Learning Analytics Icons: Easy Comprehension of Data Treatment

Daniel Amo-Filva[1]*, Marc Alier[2], David Fonseca[1], Francisco José Garcia-Peñalvo[3], María José Casañ[3]

[1] La Salle, Ramon Llull University, Barcelona (Spain)
[2] Polytechnical University of Catalonia, Barcelona (Spain)
[3] University of Salamanca, Salamanca (Spain)

* Corresponding author: daniel.amo@salle.url.edu

Received 4 March 2022 | Accepted 21 April 2023 | Published 8 April 2024

Abstract

The Learning Analytics approach adopted in education implies the gathering and processing of sensitive information and the generation of student profiles, which may have direct or indirect dire consequences for the students. The Educational institutions must manage this data processing according to the General Data Protection Regulation, respecting its principles of fairness when it comes to information gathering and processing. This implies that the students must be well informed and give explicit consent before their information is gathered and processed. The GDPR propose the usage of recognizable standardized icons to facilitate a general understanding and awareness of how personal data is deemed to be processed in each application context, like an online course. This paper presents a project that aims to provide a set of icons to inform about the treatment of educational data in the Learning Analytics processes and a survey about the student's comprehension of the icons, their meaning, and implications for their privacy and confidentiality. The result presented is a set of icons ready to be integrated into educational environments that apply Learning Analytics to increase transparency and facilitate the understanding of data processing.

Keywords

Data Treatment, Education, GDPR, Icons, Learning Analytics, Privacy, User Profiling.

DOI: 10.9781/ijimai.2024.04.001

I. Introduction

Data scientists can analyze educational data from different perspectives. On the one hand, educational data can be processed with the unique objective of extracting and discovering behavioral patterns. This process is called Data Mining [1]. On the other hand, educational data can be treated with the ultimate purpose of improving any aspect of the teaching-learning methodology. This process is called Learning Analytics [2], [3]. Therefore, Learning Analytics is an analytical approach that collects, analyzes, and visualizes student data to improve the educational context. Reasons for improvement are the processes of tutoring, evaluation, or even student follow-up [4]. This paper focuses on the possibilities offered by Learning Analytics in the educational community, focusing on the ability of institutions to be transparent and able to fulfill the various challenges it presents, especially those related to the treatment and privacy of students' [personal] data. It is worth going back to the beginnings of Learning Analytics to understand the need for transparency and trust that institutions must convey to students.

It was George Siemens [5], who in 2010 took the first steps in this approach, giving its current name and creating the first discussion groups in Google Groups to reflect on the state of the art and possibilities of educational data analysis. Over time, Siemens' extensive dissemination task managed to transcend the term to the education and scientific communities. Currently, Learning Analytics is a field considered by the scientific community to be of high interest, where a large volume of scientific contributions from different authors from around the world are published, unraveling both its underlying model [2] and its opportunities [6] and even weaknesses [7].

Regarding the education community, the use of data for decision-making is gaining adoption at all levels: those related to teachers-students, institutional and inter-institutional [8]. Learning Analytics' origins focused on supporting Massive Online Open Courses (MOOC) type courses [9], [10]. Eventually, its applications have been adapted to other educational contexts and needs. The first utility of Learning Analytics related to MOOCs and Virtual Learning Environments (VLE) was teaching support and dropout rate diminishing [11], [12]. These MOOC courses have low teaching staff and high enrollments, reaching hundreds of thousands of participants in some cases (hence they are called massive courses). Thereon, its uses in VLEs evolved from reducing dropouts [13] to meet other needs such as improving teaching methodologies, student well-being, or even shaping learning spaces. Over time, the adoption of Learning Analytics has bounced from virtual platforms to physical environments. Multimodal Learning Analytics [14] is the branch of Learning Analytics dedicated to analyzing student behavior in face-to-face context through [connected] sensors.

Please cite this article in press as:
D. Amo-Filva, M. Alier, D. Fonseca, F. J. Garcia-Peñalvo, M. J. Casañ. Learning Analytics Icons: Easy Comprehension of Data Treatment, International Journal of Interactive Multimedia and Artificial Intelligence, vol. 9, no. 3, pp. 115-126, 2025, http://dx.doi.org/10.9781/ijimai.2024.04.001

In any case, the [virtual and face-to-face] learning processes mediated by Learning Analytics collect data of all possible student interactions and academic performance, both treated as students' behaviors [15]. At the same time, data collection goes beyond what is strictly educational or academic. This additional data, considered as metadata, are complementary and may originate from heterogeneous sources such as social networks or financial data. All this data and metadata collection generate a sensitive context making data fragile in all senses [16]. Consequently, concerns arise regarding confidentiality, privacy, and security of the students' data and, in it extends, about their digital identity [17].

Despite this contextual sensitivity, the adoption of Learning Analytics in the educational context has increased due to different factors, including:

- The rapid transition from classrooms to hyperconnected classrooms.
- Deep classroom integration of connected learning devices.
- Digitization of teaching-learning materials and processes.
- The incorporation of third-party educational technology in the form of apps in the cloud.
- The rapid evolution of educational technologies based on Big Data, Artificial Intelligence, and Machine Learning as facilitators of teaching-learning processes.
- The use of cloud computing to reduce IT infrastructure costs.

However, one must not be enlightened by the rapid evolution of technologies and their promises. With the pandemic and the confinement caused by COVID-19 [18], [19], interest in Learning Analytics has increased in all educational stages and worldwide [20]. At the beginning of the pandemic, the collection and processing of educational data made it possible, in the first instance, to understand how students interacted with VLEs, and in the second instance, to give them the appropriate and necessary support. All this data collected in pre-pandemic, pandemic, and post-pandemic is stored, analyzed, and even shared between institutions and countries. Such data treatment is regulated by the General Data Protection Regulation (GDPR) [21] and other data protection laws of each EU member state. These laws exist since it is necessary to regulate any data processing to avoid improper use. Hence, the educational context must enforce these laws, such as transmitting certain aspects to students as decision-making information, even before registering for any course.

The adoption of Learning Analytics can negatively impact the confidentiality [22], privacy [23], and security [24] of student data, as well as their digital identity [25]. In the worst case, students do not realize it until it is too late, their data being misused [26], shared with third parties [27], leaked [28], or used by algorithms [29] with dire results to the students themselves. Different authors have pointed to this type of problem [17], [30]–[32]. Local technologies have even been proposed, substituting or complementary to cloud computing, to ensure that this environment of mistrust reverts to one of absolute certainty of a secure data environment [16].

