be found in (May et al., 2011).

To sum up, data indicators computed from e-learning tracking data provide means of awareness, assessment and evaluation of a learning situation. However, obtaining data indicators involves a complex procedure that starts with a tracking technique. Indeed, the latter is crucial to the whole data gathering process and always has an impact on the production of quality and substantial data indicators. Consequently, most tracking techniques that are robust and efficient in collecting users’ data are at the same time very intrusive (Corbi, 2014; Popescu and Cioiu, 2012). To gain a broader perspective on how tracking approaches could cause privacy issues, the study in the following section covers a number of research works with examples of how tracking data are used.

3 RELATED WORKS

3.1 Understanding the Correlation between User Tracking and Privacy Issues

Using data indicators in the analysis process enables one to synthesize, infer and interpret the information that it features. While it gives considerable assistance to the teachers in the tasks of monitoring online learning, it also creates major drawbacks for students. For example, some students who are cautious about privacy matters would become apprehensive of being traced and of having no control over how their personal data are being exploited. To better understand this phenomenon, we studied a number of research works where data indicators computed from e-learning tracking data are being exploited by the teachers to gain awareness, to make an assessment and to evaluate students’ activities.

The first is Argunaut (Groot et al., 2007), which is an awareness tool, built to keep track of online interactions. It provides data indicators of collaborative learning activities, thus allowing the
teachers to examine the behavioral aspects of each individual student during a collaborative task. iHelp (Brooks et al., 2006) is another awareness tool that shares the same characteristics as Argunaut. iHelp assists the teacher in supervising the communication process among the students. Besides gaining awareness, making an assessment of students’ activities is also compulsory. For that matter, (Shaul, 2007) suggested an analysis of various aspects of individual or group activities (e.g. social interactions). On the other hand, (Gibbs et al., 2006) proposed a platform that offers means to analyze the temporal and spatial dimensions of students’ interactions.

Regarding the evaluation tools that provide data indicators on students’ activities and their outcomes, (Mazza and Dimitrova, 2003) introduced CourseVis to teachers who wish to study and evaluate the social aspect of each student in a learning session. Not too far from CourseVis in terms of data analysis, GISMO, a Graphical Interactive Student Monitoring tool (Mazza and Botturi, 2007), proposes another way to visualize behavioral and social data of students’ activities. Its objective is to help teachers evaluate the involvement of the students during the course activities on a learning platform (e.g. Moodle).

For a more complex learning environment like MOOC, (Coffrin et al., 2014) made used of data indicators to classify student types and to analyze students’ engagement and performance throughout their learning process. Last but not least, Glass (Gradient’s Learning Analytics System), a Web-based platform by (Leony et al., 2012), offers the possibility to keep track of students’ activities and to evaluate their performance in a given learning context.

Thus far, our primary observation regarding the existing tools is that most of them aim to better support the teachers in exploiting students’ tracking data. To do so, they explore every possible piece of information related to the student activity in order to accordingly generate significant data indicators on the latter. The intrusive characteristics of each tool allow, on the one hand, a pertinent analysis on students’ activities, but causes major privacy concerns on the other hand (Bandara et al., 2014). While some recent research efforts like (Esposito, 2012) and (Ivanova et al., 2015) took into account the need of users in controlling how their personal data are being used, only a few are accessible by the students. This is due to their restricted user rights from a technical standpoint, as well as their roles in the learning process. As a result, the students always comply with the regulations of a learning platform and put their trust on the latter (Juan Carlos Roca et al., 2009). Consequently, the privacy concerns remain to be addressed as the students are not always in the position to determine what data to share and who to share with.

Another observation is relevant to the strong focus of the previous works, placed on the efficiency of the tracking approach and the data indicators. The privacy concerns seem to be overlooked even though they have a direct impact on student behavior in e-learning as studied by (Brown, 2008). Sharing the same concerns as (Kanuka and Anderson, 2007), our research team attempts to study how privacy issues perceived by the students could make changes to the behavioral aspects of their activities. For instance, the confidentiality and anonymity can alter student engagement and performance as an individual or a group. Without necessary protection measures, students are becoming too afraid of being exposed to what meant to help them learn in the first place as found in the studies of (Bandara et al., 2014).

### 3.2 Identifying the Privacy Concerns

According to (Klobucar et al., 2007), privacy concerns are mainly caused by the use of technologies. As a matter of fact, with the growth of new platforms, new learning opportunities can be created along with new problems. The participants usually require technical knowledge of how technologies work in order to understand the privacy levels and threats. Nevertheless, the lack of information and technical support in that matter causes the most privacy concerns.

Learning service and content providers also bear the responsibility of intensifying the privacy issues. The students frequently ask the question of how their personal data are stored and protected by the learning content providers. As found in the study of (Anwar and Greer, 2006), the students are primarily concerned with trust assessment of learning environments they are using, and with the protection of their sensitive data.

