Je-LKS
PEER REVIEWED PAPERS
LEARNING ANALYTICS: FOR A DIALOGUE BETWEEN TEACHING PRACTICES AND EDUCATIONAL RESEARCH
Vol. 15, n.3, 2019
ISSN: 1826-6223 | eISSN: 1971-8829
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal

DATA MANAGEMENT IN LEARNING ANALYTICS: TERMS AND PERSPECTIVES

Claudia Bellini, Annamaria De Santis, Katia Sannicandro, Tommaso Minerva

University of Modena and Reggio Emilia
{claudia.bellini; annamaria.desantis; katia.sannicandro; tommaso.minerva}@unimore.it

Keywords: Learning Analytics, Data Protection, Distance Education, Data Management, Ethics

Online teaching environments acquire extremely high granularity of data, both on users' personal profiles and on their behaviour and results. Learning Analytics (LA) is open to numerous possible research scenarios thanks to the development of technology and the speed of data collection.
One characteristic element is that the data are not anonymous, but they reproduce a personalization and identification of the profiles. Identifiability of the student is implicit in the teaching process, but access to Analytics techniques reveals a fundamental question: "What is the limit?" The answer to this question should be preliminary to any use of data by students, teachers, instructors and managers of the online learning environments.
In the present day, we are also experiencing a particular moment of change: the effects of the European General Data Protection Regulation (GDPR) 679/2016, the general regulation on the protection of personal data that aims to standardize all national legislation and adapt it to the new needs

for citations:
Bellini C., De Santis A., Sannicandro K., Minerva T. (2019), *Data Management in Learning Analytics: terms and perspectives,* Journal of e-Learning and Knowledge Society, v.15, n.3, 133-144. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1135021

dictated by the evolving technological context.
The objective of this work is to propose a three-point checklist of the questions connected to the management and limits of teachers' use of data in Learning Analytics and students' right of transparency in the context of Higher Digital Education, to take into account before conducting research.
To this end, the paper contains an examination of the literature on privacy and ethical debates in LA. Work continues with legislative review, particularly the Italian path, and the discussion about online data management in our current universities' two contexts: technology and legislation.

1 Introduction

According to the definition provided in the first International Conference on Learning Analytics (LA) and Knowledge held in Alberta in 2011, "learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs."

Learning efficacy is a principal goal of Higher Education institution didactic strategies, especially now that their attention is becoming focused on learner-centred pedagogical approaches. Now that they can design whole courses embedded in Learning Analytics, universities are forced to adopt new strategies in the way these goals are achieved, and it must have a clear idea of how to go about doing this.

In the current education context, "this call is gaining a new level of urgency" (Slade & Prinsloo, 2013, p. 31), especially in the last few years with the emergence of educational platforms, mobile-learning, micro-learning, and an increasing use of video resource as didactic tools. Moreover, understanding the potential of Learning Analytics is urgent – "as well as the changes that may be required in data standards, tools, processes, organizations, policies, and institutional culture" (Campbell *et al.*, 2007).

In line with Slade and Prinsloo (2013), "approaches taken to understand the opportunities and ethical challenges of Learning Analytics necessarily depend on a range of ideological assumptions and epistemologies" (p. 3). At the same time, they depends also on who is managing the data: leaders of institutions, teachers and academic staff (database administrators, educational researchers, programmers, instructional designers, and institutional researchers). Each has different interests in collecting data with their particular goals, with consequent privacy and ethical issues concerning the ownership of data and users' consent, etc. What binds them is the fact that everyone is required to have more than traditional digital skills.

Below we propose a scenario in a wide variety of potential cases contained in the 2018 edition of the EU Handbook of privacy with the type of question on which this paper aims to focus:

> "A university research department conducts an experiment analysing changes of mood on 50 subjects. These are required to register in an electronic file their thoughts every hour, at a given time. The 50 persons gave their consent for this particular project, and this specific use of the data by the university. The research department soon discovers that electronically logging thoughts would be very useful for another project focused on mental health, under the coordination of another team in another university." (p. 119)

The first questions that researchers must ask are:

- What must be written in formal consent before data can be collected and/or analysed?
- Do students have the option to "opt out" from the analytics project?
- Is a new student formal consent needed for sharing data with a new team?