In any case, whether using local or cloud technologies, the use of Learning Analytics implies a great responsibility regarding the collection, storage, treatment, and sharing of educational data, especially when the data is from minors. For all the above reasons, strict law enforcement is necessary as a tool for preserving data privacy.

A. General Data Protection Regulation

In 2016, the GDPR was approved, however, its entry into force was not scheduled until two years later, specifically, on May 28, 2018. The GDPR establishes the obligations that contract the entities that process and manage personal data, those that by themselves can identify a person. At the same time, it defines five fundamental rights of citizens before such entities: the right to know, the right to request the data controller, the right to rectify data, the right to delete data, and the right to oppose data processing [33]. These rights allow any person to suspend the data processing, facilitate data portability to third parties, revoke the consent given, or even oppose automatic processing.

The study tackles the right to be informed, which includes other rights such as knowing: the purposes of data use, the period of data conservation, and even if there are automated decisions or profiling. Recital 60 [34] of the GDPR establishes that the interested party must be informed:

- "...of the existence of the processing operation and its purposes".
- "...with any further information necessary to ensure fair and transparent processing".
- "...of the existence of profiling and the consequences of such profiling".
- "...whether he or she is obliged to provide the personal data and of the consequences, where he or she does not provide such data".

Considering the above and the educational context, fast, transparent, and easy to understand forms of information are required to:

- Raise awareness of current student data treatment processes in any educational context.
- Let the students know how their data is treated and for what purpose, preferably previous registration to any course.
- Establish a standardized information system and transmission medium in any educational context for obvious reasons.

B. Icons & Learning Analytics

Recital 60 of the GPDR informs about the possibility of using standardized icons to combine with textual information. The purpose of the complementary use of standardized icons is to give a clear, intelligible, and legible view of the intended processing. Besides, in point 7 of Article 12 of the GDPR, the possibility of an iconified representation of Articles 13 and 14 is exposed. These two articles consider all the information shown at Recital 60. Both Recital 60 and Article 12.7, require that electronically presented icons must be machine-readable, in other words, data structured in formats such as P3P, JSON, or XML should accompany the icons.

The use of the icons can fulfill the three points of the previous section. Icons can inform how the data is processed in Learning Analytics and, therefore, generate awareness to students by facilitating access to this type of information. Consequently, we propose as objectives of the study:

- Design descriptive icons of those parts of the Learning Analytics processes that must be reported to students, and that other authors have not designed in their work.
- Develop and make available a tool such as Creative Commons where any VLE administrator can create the appropriate icon packs to inform students of the different treatments of their data from the LMS itself.

The structure of this document is organized into four sections. All four sections show how our work makes available a set of icons regarding the data treatment in education, where methods such as Learning Analytics [or other kinds of as Academic Analytics or Educational Data Mining] are applied to facilitate its comprehension by students. Section I is the introduction. Section II gives the used methodology and fundamentals to design the icons. Section III exposes the results of the different design phases concerning questionnaires results. Section IV presents the conclusions.

II. Methods

The methodology of the study is mixed and of two phases whose purposes are adjusted to the objectives of the study.

A. Phase 1

We propose to start with a documentary methodology using a qualitative-quantitative approach. The purpose of this phase is to review and understand the work done by other authors regarding the iconifying of Articles 13 and 14, both in general and specifically for Learning Analytics. For this purpose, we designed a basic systematic literature review (SLR) [35], [36] using Web of Science and Scopus as database indexes. However, we begin with a mapping of the context before delving into the review itself. These are the mapping questions:

- MQ1. How many domain-related studies have been published in the last six years?
- MQ2. In which media have the articles been published?
- MQ3. Are there authors in common among the selected articles?

We aim to answer the following research question:

- RQ1. What kind of icons has been created related to GDPR?
- RQ2. What icons have been created related to analytical actions?

The answer to these questions will help understand how many Learning Analytics actions have been iconified and if there are any that remain to be defined.

We establish the following inclusion criteria for the search of manuscripts:

- (IC1) The results must be scientific publications.
- (IC2) The results must contain icons related to legal aspects defined in the GDPR.
- (IC3) The results can refer to icons related to Learning Analytics processes.
- (IC4) Results must be published before the enactment of the GDPR.
- (IC5) The language of the results must be in English or Spanish.
- (IC6) The results must be published in scientific conferences or journals without the need for impact.
- (IC7) The results must have been published through a peer-review process (double-blind).

We establish the following exclusion criteria for the search of manuscripts:

- (EC1) The results are not scientific publications.
- (EC2) The results do not contain icons related to legal aspects defined in the GDPR.
- (EC2) The language of the results is different to English and Spanish.
- (EC3) The results are not published in scientific conferences or journals (with or without impact).
- (EC4) The results are not published through a peer-review process (double-blind).

We define database-related search strings for each database index as:

- Web of Science: "GDPR icon*"
- Scopus: TITLE-ABS-KEY (gdpr AND icons)

We conduct the SLR regarding the PRISMA declaration . Thus, the flow of information through the different phases of the systematic review is shown in the flow diagram available in Fig. 1.

The search returns only 12 results. Only 6 are considered valid after removing duplicates and non-GDPR nor icons related.

In response to MQ1, six related articles have been published in the last six years, one in 2017, one in 2018, one in 2019, two in 2020, and one in 2021. In response to MQ2, two articles have been published in a scientific journal and 4 have been published in scientific conferences. In response to MQ3, the authors who publish these works are represented in Table I, with Rossi and Palmirani being the most active and repeated actors in the works found:

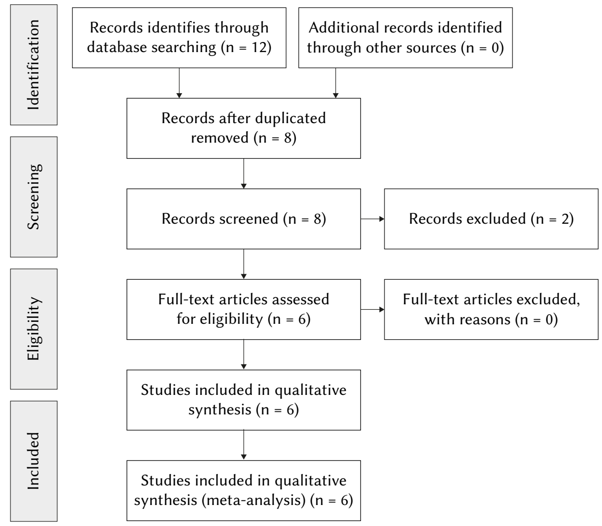


Fig. 1. PRISMA's flow diagram of the SLR conducted.