The study of (Kanuka and Anderson, 2007) pointed out that privacy issues are also related to the participants’ consent, data ownership, confidentiality and anonymity. The students expressed their concerns regarding how information collected throughout the learning process would be kept secure and private. The confidentiality is part of the privacy protection that refers to students’ rights to control the access of their tracking data as well as other information about them. Regarding the
anonymity, most students are unaware of their right to request a removal of any characteristics (i.e., name, address, affiliated institution, geographical area, etc.) that would allow them to be identified.

Our findings reveal that most students regret not being part of the decision making on what information to be collected, what to be used, and for what purpose. Despite the compromise they have to make when consuming resources on an e-learning platform, students expect to have a choice to accept to be traced, to deny the use of their data or to limit access to some users.

Privacy issues also concern the security threats of technologies we are using. Indeed, students are exposed to a risk of data and identity theft. Such issues make students doubt the confidentiality and data protection measures proposed by their affiliated institutions or learning content providers. Research evidence can be found in the study of (Ben Arfa Rabai et al., 2012).

Our study also takes an interest in mobile e-learning where security and privacy threats remain challenging despite technological progress made on ubiquitous learning. This is due to the fact that granting access to mobile devices on learning contents opens doors to security threats that have not been taken into account by the learning service providers. The diversity of mobile devices and their security protection measures are varied in accordance with the operating system, the application used and the users’ own measures to protect their privacy. Research data from the study of (Yong, 2011) cover the privacy preservation for mobile e-learning. Yong pointed out the security threats regarding ubiquitous learning and the privacy preservation techniques for the students.

To summarize, security and privacy levels differ in various learning environments and depend on types of learning activities being conducted by the participants. In practice, it is not always straightforward or simple to promise absolute privacy, confidentiality and anonymity when using open e-learning environments. However, identifying clearly the privacy levels and their relative protection measures allows us to set rules and policies in terms of student tracking.

4 CASE STUDY

4.1 Set-up and Participants

On top of the study we made on existing research findings, we have conducted a semi-controlled experiment where TrAVis, the tool mentioned earlier in section 2.2, was used to analyze and visualize tracking data collected on a Moodle learning platform. Our main goal is to acquire an overview of student perception on privacy issues when their personal data are being used in an authentic learning situation. Our clear intention is to consult the students who are naturally concerned about their data in an actual practice setting instead of interviewing some random students. As a matter of fact, every student who participated in our experiment uses online learning platforms on a regular basis. A total of 178 students from three universities in France and one university in Germany participated in our study.

4.2 Procedure

The participants were assigned the task of using Moodle to organize their group activities. They were also encouraged to use a discussion forum, already integrated into Moodle, to perform their communication activities. Depending on their affiliated university, the participants had between one and two weeks to complete the assigned task. They were then asked to use TrAVis to analyze their personal data gathered during the period of the experiment. We also provided technical assistance to the participants in choosing tracking data to analyze and in visualizing data indicators on their past activities.

At the end of the experiment session, the participants were solicited to answer a questionnaire, which emphasizes on two main points: (i) their perception of privacy issues in e-learning, and (ii) their request for privacy protection measures. Other aspects of the questionnaire will be studied in our future work.

4.3 Experimental Data

Having acknowledged the necessity of protecting the participants’ personal data used in our study, we choose to discuss in this paper the findings from a general perspective. Experimental data will be presented without distinguishing the groups of participants, their respective university, and academic background.

The most significant data from the questionnaire that reflect how students perceive the privacy issues in e-learning are illustrated in figure 2. What gets our attention the most is the belief of the participants that an anti-virus or an anti-spyware would help them overcome the privacy issues in
e-learning. Indeed, 34% of the participants claimed to have a good protection system that prevents their personal data from being collected. Such misunderstanding is a big part due to the lack of technical knowledge on how a tracking mechanism works. It is also because the privacy issues are very confusing for the participants. For instance, the most frequently asked question during our experiment was whether or not they were tracked when using a browser in incognito or private mode to access their online learning environment. In fact, most participants do not have a clear perception on the tracking process and its correlation to the privacy concerns.

Figure 2: Student perception on privacy issues in e-learning.

Although the majority of the participants seem to figure out the most common security aspects in e-learning technologies, they still have difficulties identifying the related privacy threats. As confirmed the data from the fourth and fifth rows in Figure 2, only 21% claimed to have knowledge of the tracking process and considered it without harm to their private data. Respectively, 29% admitted that they have neither privacy nor security preoccupation in e-learning. On the contrary, over 78% disagreed with the previous claims and felt unsafe when using e-learning regardless the tracking process being deployed or not.

Figure 2 also reveals interesting data that supports our hypothesis regarding the impact of user-tracking approach on user behavior in e-learning. 68% of the participants expressed their fears towards a learning environment with an integrated tracking system. The participants recognized that the latter had sometimes affected how they perform certain types of activities. For example, they suggest limiting private activity or to reduce public intervention like on a discussion forum, so that they would leave the least of their traces possible on an open e-learning environment.