This is a short example of how data analytics generates a privacy debate in common situations for Higher Education Institutions.

Our goal is to contribute to this debate by providing suggestions adapted to the General Data Protection Regulation (GDPR). With this aim, we start with a literature review on Learning Analytics, privacy, and ethical issues. We continue our discourse with a focus on the Italian context in these fields. The paper concludes with open questions on privacy policies and learning analytics linking three overlapping categories:

1. management of personal data;
2. limits of teachers' use of students' learning data;
3. students' right of transparency.

2 Related Works

In the debate during the past 10 years on privacy and ethics in education before the GDPR, several authors and pioneers in distance education referred to privacy and ethical issues as parts of a Learning Analytics system (Hoel *et al.*, 2017; Ferguson *et al.*, 2016; Drachsler & Greller, 2016; Slade & Prinsloo, 2013; Pardo & Siemens, 2014; Drachsler *et al.*, 2015; Campbell *et al.*, 2007) that has not yet been influenced by another major debate on privacy coming out of the GDPR. Each one faces the topic from a different point of view and with a different series of principle analyses.

Pardo and Siemens (2014) define ethics in a digital context as "the systematization of correct and incorrect behaviour in virtual spaces according to all stakeholders" (p. 439). In the same context, the concept aligns with the definition of privacy formulated by Drachsler and Geller (2016): "a living

concept made out of continuous personal boundary negotiations with the surrounding ethical environment" (p. 91).

In the Learning Analytics scenario, giving a possible reply is not straightforward. As Ferguson and colleagues note in their work, "the ethical and privacy aspects of learning analytics are varied, and they shift as the use of data reveals information that could not be accessed in the past" (2016, p. 5). What is sure is that Higher Education institutions have an obligation to protect students' data on the institutional platform and to inform them of possible risks when research data are sent outside the boundaries of national jurisdiction.

Teachers (and institutions) must know the current responsibility they have if they want to use data collected from students for a specific purpose. At the same time, designers are encouraged to include privacy and security issues in the early stages of their work and to comply with the requirements from both technological and legal aspects (Pardo & Siemens, 2014, p. 444).

Higher education institutions have always collected and analysed data of students in class through assessments and questionnaire. What has changed today is the volume of data that continues to rise along with tools' digital development, the diffuse use of Learning Management Systems (LMS), and the increasing need for the exploitation of data for educational goals (predictive learning, cases of special student, significant learning).

Quantity changes the methods and approaches that we use to interact with students and their data (Siemens & Long, 2011, p. 32) in ways that were not possible in the past without current technology. Google's Mayer (2010) suggested three "S"s (Ivi, p. 33). We propose something similar in an education context:

1. *Speed:* increasing available data in real time. Download speed from LMS allows for a larger range of research in a shorter time span.
2. *Scale:* increase in computing power. The diffusion of digital competence for both teachers and students produces data interaction and types of collaboration that change didactically, predicting the success of students and proposing new methodologies.
3. *Sensors:* new types of data. The information on student learning (analysing discussion messages posted, time spent watching videos, assignments completed, interaction with peers, etc.) serves the purpose of situated teaching and predictive modelling.

As a result of these new possibilities, there are a growing number of ethical issues regarding the collection and analyses of educational data. In this scenario, students (and society in general) are in a delicate situation in which the exchange of personal data is normal, but a balance between control and limits needs to be achieved yet (Pardo & Siemens, 2014, p. 440).

Generally, students know of the growing prevalence of data mining to monitor their behaviour on social media and shopping, but they might not be equally aware of when this occurs within an educational environment. In any case, they should be able to feel safe when they study and learn through an online system for distance education.

3 The Italian Path

"The data that is collected and analysed may be protected by federal, state, and institutional privacy regulations" (Campbell, 2007, p. 8).