TABLE I. Authors

Author	Nº of papers found	Reference
Rossi, A.	5	[38]–[42]
Palmirani, M.	4	[38]–[40], [42]
Lenzini, G.	1	[41]
Martoni, M.	1	[40]
Hagan, M.	1	[40]
de Jong, S.	1	[43]
Spagnuelo, D.	1	[43]

Reading the six references allows answering RQ1 and RQ2 in terms of classification of icons related to the GPDR and Learning Analytics. From the scientific literature found, we extract that some authors have based their work on different proposals for iconified representations, including some before the enactment of the RGPD. Jong and Spagnuelo [43] classified icons into two main groups, dividing them into many subcategories:

- Data collection: personal data, sensitive data, sharing with third parties, data security, and data retention.
- Processing purposes: privacy settings, policy changes, legal obligations, user tracking, and profiling.

JJong and Spagnuelo are not the only ones to create icon taxonomies related to data privacy. Rosi and Palmirani [38]–[42] also classified icons but in multiple categories, being their taxonomy the most complete and exhaustive work among all the search results, which even considers proposals before the enactment of GDPR:

- Types of data: processed data, inferred data, etc.
- Functions of the agents: owner of the data, data controller, etc.
- Processing operations: copy, transfer of data outside the EU, etc.
- Rights of the interested parties: the right of deletion, the right of rectification, etc.
- Processing purposes: statistical purposes, research purposes, security purposes, service provision purposes, and service improvement purposes.
- Legal bases: consent, contract, legal obligation, etc.

Despite the above, not all authors equally agree about using icons to complement the information for data subjects. Institutions that use icons or other nuances to facilitate self-determined choices must consider associated risks. The data subjects must accept the terms and conditions of the services once they are fully informed; therefore, consent must be informed. However, as Efroni et al. [44] state, "the process of giving consent is often uninformed and does not encourage self-determination"; and continues that "one of the key reasons for the lack of informed consent is that users do not adequately assess or even recognize the risks (or the possible negative consequences) involved in the treatment of your data". The risk-based approach in the design of "privacy icons" as stated by Efroni et al. must consider both individual and societal risks. Consequently, there is a disparity of opinions regarding whether the icons are complementary tools enough to allow a self-determined choice before consent, or additional ones are required to fulfill this task.

Despite the legal connotations of the GDPR regarding the use of icons and the efforts made by different authors, the current proposals for the representation of Articles 13 and 14 in icons (Privacy Icons as established by Efroni et al.) do not represent some of the "processing purposes" of Learning Analytics. Rossi and Palmirani define a subcategory inside "processing purposes" as "statistical purposes". However, "statistical purposes" is a category too broad to fully inform the data subject about the treatment detail of his or her data in Learning Analytics processes. A Learning Analytics process in education refers, and not only, to the analytical treatment of student data, where different techniques and methods can be used, such as:

- Predictive Analytics
- Descriptive Analytics
- Diagnostic Analytics
- Prescriptive Analytics
- Machine Learning
- Deep Learning
- Big Data
- Artificial Intelligence
- Neural Networks
- …

These actions are only those related to data analysis. However, there are other related actions such as the use of cookies, the internal transfer of data between departments, the storage of data for a certain time, or the processing of data for a certain period which, in part, are already included in the studies found, but that in Learning Analytics require adaptations beyond the analytical purpose, due to their connotations of fragility and sensitive context.

We confirm with this literature review that the icons designed to date include the generalities identified in the GDPR. However, they do not reach all the actions derived from using Learning Analytics. For both legal and ethical reasons, it is necessary to expand the scope of the icons with new designs to provide the maximum amount and accuracy of information to students about data processing in specific situations and contexts of Learning Analytics.

B. Phase 2

The project aims to provide students with icons to 1) generate awareness and knowledge about data processing by academic institutions and 2) make decisions based on accurate information. As stated in the introduction, this awareness and decision-making are possible if visual and standardized information is delivered clear, quick to understand, and intuitive. Considering the latter, we establish a fundamental requirement for the design of standardized icons between the execution of the study: users must know what the icon intends to inform them with only their observation. After the subsequent results of the analysis, we found that fulfilling this task depends on the subjectivities and beliefs of the participants, thus making standardization difficult and almost an impossible task.

We follow a qualitative-quantitative methodology based on conducting surveys to achieve the objectives of this phase. These surveys are executed in an iterative process, considering each iteration as a stage that depends on the previous one.

1. Materials

The materials used to develop the methodology of this second phase are mainly research instruments based on questionnaires. These questionnaires aim to collect the participants' perceptions regarding a series of icons associated with actions within Learning Analytics processes. Knowing the perception of the surveyed participants allows us to accept or discard icon designs considering consensus among responses.

Considering the iterative methodology, at each stage, one or more of these actions are performed:

- Survey the participants using multiple, open-ended, or drawing questions.
- Modify or create icons after analyzing the survey results.
- Create a new questionnaire from the new icons.

In the first iteration, we start with icons designed considering the results of the literature review carried out in phase 1. At the end of each stage of phase 2 and after analyzing the results, we create new icons or adapt those presented in the questionnaires. The questionnaires are made up of questions such as:

- Multiple answers. The question presents an icon, and participants must choose between five answers.
- Open-text answer. The question presents an icon, and participants must explain what the icon represents. Participants can indicate what modifications should be applied to make the icon clearer.
- Drawing response. The question shows an icon description, and participants must draw its graphic representation.

The resulting icons are presented in the following iterative stage. This cyclical methodology allows making changes justified by the participants' perceptions in a user experience loop.

2. Participants

The population is considered representative as both teachers and students are surveyed. Both are mainly interested in the use and visualization of the icons. A total of 103 people make up the surveyed population. Considering numbers, 15 are professor-researchers, and 88 are students. These amounts exceed the studies by Rossi and Palmirani, whose population sample is approximately 30 participants.

