

## ARTICLE SUMMARY

### Learning analytics and the processing of student data in universities

Juuso Ouli and Tomi Voutilainen

Full article: Edilex 2019/37, Edilex 30 September 2019, [www.edilex.fi/artikkelit/20064](http://www.edilex.fi/artikkelit/20064)

---

#### 1 Background and goals of the study

AnalyticsAI is a project funded by the Finnish Ministry of Education and Culture, coordinated by the University of Oulu and involving seven universities. The aim of the project is to facilitate students' studies and transition to working life by using learning analytics to support the planning and monitoring of studies, teaching and student guidance and university governance. It collects and defines needs for learning analytics, as well as develops and pilots digital applications and practices that support studying, teaching, guidance and governance. As part of the project, Tomi Voutilainen, Professor of Information and Information Technology Law at the University of Eastern Finland Law School, and Juuso Ouli, Research Assistant, conducted a study and wrote an article on learning analytics and the processing of student data in universities. In this article, they analyse issues related to the legal bases of the personal data processing relevant for the project and investigate the legal requirements imposed upon personal data processing and the functionalities of information systems. They also examine the possibilities and restrictions imposed by the legislation and guidelines based on it on the use of student data in the development of learning analytics, the data analysis of student data, the development of teaching technology and the use of artificial intelligence. In the article, these issues are studied from students' perspective as well as that of the legal protection of teachers and other university personnel with regard to how they are allowed to process student data. The article mainly focuses on the regulation of data protection but also introduces other legally relevant perspectives.

#### 2 The data processed in learning analytics

Various types of data can be used in learning analytics. In addition to students' identifiers, it is possible to use data concerning the progress and planning of their studies and the setting of personal goals. From a legal perspective, it is essential to distinguish between the following categories of data based on the personal data protection legislation: data other than personal data, 'ordinary' personal data, specific categories of personal data and unique identification numbers. Since the provisions of the EU's General Data Protection Regulation (GDPR) only apply to the processing of personal data, it is essential to identify anonymous data and distinguish them from personal data. Anonymity must be ensured effectively. The authors of the article point out that making the data necessary for identification unavailable to the holder of pseudonymised data is not sufficient to make data processing anonymous. They recommend that to make data processing anonymous, all possible data that can be used or combined to enable identification should be destroyed.

Furthermore, they also recommend that developers of learning analytics should thoroughly evaluate which of the data to be used in learning analytics are truly anonymous and which are in fact pseudonymised or other types of personal data to which the GDPR applies. In the event of any

ambiguities, it is recommendable to comply with the requirements of the GDPR. However, in line with the risk-based approach taken by the GDPR, it is possible to apply lighter safeguards in borderline situations in which it is very difficult to identify a person. In learning analytics, attention must be paid to confidential data as well as the provisions of the Act on the Openness of Government Activities on the form and method of disclosing data when disclosing data from a filing system of a public authority. Attention must also be paid to data belonging to specific categories of personal data, such as data concerning health. Data concerning health include data about students' life habits, for example. It is also necessary to take into account the regulation related to identification numbers, or students' student or learner numbers.

### **3 Responsible bodies and obligations in the use of learning analytics**

In the article, the authors discuss the regulation concerning the responsible bodies under the GDPR, i.e. the controller, the processor and the joint controller. As education and guidance leading to a degree are universities' legal duties, universities determine the purposes and means of the personal data processing related to the implementation of these duties and thus act as controllers. Universities may also act as joint controllers if they jointly determine the purposes and means of personal data processing. The actual processors are often service providers to whom the personal data processing, such as storage and analysis, is contracted out. In such contracting out, attention must also be paid to data protection issues; for example, to the provisions concerning the contract between a processor and a controller in Article 28 of the GDPR. The authors recommend that universities' data protection officers should thoroughly acquaint themselves with the functions of learning analytics because it may require the introduction of a new technology or an impact assessment. They note that to demonstrate their accountability, controllers may use the codes of conduct referred to in Article 40 of the GDPR in learning analytics. These codes of conduct should be prepared jointly by universities and other institutions of higher education. The authors discuss in detail the regulation of risk assessment, impact assessment and prior consultation from the perspective of learning analytics.