As Hoel and colleagues said in their work "The Influence of Data Protection and Privacy Frameworks on the Design of Learning Analytics System" (2017), national data protection acts influence LA tools and systems. They propose an international study in a different context (OECD, APAC, and European GDPR) aimed to design a general privacy framework related to privacy processes and pedagogical LA requirements.

In Italy, the path to digitalization of services began in 2005 with the publication of the Digital Administration Code (CAD), a text that laid the foundations for the digitalization of the Italian public administration. Numerous revisions of the code have taken place over the years, most recently in December 2017. In it, the Agency for Italian Digitalise (AgID) introduced (Annex B) the minimum-security measures for public administrations to better protect the archival heritage and digital data of them. In addition, the AgID has published the three-year plan for the ICT of the Public Administration (PA), indicating the rules for a coherent development of systems.

Regarding the management of data, the Public Administration referred to the Legislative Decree 196/2003 before the GDPR. The privacy code, text that appears original for the period in which it is issued, implements its sanction based on the minimum-security measures contained in an annex.

The annex contains a list of minimum-security measures that all controllers or processors (regardless of size), features, and peculiarities of represented institutions must comply with. However, technological evolution has demonstrated the ineffectiveness of this system, in particular from two points of view: 1) to have imposed equal measures for all, which therefore does not take into account the characteristics of each controller or processor (e.g., the 8-character password for both a small company and a large hospital); and 2) these measures are not in line with the times because they were based on the technological contest of 2004 that has now completely changed (e.g., at that time, a 5-character password could be violated in one minute. After only four years, the time to violate a 6-character password dropped to 0.0224 seconds) (Re Garbagnati, 2012). As a result of this, the GDPR is no longer based on

standard measures but on a self-assessment of the owner in accordance with the accountability principle.

The GDPR was issued on 25 May 2016 as a general regulation on the protection of personal data. It has the aim of standardizing all national legislation and adapting to new needs based on the evolution of the technological context. Directly applicable in all member states (as an "self executive" Regulation), each national legislator has had a deadline to adapt to the new European legislation. On 25 May 2018, the two-year time limit expired, and the GDPR began producing its effects concretely.

3.1 What is going on in the Italian universities?

This discourse has emerged in Italian University debates only in recent years, in particular with the National Plan for Digital University edited by the Rectors Conference of Italian University (CRUI). During the Rectors meeting in Udine (2018), they organized a series of "work tables" in which professors and sector experts debated on a specific topic regarding distance education including digital environments for the innovation of teaching, technology and cybersecurity, MOOCs, and so on. This event suggests that university institutions finally understand the critical value of education technology (Siemens & Long, 2011, p.33) in addition to the monetary value (Margoni, 2007). The different scope that exists between *learning analytics* and *academic analytics* affirms this trend[1].

Moreover, the CRUI has turned on a GDPR regulation web page on the national territory aimed at monitoring the development and speed of the digital transition process. The survey involved 60 universities. In most cases (44.29%), the appointment of the Data Protection Officer (DPO) is still ongoing and little more than the Digital Transition Manager (55, 71%). In at least half of the cases (45.71%), the minimum-security measures proposed by the AgID (Agency for Digital Italy) were not activated, nor was the software implemented for the management of the registers related to the processing of personal data (48.57%)[2].

It is evident that Italian universities are in delay with this accountability process. The result is that academic staff could be in trouble without a clear idea of the new regulation and aspects linked to research.

With the purpose to clarify some new approaches and based on research evidence of existing research on Learning Analytics and privacy issues, this paper adds an integrated overview of European GDPR principles from an online

[1] The Academic Analytics focus on political/economically challenge on the potential to create actionable intelligence to improve teaching, learning, and student success. In the first time they were corporate in education sector as "business intelligence".

[2] Data update in July 2019. Direct: http://bit.ly/gdpr-crui

education perspective.

4 The Need for the GDPR in Online Education

Universities are increasingly damaged by cyber-attacks (Cisternino *et al.*, 2018, p. 2). This represents a danger not only for the intellectual property of the contents present in the university databases but also for the personal data recorded about the numerous persons that work in the academic context (students, teachers, administrative staff). Added to this is the sanctioning provide of GDPR that can concern institution that failure to provide high levels of data protection.