The project so far is divided into nine iterative stages:

- Stage 1: Creation of the first sketches. The population surveyed: 15 professor-researchers from the authors' research groups and 2 university students.
- Stage 2: Changes in some of the designs of the first phase and the addition of new ones. The surveyed population: 2 university students.
- Stage 3: Questionnaire fulfilling to validate the icons created in the previous stage. The surveyed population: 10 students.
- Stage 4: Modification of the icons considering the results of the previous stage. Conduction of questionnaires to validate the modifications. The surveyed population: 12 students.
- Stage 5: Conducted a questionnaire in which students must draw

icons based on a brief description. The objective is to acquire new design perspectives from the participants on some icons that pose problems of interpretation. The surveyed population: 15 students.

- Stage 6: Creation and modification of the icons obtained in the two previous phases.
- Stage 7: Questionnaire fulfilling to validate the icons created in the previous stage. The surveyed population: 30 students.
- Stage 8: Creation and modification of the icons obtained in the previous stage. Completion of a questionnaire to validate the newly designed icons. The surveyed population: 19 students.
- Stage 9: Last modification of the icons obtained in the previous phase (pending validation).

C. Results

Each stage of the methodology yields a series of icons validated by questionnaires. We expose some of those resulting icons for every stage. At the end of this section, we show a summary table with all the last icons resulting from the work done.

1. Stage 1

Fig. 2 exposes the first icons designed for validation in stage 1. In this stage, the very first icons are validated. The data transfer icon has 100% of consensus among the surveyed participants, not being as this regarding the icon representing the analytical treatment of data. The icons about predictive and prescriptive analytics icons needed iterative design changes to get interpretable.

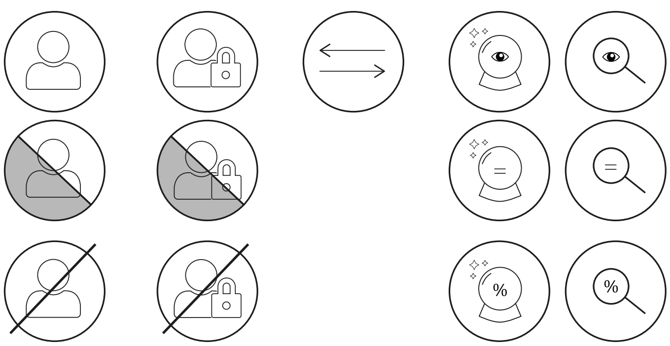


Fig. 2. First icons represent the collection of open and anonymized user data, no data collection, data transfer of any kind, and two variants of descriptive, prescriptive, and predictive data analysis using a magnifying glass and a crystal ball.

2. Stage 2

Fig. 3 shows some of the icons' modifications in stage 1 after surveying the participants. In this stage, the shape that identify a user is filled in black, and the document icon begins to be used to design icons regarding data collection.

3. Stage 3

Fig. 4 shows the designs elaborated after analyzing the results of the stage 2 survey. Those icons regarding data storage are designed considering a period. Numbers are added to the icons to identify the duration of data stored or treated.

4. Stage 4

Fig. 5 shows some of the (re)designed icons considering the results of the stage 3 survey. Changes to icons regarding analytical processing of data are made, where a magnifying glass is used instead of a crystal ball.

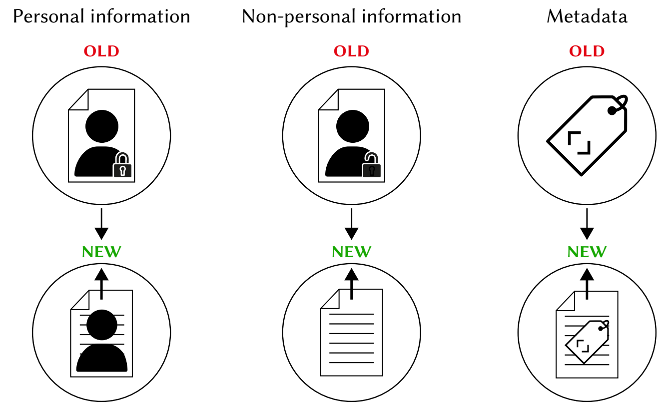


Fig. 3. Icons that identify the collection of "personal information", "non-personal information", and "metadata".

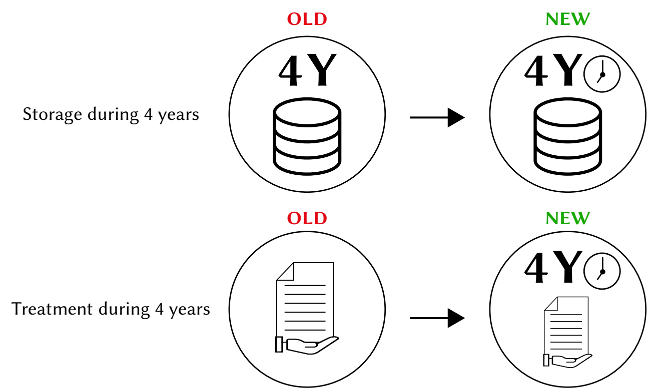


Fig. 4. Icons that identify, in this order, "data storage for 4 years" and "data treatment for 4 years".

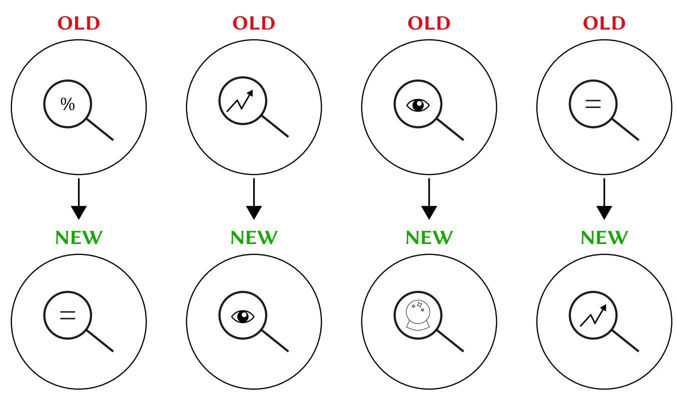


Fig. 5. Icons that identify, in this order, "descriptive analytics", "diagnostic analytics", "predictive analytics", and "prescriptive analytics".

5. Stage 5

Due to the inconclusive results, we ask the participants to draw a graphic representation of the icons. Fig. 6, Fig. 7, and Fig. 8 show some examples of the drawings made by the participants. Depending on the icon the results are more consistent, but others have pronounced differences.