### **4 Principles of data protection in the use of learning analytics**

In this chapter, the authors discuss the principles relating to personal data processing set out in Article 5 of the GDPR in the context of learning analytics. First, they examine the legal bases for the processing of personal data. In the use of learning analytics, the most important and primary legal basis for personal data processing is the necessity of complying with universities' legal obligation (the provision of teaching, which also involves the guidance of students). Depending on its purpose, the development of universities' activities may also be regarded as their legal duty. Universities' tasks that are carried out in the public interest include tasks related to research and the compilation of statistics. A legitimate interest can be used as a legal basis for processing, for example, in the case of development of activities that is not based on universities' legal duties. A legitimate interest may also provide a basis for offering working life-related guidance, for example, which is not directly related to universities' duty to provide education that leads to a degree. An example of contract-based processing is the processing of a university library's customer data or the provision of commissioned training. Universities should adopt a

negative view of consent-based personal data processing. The authors argue that if the processing concerns data belonging to specific categories of personal data, such as data concerning health, there is no legal basis for processing the data for the purposes of learning analytics apart from scientific research.

With regard to the requirement of fairness, it must be noted that no data should be collected for learning analytics that are not otherwise necessary for the implementation and evaluation of students' study attainments. The authors recommend that to fulfil the principle of transparency and the obligation to inform, the necessary information should be provided in the digital services in which data are collected from students themselves. If the processing involves profiling or automated decision-making, it may be necessary to explicitly explain the logic of processing and reasoning to fulfil the requirements of transparency. The purpose of using personal data in learning analytics must be specified sufficiently so that students and members of personnel who process the data truly understand the reason and need of the data processing and for what purpose the data can be used. If a new purpose is planned based on learning analytics data, it must be evaluated according to the GDPR.

No unnecessary data must be collected for learning analytics; there must be an appropriate basis for all data in relation to the purpose for which they are used and to the controller's activities. When introducing learning analytics, it must be ensured that time limits for erasing students' personal data are planned in advance. The requirement of accuracy also applies to learning analytics and the data produced by it. When personal data are no longer needed, they must be erased. If the data are needed, but it is no longer necessary to identify the data subject to fulfil the purpose of the processing, the data must be anonymised. With regard to the requirement of integrity and confidentiality, it must be noted that the implementation of technical and organisational measures is provided for in more detail by the Act on Information Management in Public Administration, which applies to information management, including personal data processing, in universities.

## **5 Profiling and automated decision-making**

In this chapter, the authors discuss in detail the definition of profiling in Article 4(4) of the GDPR and its meaning in the context of learning analytics. In the case of learning analytics, profiling means that the analytics automatically produces information about each student's personal aspects. These types of data may be useful for teachers, for example, who can use them when evaluating or guiding students or developing their teaching. These data may also be useful for students themselves to develop their learning. In principle, profiling is not forbidden. It should comply with the general requirements concerning the legal bases, purpose limitation, data minimisation, transparency and the rights of the data subject set out in relation to processing in the GDPR. The authors also further explain these requirements in the context of learning analytics.

In addition, the chapter introduces a definition of automated decision-making, which is indirectly implied in Article 22(1) of the GDPR. The authors also examine issues related to the application of Article 22(1) of the GDPR in the context of learning analytics. In order to fall

into the category of automated decision in the GDPR, such decisions should be based solely on automated processing and such decision-making should produce legal effects or similarly significant effects concerning the data subject. They suggest that in the university environment, these legal effects may be produced by decisions concerning the admission of students, granting of degrees, evaluation of study attainments and admission to study units, for example. The authors put forward that automated evaluation involves very complex legal issues. In principle, automated decision-making is forbidden, and use of other automated processes in student evaluation should be regarded with great caution due to the higher data protection requirements. As yet, there is no national or Union legislation that would allow for automated decision-making within learning analytics, which means that learning analytics cannot be used for automated individual decision-making pursuant to the GDPR in administrative matters.