The GDPR has as its principal goal to protect the personal data of the subjects (i.e., the natural person to whom those data belong) up to a coherent engineering of the data management system that avoids the need to protect them afterward. With the evolution of digital education, in the field of public infrastructures, the process of transition and review of the functioning of an organization through ICT services and digital management (Sperduti *et al.*, 2018, p. 2) must fall within the current academic policies.

In this scenario, which continues to experience rapid technological evolution (but not as rapidly evolving legislation), the universities are now facing the heavy task of updating administration, guaranteeing compliance, and innovation of services, even with respect to performing consistently toward quality assurance.

In the educational environment, this particularly concerns distance learning issues and those regarding the work done on e-learning platforms in the current need to respect the new legislative provisions.

Whenever a user interfaces with the platform, be it a teacher or a learner, it variously transfers his personal data through multiple actions: registering, uploading a course, sending requests, managing the materials, following a course, and providing its access and use data. On the other hand, technology makes education more personal, and it empowers academic staff and students to make better decisions (Oblinger, 2012).

The data flow in the e-learning platform can be long and complicated to manage. We propose a typical one in Figure 1.

In this flow, the personal data of students, teachers, external users, and operators must always be acquired, managed, processed, and preserved following the principles sanctioned by the GDPR. The distance education distributed by e-learning platforms, as well as the whole traditional didactic context, can no longer avoid it. This leads to the emergence of new "privacy" problems, the use of LMS helps researchers in the design and delivery of

learner-centred courses, and students have greater access to more flexible options for engaging with peers and instructors (Macfadyen & Dawson, 2010, p. 589; Drachsler & Geller, 2016). However, every decision in the flow must be made with sufficient information to reduce risk.

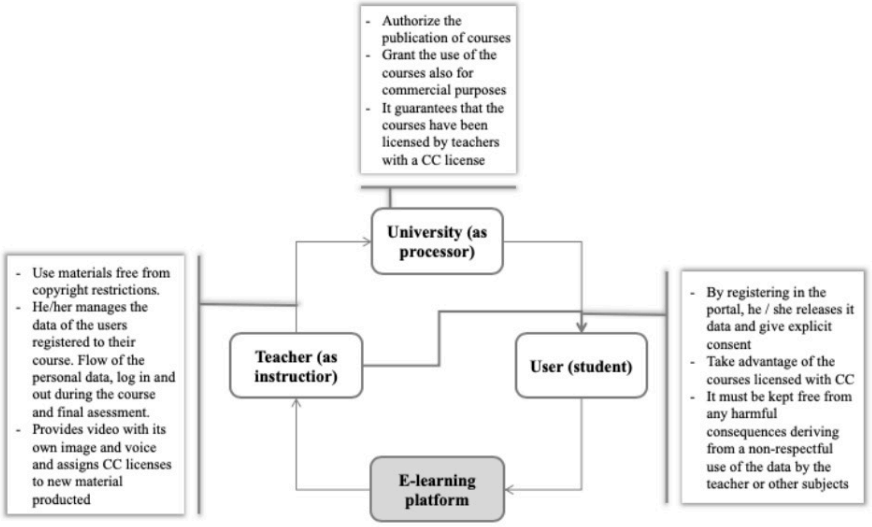

Fig. 1 - *Data flow in e-learning systems.*

To fully understand the scope and the novelties of the new regulation, we must clarify the meaning of "personal data":

> "Personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." (Art.4, GDPR)

It is also important to understand the difference between types of personal data: name and surname, fiscal code, address, e-mail address, telephone number (common data); data revealing racial and ethnic origins, religious beliefs (particular data); data revealing of the quality of a suspect or accused person (judicial data).

The controller or processor who manages the e-learning platform protects personal data in the simplest and most stringent way possible, first by following

and respecting the fundamental principles contained in article 5 of the GDPR: lawfulness, transparency, relevance, accuracy, conservation and security, in order to achieve privacy in compliance with the new European regulation.