6. Stage 6

Fig. 9 and Fig. 10 present the evolution of some icons between stages 5 and 6. Magnifying glasses are forgotten, and new icons are designed based on a collage. The collage idea is considered after analyzing the icons drawn by users.

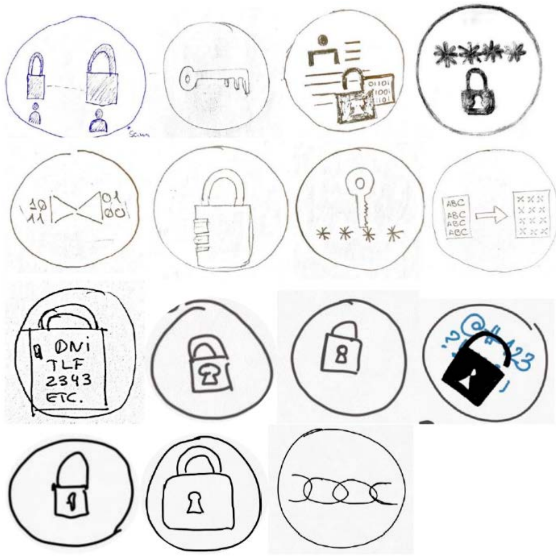


Fig. 6. Icons that identify "data encryption" drawn by participants.

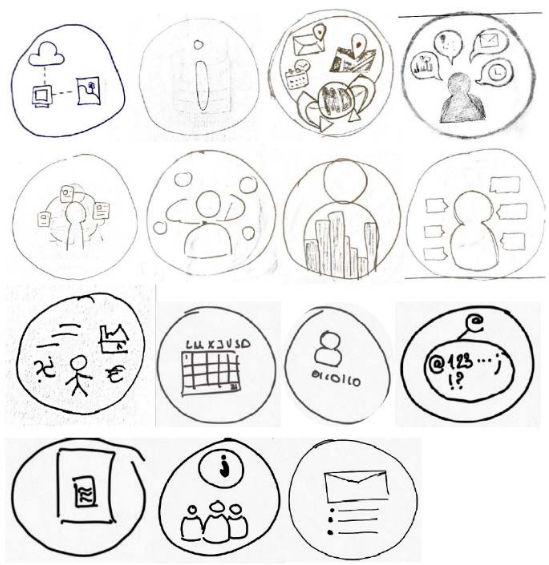


Fig. 7. Icons that identify "metadata collection" drawn by participants.

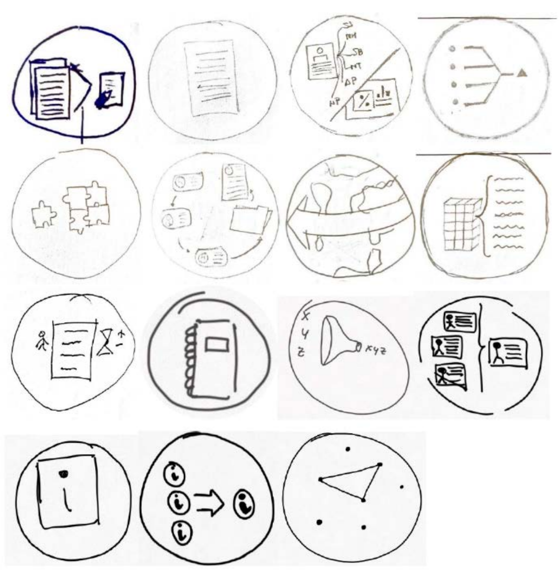


Fig. 8. Icons that identify "descriptive analytics" drawn by participants.

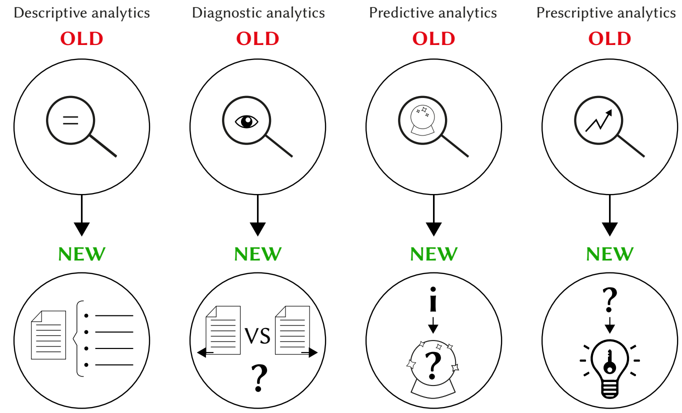


Fig. 9. Icons that identify, in this order, "descriptive analytics, "diagnostic analytics", "predictive analytics", and "prescriptive analytics".

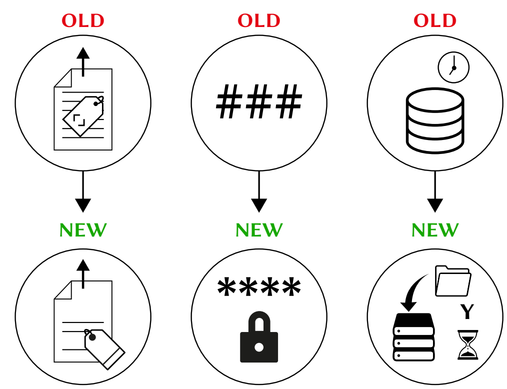


Fig. 10. Icons that identify the collection of "metadata collection", "data encryption", and "data storage for years".

7. Stage 7

Fig. 11 presents the evolution of some icons between stages 6 and 7. The idea of a crystal ball returns and icons are redesigned. The shape that identifies a document is used as the base for the icons that represent data analysis.

8. Stage 7

Fig. 11 presents the evolution of some icons between stages 6 and 7. The idea of a crystal ball returns and icons are redesigned. The shape that identifies a document is used as the base for the icons that represent data analysis.

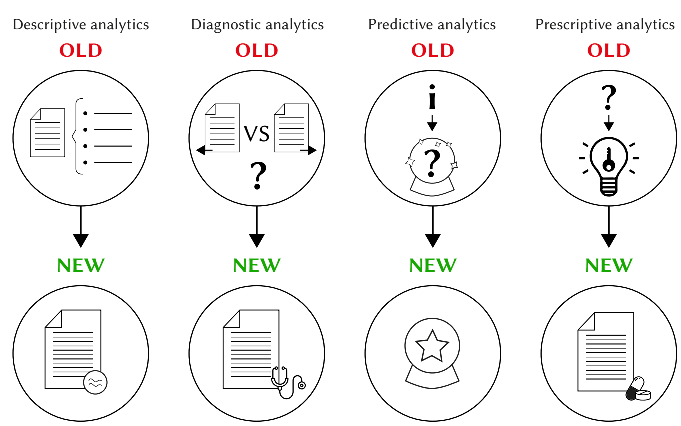


Fig. 11. Icons that identify, in this order, "descriptive analytics, "diagnostic analytics", "predictive analytics", and "prescriptive analytics".