The second analysis we made on the data from the questionnaire focuses on privacy protection measures as seen by the participants, to help them get beyond privacy concerns. Table 1 shows the most demanding features regarding personal data protection, consent agreement, anonymous use of learning services, ethics legislation, and awareness raising. On a scale of 0 to 5, the participants expressed the least and the most important privacy protection measures.

Table 1: Most requested features in terms of privacy and personal data protection.

<table>
<thead>
<tr>
<th>Importance level</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness raising</td>
<td>2%</td>
<td>5%</td>
<td>10%</td>
<td>15%</td>
<td>33%</td>
<td>49%</td>
</tr>
<tr>
<td>Avoidance of personal data</td>
<td>2%</td>
<td>5%</td>
<td>10%</td>
<td>16%</td>
<td>18%</td>
<td>49%</td>
</tr>
<tr>
<td>Data protection</td>
<td>6%</td>
<td>11%</td>
<td>15%</td>
<td>22%</td>
<td>18%</td>
<td>38%</td>
</tr>
<tr>
<td>Anonymous access</td>
<td>10%</td>
<td>7%</td>
<td>18%</td>
<td>13%</td>
<td>33%</td>
<td>19%</td>
</tr>
<tr>
<td>Consent agreement</td>
<td>0%</td>
<td>3%</td>
<td>12%</td>
<td>33%</td>
<td>14%</td>
<td>38%</td>
</tr>
<tr>
<td>Ethics legislation</td>
<td>19%</td>
<td>29%</td>
<td>14%</td>
<td>13%</td>
<td>13%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Interesting information can be retrieved from Table 1. Examples include “avoidance of personal data” and “consent agreement”, which are both strongly relevant to privacy concerns in e-learning. In fact, consent is one of the keystones of privacy research practices in e-learning. Somehow, we were surprised to learn that most of the participants had never been reached out by anyone to sign a consent form. Yet, they have been regularly using Moodle, and their tracking data have been exploited in both educational and research settings. The data from Table 1 also shows that the participants consider “user data protection” and “anonymity” among the most important privacy provisions. As for “personal data protection”, the participants requested to be informed of the tracking process. According to the participants, being aware of the latter is one of the key facts to reduce privacy concerns.

If taken a closer look at how the participants perceive tracking approach in e-learning, while 49% of them claim that user tracking as a big threat, only 12% believe that ethics legislation could help them control the visibility and the use of their sensitive data. Interestingly, we have found similar results in our previous study (May and George, 2011a) that user tracking is not welcome even when users receive personalized content and assistance in return.

To wrap up, this study enables us to gain a broader perspective of the most crucial aspects regarding privacy concerns in e-learning. While we still need to conduct more analysis on the experimental data we have acquired, the early
findings point to the most critical measures to undertake to keep users informed of the privacy issues and to help them avoid confronting one. The study we made thus far also inspires us to explore a proper solution for our research work, which implicates user tracking and data analysis.

5 CONCLUSIONS

The research effort we presented in this paper analyzes existing findings and experimental data obtained from a case study on privacy issues in e-learning. While attempting to demonstrate, with research evidence, the benefits of a user tracking approach to e-learning enhancements, we also point out the necessity of gaining an insight on the privacy concerns that most participants encounter in their daily learning activities. Therefore, we hope to raise awareness of researchers, pedagogical teams and other e-learning practitioners in terms of user tracking and data analysis.

We also address the lack of guidance for the participants to acquire a better understanding on privacy levels and threats. Data from our study reveal that the participants have a very negative perception on e-learning technologies when it comes to privacy and data protection.

We recognize that avoidance of personal data is the most requested privacy provision, enabling students to anonymously access to e-learning platforms. However, we should also point out that a learning application aims at assisting students and so they cannot act in full anonymity. For that reason, participants in our study were always informed of the tracking process and given the right to control access to their data. On top of that, we always have a clear policy regarding the use of student tracking data in research and instructional purposes. For instance, the consent agreement is compulsory for the students and only authorized and anonymous data are used in our publications.

Our future work will focus on a more in-depth analysis of the current experimental data to explore other aspects like ethics in e-learning. We are also attempting to quantify and qualify the impact of the privacy issues on the behavioral, social and cognitive aspects of online learning. To do so, research colleagues from France, Germany and Greece are collaborating on an experiment to study the evolution of privacy questions and their associated threats by taking into account both ethical and cultural points of view.

REFERENCES


Juan Carlos Roca, Juan José García, Juan José de la Vega, 2009. The importance of perceived trust, security and privacy in online trading systems. Info Management & Computer Security 17, 96–113.


May, M., George, S., Prévôt, P., 2010b. TrAVis to Enhance Online Tutoring and Learning Activities: Real Time Visualization of Students Tracking Data, in: IADIS International Conference on E-Learning. Freiburg, Germany, pp. 57–64.