What establishes the limit of what teachers, Instructional Designers, and Higher Education institutions in general can do with students' learning data depends on which data will be used.

In the process of acquiring personal data, as in the LA survey, it is always necessary to request and subsequently retain the consent of students. Researchers must guarantee the right to the revocation of consent and map the databases to respond promptly and adequately to any requests received. The databases like the servers or Learning Management System must be adequately protected. Exposure to data breach could result in serious losses of trust in the users.

A big challenge for Learning Analytics in this respect is the complexity of the data collection (Drachsler & Greller, 2016) and variety of use for the researchers. The best method to manage data in online education is to be clear and available when explaining the purpose of data collection and to do it in compliance with the existing legal frameworks (Ivi, p.95).

Going beyond these terms, "using analytics requires that we think carefully about what we need to know and what data is most likely to tell us what we need to know" (Siemens & Long, 2011).

Conclusions and future perspectives

In the Learning Analytics process, selected questions, quality data, sound practices, and prudent processes mitigate risks (Oblinger, 2012) that are inherent in making any decisions.

The GDPR introduced a fundamental revolution with respect to the past: the principle of accountability, privacy by design and by default. Those require constant and continuous self-assessment and previous knowledge of the treatment to be activated.

The GDPR today seems to be the more complete law with respect to digital transformation and consequent needs within various sectors (Hoel *et al.*, 2017, p. 3) for these two principles. With the GDPR, we have gone from the typical Italian logic, which permeated the whole regulatory framework of the privacy code, to a more Anglo-Saxon logic oriented to self-assessment. This is the real novelty of the European regulation — no longer a detailed set of rules "dropped from the top" but a series of principles to which the processor must adapt, evaluate, and implement.

Once we overcome the logic of static and equal security measures for all and a dynamic and flexible approach has been inaugurated, a fundamental question

emerges: what if I have to know, and what is the limit to manage the personal data in a research project?

A checklist of questions can help teachers who intend to proceed with a treatment to simply and quickly assess the intrinsic criticalities of the treatment itself.

Drachsler and Geller proposed a checklist called "DELICATE" (2016) in order to reach the goal of providing a largely self-explanatory practical tool for designing Learning Analytics surveys within any data-driven educational organisation.

We propose below a short set related to three categories, as indicated in the introduction:

Management of personal data	• What should I do? • Have I carefully evaluated the operations to be undertaken with personal data? • Under which conditions do I want to use these data? • Were the personal data of the interested party (or third party) collected legitimately? • Are there special categories of personal data be processed?
Limit of teachers' use of students' learning data	• To whom will I send the data? • Are there other subjects (colleagues) to whom I will send the data? • Have the purposes of my processing been clearly defined? • Are there purposes that require special additional information (e.g., research, industrialization, data transfer)?
Students' right of transparency	• Have I decided the measures through which individuals' identity will be protected? • Can my processing in any way compromise the interests or fundamental rights and freedoms of the data subjects? • Can data subjects access the data if they wish? • Are there any obstacles to guaranteeing the subjects' right to rectify and/or delete data or to oppose their processing and their portability?

Based on these three macro-areas, it will be possible to draw up other questions that are common for the activities of Learning Analytics with the critical issues that will need to be resolved before starting the research.

The final part of the case proposed in the Handbook of Privacy is:

> "Even though the university, as controller, could have used the same data for the work of another team without further steps to ensure lawfulness of processing that data, given that the purposes are compatible, the university informed the

> subjects and asked for new consent, following its research ethics code and the principle of fair processing." (p. 119)

This case demonstrates how sensitive the issue of privacy use on Learning Analytics is, particularly when dealing with other groups. This practical example could also help to understand the difference between the previous legislation on privacy and the GDPR that we can synthesize, saying everything we need to know and do for privacy compliance and what must be done before every decision (privacy by default) for the principle of "prevent, not correct." The university, as a data controller, puts in place appropriate measures to ensure that only the personal data necessary for each specific purpose of the processing are treated by default.

Reflecting on future directions for this research, we aim to analyse more cases related to Learning Analytics and data management in distance education, to create a comprehensive framework to address all types of data used in possible scenarios and propose a grid and parametres to respond, with an original point of view, to the questions presented in the table.