9. Stage 8

Fig. 12 and Fig. 13 present the evolution of some icons between stages 7 and 8. In general, the use of a shape that identifies a document facilitates the comprehension of the icon regarding data collection or treatment. The icon identifying the collection of metadata needs to be redesigned and the tag shape is used to identify "other data" of users.

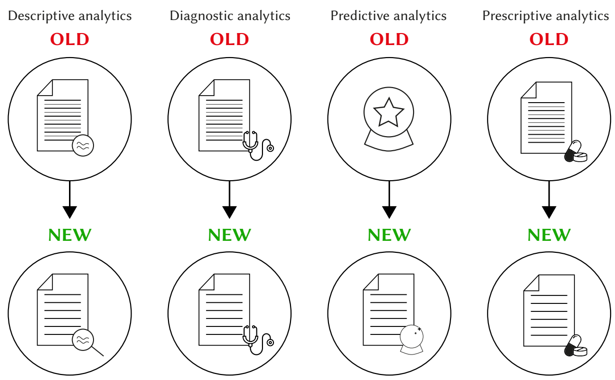


Fig. 12. Icons that identify the evolution of "descriptive analytics, "diagnostic analytics", "predictive analytics", and "prescriptive analytics".

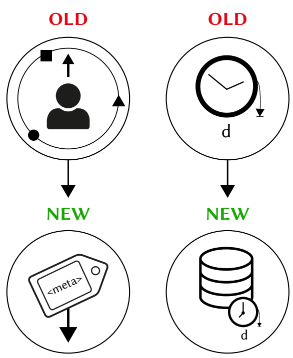


Fig. 13. Icons that identify, in this order, "metadata collection" and "data storage for days".

10.Stage 9

In this stage, we show resulting icons throughout the execution of the project using categories proposed by Jong and Spagnuelo, and Rosi and Palmirani (data collection, data storage, and processing operations). However, this is an approach that we consider not definitive and in which we are working to receive more feedback from students. For instance, there is no consensus in icons regarding "predictive analytics", where Fig. 14 shows both proposals being validated. The same happens with "cache technics", where the type of graphics inside icons seems to generate divergence, as shown in Fig. 15.

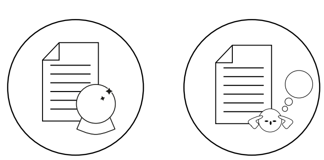


Fig. 14. Icons that identify "predictive analytics".

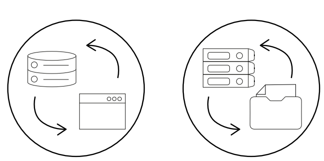


Fig. 15. Icons that identify "cache technics".

The icons extracted so far related to Learning Analytics processes are available in Table II and Table III, where its actions representation regarding data collection, data storage, and processing operations are exposed:

- Data is encrypted. Data encryption during data collection or storage.
- Data is anonymized. Anonymization of data during data collection or storage.
- Data is pseudonymized. Pseudonymization of data during data collection or storage.
- Personal information. Data collected or stored can identify data subjects.
- Metadata. Metadata collected or stored where metadata could be any non-personal data.
- Cookies. Use of web browser cookies.
- Cache technics. Use any cache technic in user devices (such as browser database), servers, or cloud computing.
- Storage for second(s), minute(s), hour(s), day(s), month(s), or year(s). Data storage of any kind for an estimated period.
- Descriptive analytics. Automated or manual descriptive data analytics approaches.
- Diagnostic analytics. Automated or manual diagnostic data analytics approaches.
- Predictive analytics. Automated or manual predictive data analytics approaches.
- Prescriptive analytics. Automated or manual prescriptive data analytics approaches.
- Data transfer to third parties. Data transfer outside the institution both in the European space or another accepted country outside European space.
- Internal data transfer. Data transfer inside the academic institution, such as between departments.
- Data treatment for second(s), minute(s), hour(s), day(s), month(s), or year(s). Data treatment of any kind for an estimated period. It differs from data storage, due data can be stored longer than treated, or vice versa.

For the "data collection" and "data storage" categories, 14 icons have been accepted by participants and are shown in Table II.

For "processing operations", 12 icons have been designed and validated by participants and shown in Table III.

D. Discussion

Designing, validating, and adapting icons that report data processing in Learning Analytics processes is an arduous task. After eight iterations, there is still a long way to go. The results obtained are very encouraging, and the icons achieved can be almost considered definitive. However, resulting icons will be subject to change as the laws and the GDPR are constantly evolving. Thus, this project will continue iterating icon designs to adapt to future law considerations.

After the collection and analysis of the results throughout the eight iterations, we affirm that:

1. It is very complex to create an icon that the affected subject knows what it is referring to just by viewing it. In this sense, it is necessary to accompany the icons with descriptive text with all the details to facilitate the comprehension of how the data will be treated.
2. It is possible to create icons that identify specific tasks and types of data analysis. Despite the difficulties expressed in the previous point, some icons show a 100% agreement among population responses; in other cases, the population agrees in 80%-95%. These