## **6 The legal protection of students**

This chapter shortly describes the regulation related to the legal protection of students. Ensuring the legal protection of the data subject when personal data are processed is part of the fulfilment of data protection. The rights and obligations of university students are determined in general administrative acts as well as the Universities Act applied as a special act. Legal protection is a fundamental right that applies to the administrative decisions made in universities as well as their actual administrative activities, such as teaching and the evaluation of study attainments. In the use of learning analytics, too, attention must be paid to the requirements concerning the legal protection of university students. The chapter describes in detail the legal principles of administration and their meaning in learning analytics. It also describes the legislation relating to liability for acts in office and how it is determined in the use of learning analytics.

## **7 Information management**

In this chapter, the authors introduce central issues concerning the application of the Act on Information Management in Public Administration ('Public Information Management Act') in relation to universities and their activities. In Section 3(1) of the Act, universities are equated with authorities, and thus the provisions of the Act are also applied to their activities. However, the provisions in Chapter 3 of the Act are not applied to universities. The subject matter of the Public Information Management Act is tied to the concept of official document provided for by the Act on the Openness of Government Activities. Thus, the Public Information Management Act will not be applied if the material to be processed does not fall within the scope of the concept of official document. In principle, the student data contained in student information systems and various learning platforms are documents submitted to an authority or prepared by an authority. Similarly, documents produced by a learning analytics application or other similar data can in principle be regarded as documents prepared by an authority. The introduction of learning analytics in universities entails a change in the way information is managed in the processing of student data, and this change requires an impact assessment conducted as referred to in the Public Information Management Act. The introduction of learning analytics means that a new technology is introduced, which also requires an impact assessment referred to in Article 35 of the GDPR. These assessments referred to in the GDPR and the Public Information Management Act can be combined into one overall assessment. In this chapter, the authors also

discuss the legal bases of the protection of student data from the perspectives of the GDPR, the Data Protection Act and the Public Information Management Act.

## **8 Conclusions**

In the final chapter of the article, the authors present their conclusions. In the use of learning analytics, it is important to identify the purpose determined for the data collection before any data is collected, because the purpose determines the possibilities of using the data. The purpose of research on learning analytics and the data produced by it must be distinguished from the purpose of using learning analytics to fulfil universities' legal obligations, such as the provision of teaching and guidance. Universities cannot develop learning analytics solely on the basis of the data protection legislation, because their activities are also governed, within the national margin of manoeuvre, by the general administrative laws, of which the most important are the Administrative Procedure Act, the Act on the Openness of Government Activities and the Public Information Management Act as well as the Universities Act as a special act. If universities wish to use learning analytics data for activities that are not part of their legal duties, the data processing must, in practice, be based on either statistical purposes or the universities' legitimate interest.

One aspect that must be considered in relation to learning analytics data is their storage period. Data cannot be stored forever for the purposes of potential future research, new uses or analytics tools to be developed in the future, for example, unless the National Archives of Finland has decreed that the data will be archived. In principle, learning analytics functions that include student profiling entail a higher risk, but if implemented correctly, they are functionalities that can be included in learning analytics. In situations in which universities exercise official authority – for example, when evaluating study attainments or deciding on student admissions – it is not possible to use automated decision-making without the support of special legislation, which does not presently exist in Finland. When learning analytics is used, attention must also be paid to the legal protection of students, which in the case of learning analytics means, above all, that the quality of the data must be ensured. Often the final responsibility lies with the teacher or other member of university personnel who made the decision based on learning analytics.