What we can share now is the consciousness of the need to update university privacy policy, in line with new content introduced by the GDPR, ensuring an effective governance and data management in every work sector, from research to administrative issues.

For teachers, it is important to establish research goals, taking care to associate the relative legal basis of each purpose that makes the processing legitimate.

Numerous questions exist around Learning Analytics, privacy, and ethical issues, so it is important to have full knowledge of the present state of things to be sure of the future. The proposed considerations and questions in this paper provide practical support for higher education actors to clarify the Italian legal context and to increase the quality and effectiveness of Learning Analytics.

REFERENCES

1st International Conference on Learning Analytics and Knowledge. Banff, Alberta: Canada. February 27–March 1, 2011. Retrived from <https://tekri athabascau.ca/analytics/>.

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

Cisternino, A., Baldi, M., Longhi, S., Paganoni, M., Ruggieri, F. (2018), Infrastrutture tecnologiche and cybersecurity. In *I Magnifici incontri CRUI*, *Tavolo 2B*. (in italian) Retrived from http://www2.crui.it/crui/magnifici_incontri_crui_2018/Tav2B%20-%20Infrastrutture%20tecnologiche%20e%20cybersecurity.pdf

Drachsler, H. & Greller, W. (2016), 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). ACM.

Drachsler, H., Cooper, A., Hoel, T., Ferguson, R., Berg, A., Scheffel, M., Kismih´ok, G., Manderveld, J. and Chen, W. (2015), Ethical and privacy issues in the application of learning analytics. In *5th International Learning Analytics & Knowledge Conference (LAK15): Scaling Up: Big Data to BigImpact*, 16-20 Mar 2015. Poughkeepsie, NY: USA.

European Union Agency for Fundamental Rights and Council of Europe, (2018), Handbook on European data protection law. Retrived from https://fra.europa.eu/en/publication/2018/handbook-european-data-protection-law

European Union (2016), General Data Protection Regulation. Retrived from https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:32016R0679

Ferguson, R., Hoel, T., Scheffel, M., Drachsler, H. (2016), *Guest editorial: Ethics and General Data Protection Regulation* (2016). Retrived from https://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4912/5426

Hoel, T., Griffiths, D., & Chen, W. (2017), The influence of data protection and privacy frameworks on the design of learning analytics systems. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 243-252). ACM.

Macfadyen, L. P., & Dawson, S. (2010), Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2), 588-599.

Margoni, T. (2007), E-learning, corsi online and diritto d'autore. *Diritto dell'internet*, 6 (in Italian).

Mayer, M. (2010), *Innovation at Google: The Physics of Data*. PARC Forum, retrived from <http://www.slideshare.net/ PARCInc/innovation-at-google-the-physics-ofdata>.

Oblinger, D. (2012), *Analytics: What we're hearing?* Retrived from https://er.educause.edu/articles/2012/11/analytics-what-were-hearing

Pardo, A., & Siemens, G. (2014), Ethical and privacy principles for learning analytics. *British Journal of Educational Technology,* 45(3), 438-450.

Re Garbagnati, E. (2012), *Quanto impiega un hacker a bucarvi la password?* (in italian) Retrived from http://www.ictbusiness.it/cont/news/quanto-impiega-un-hacker-a-bucarvi-le-password/29170/1.html#.XSxAjZMzZN1

Siemens, G., & Long, P. (2011), Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.

Slade, S. & Prinsloo, P. (2013), Learning analytics: ethical issues and dilemmas. *American Behavioral Scientist,* 57(10) pp. 1509–1528.

Sperduti, A., Vannozzi, D., Mingarelli, D., Micheloni, C. (2018), Mappa delle piattaforme and loro inter-operabilità. In *I Magnifici incontri CRUI*, *Tavolo 2A*. (in italian) Retrived from http://www2.crui.it/crui/magnifici_incontri_crui_2018/Tav2A%20-%20Mappa%20delle%20piattaforme%20e%20loro%20inter-operabilit%C3%A0.pdf