TABLE II. Icons Relative to Data Collection & Storage

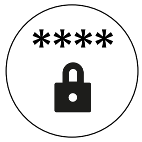
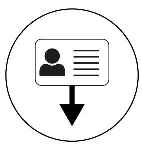
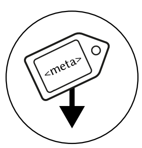
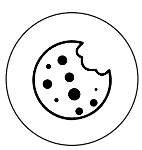
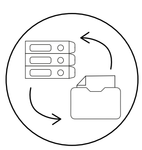
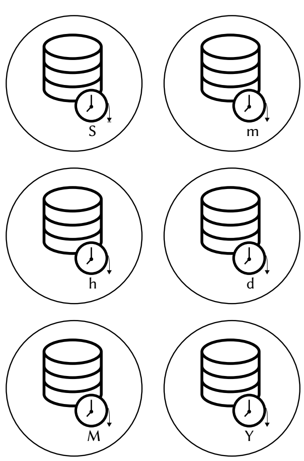
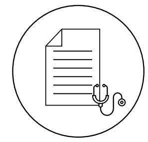
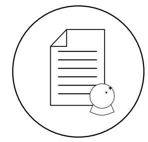
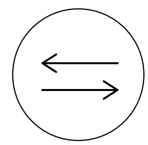
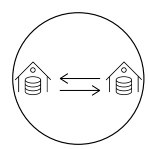
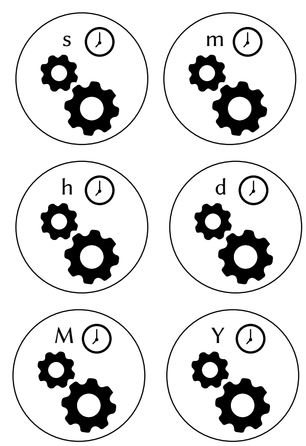
Icon description	Icon image
Data is encrypted	
Data is anonymized	
Data is pseudonymized	
Data is non-anonymized	
Personal information	
Metadata	
Cookies	
Cache technics	
Storage for second(s), minute(s), hour(s), day(s), month(s), or year(s)	

TABLE III. Icons Relative to Processing Operations

Icon description	Icon image
Descriptive analytics	
Diagnostic analytics	
Predictive analytics	
Prescriptive analytics	
Data transfer to third parties	
Internal data transfer	
Data treatment for second(s), minute(s), hour(s), day(s), month(s), or year(s)	

results indicate that the affected subjects can extract the general purpose of the icon, generating enough interest to end up reading the informative texts that accompany them if detail is needed.

We have found limitations throughout the execution of the project. The most important has been the COVID-19 pandemic. This project started in the 2019-2020 academic year. However, the COVID-19 pandemic stopped the project until we restarted it in the 2020-2021 academic year. Another limitation is the data subject's perception regarding legal terms, data processing, and Learning Analytics. These subjective perceptions lengthened the execution period of the project since there was not much consensus on the answers in the first icons iterations. Until the sixth iteration, the

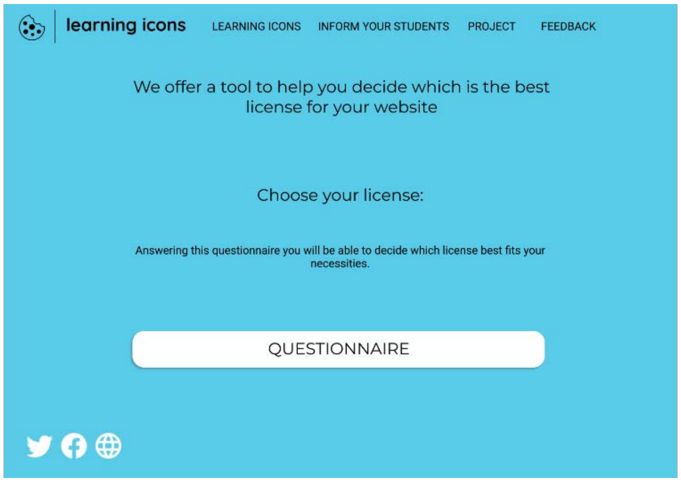


Fig. 16. The platform facilitates the creation of grouped icons with a questionnaire.

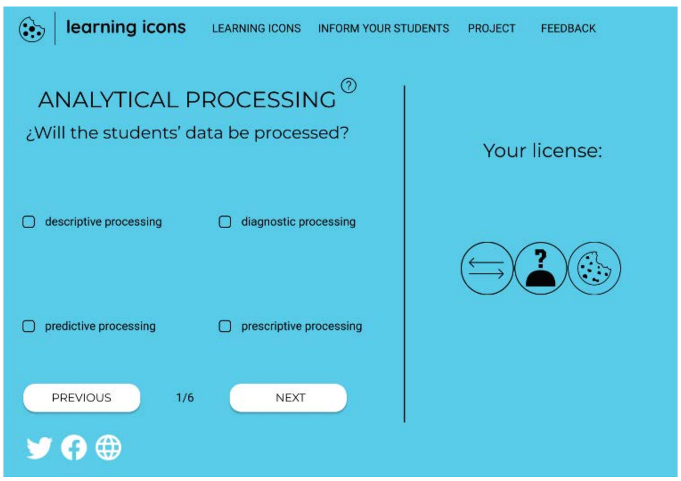


Fig. 17. The platform allows choosing what data treatment will be conducted.

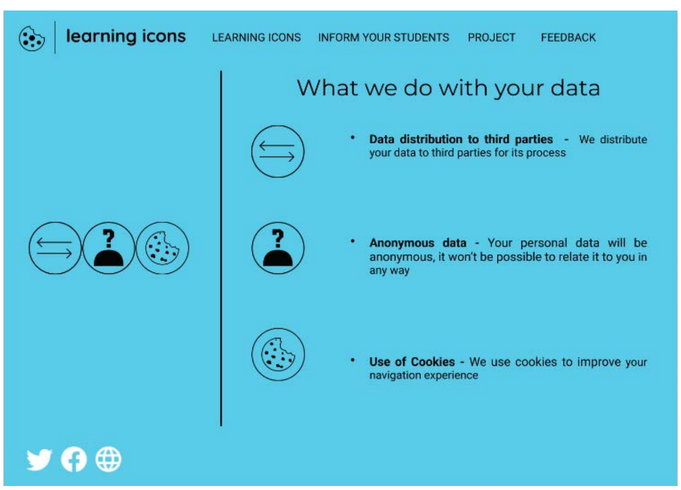


Fig. 18. The platform allows showing extended text for data treatment details.

redesigns were focused on creating as much consensus as possible. Afterward, we focused on enhancing the details of the icons to improve the accuracy of meaning.

The delay in achieving the first objective implies that we could partially complete the second objective. We have begun to develop and move towards its achievement. In this sense, we present in Fig. 16, Fig. 17, and Fig. 18 the screens of the web application as a platform to facilitate anyone to easily create icons regarding data treatment of students in Learning Analytics processes. The platform allows icons to be grouped into a single image linked to an informative space displaying its meaning regarding the educational context of data processing. Considering the Article 12.7 of the GDPR, each image will be accompanied by a JSON file so icons can be machine-readable.

E. Conclusion

The GDPR provides the possibility to accompany the information provided to the data subjects identified in articles 12, 13, and 14 and recital 60 with icons that can "provide in an easily visible, intelligible and legible manner, a meaningful overview of the intended processing" [34]. Different works before the enactment of GDPR and some after identifying categories and subcategories related to data collection, purposes of the processing, types of data, functions of agents, processing operations, rights of data subjects, purposes of the processing, the legal bases.

The icons resulting from these works identify different aspects of data processing indicated in the GDPR. However, they do not cover the full range of possibilities in other contexts where data is processed constantly in different manners. This is the case in the educational context, where analytical processes such as Learning Analytics are applied. Learning Analytics is an approach in which confidential student data is processed, generating profiles and aggregated data. According to the GDPR, students must be well informed, even before enrolling in any academic course.

Our work aims to generate a series of icons, which, complemented with the work of other authors, cover specific aspects of Learning Analytics. In this way, we hope to facilitate the understanding of the treatment of student data, considering the peculiarities of each educational institution. In the manuscript, we present a two-folded methodology. On the one hand, a documental methodology based on a systematic literature review with very limited results. These very limited results indicate that there is an open field for research, especially when regulations in Europe are susceptible to recurrent changes in terms of privacy. On the other hand, the qualitative-quantitative methodology used in the design of icons for the information students' data treatment in Learning Analytics processes. Participants are part of the academic field, in specific teachers and students; we used questionnaires to collect participants' perceptions to validate icons. We present all the icons generated in the results of the applied methodology. However, the surveyed population is Spanish, and in the following iterations, we will consider participants from other European countries.

As a second objective, we set the development of a platform to facilitate the integration of icons in any VLE. This platform will allow the creation of grouped informative icons linked to an explanatory text. This text is personalized considering the context of data processing -descriptive analyzes can be conducted in one course and predictive analyzes in another-. The platform is currently in development. Thus, the second objective is not accomplished. However, we present screens to appreciate some characteristics and functionalities.

Future work focuses on two phases. First, finish the Learning Analytics Icons platform. Second, extend participation from other countries. Subjectivities of participants are associated with cultural implications. Icons designed should be tested in countries different from Spain to create standard icons or adapted ones to the countries if participants' consensus is very divergent.

Funding

This research was funded by EU within the framework of the Erasmus+ Programme. Project Security and Privacy in Academic

Data management at Schools (SPADATAS) Grant Agreement n° 2022-1-ES01-KA220-000086363.

Acknowledgment

Many thanks to Aleix Ollé, Silvia Blasi, Javier Geli, and Rogelio Sansaloni who have helped in undertaking the research work, specially in the execution of surveys and icon design.

References

[1] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. J. D. Baker, *Handbook of educational data mining*. CRC press, 2010, p. 503. doi: 10.1201/b10274.

[2] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, "A reference model for learning analytics," *International Journal of Technology Enhanced Learning*, vol. 4, no. 5–6, pp. 318–331, 2012, doi: 10.1504/IJTEL.2012.051815.

[3] F. J. García-Peñalvo, "Learning Analytics as a Breakthrough in Educational Improvement," in *Radical Solutions and Learning Analytics*, Springer, 2020, pp. 1–15. doi: 10.1007/978-981-15-4526-9_1.

[4] D. Amo, M. Alier, M. J. Casan, and M. J. Casañ, "The student's progress snapshot a hybrid text and visual learning analytics dashboard," *The International Journal of Engineering Education*, vol. 34, no. 3, pp. 990–1000, 2018.

[5] G. Siemens, "What are Learning Analytics?," *Elearnspace*, no. 1, pp. 1–1, 2010. Accessed: Jan. 01, 2015. [Online]. Available: https://web.archive.org/web/20100827114932/http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

[6] D. Amo *et al.*, "Using Web Analytics Tools To Improve the Quality of Educational Resources and the Learning Process of Students in a Gamified Situation," in *INTED2018 Proceedings*, 2018, vol. 1, pp. 5824–5829. doi: 10.21125/inted.2018.1384.

[7] M. Alier, M. J. Casany, C. Severance, and D. Amo, "Learner Privacy, a pending assignment," in *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality*, New York, NY, USA, Oct. 2020, pp. 725–729. doi: 10.1145/3434780.3436635.

[8] M. G. Alonso de Castro and F. J. García-Peñalvo, "Successful educational methodologies: Erasmus+ projects related to e-learning or ICT," *Campus Virtuales*, vol. 11, no. 1, pp. 95–114, 2022, doi: 10.54988/cv.2022.1.1022.

[9] D. Amo, "MOOCs: Experimental approaches for quality in pedagogical and design fundamentals," in *ACM International Conference Proceeding Series*, 2013, pp. 219–223. doi: 10.1145/2536536.2536570.

[10] Z. N. Khlaif, M. Ghanim, A. A. Obaid, S. Salha, and S. Affouneh, "The Motives and Challenges of developing and delivering MOOCs courses," *Education in the Knowledge Society*, vol. 22, p. art. e23904, 2021, doi: 10.14201/eks.2390.

[11] M. Á. Conde-González, F. J. García-Peñalvo, M. J. Rodríguez-Conde, M. Alier, and A. García-Holgado, "Perceived openness of Learning Management Systems by students and teachers in education and technology courses," *Computers in Human Behavior*, vol. 31, pp. 517–526, 2014, doi: 10.1016/j.chb.2013.05.023.

[12] Á. Fidalgo-Blanco, M. L. Sein-Echaluce, F. J. García-Peñalvo, and M. Á. Conde, "Using Learning Analytics to improve teamwork assessment," *Computers in Human Behavior*, vol. 47, pp. 149–156, 2015, doi: 10.1016/j.chb.2014.11.050.

[13] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses," in *ACM International Conference Proceeding Series*, 2013, pp. 170–179. doi: 10.1145/2460296.2460330.

[14] X. Ochoa, N. Weibel, M. Worsley, and S. Oviatt, "Multimodal learning analytics data challenges," in *6th International Conference on Learning Analytics and Knowledge, LAK 2016*, 2016, pp. 498–499. [Online]. Available: https://sci-hub.se/https://nyu-staging.pure.elsevier.com/en/publications/multimodal-learning-analytics-data-challenges

[15] A. Álvarez-Arana, M. Villamañe-Gironés, and M. Larrañaga-Olagaray, "Improving Assessment Using Visual Learning Analytics," *Education in the Knowledge Society*, vol. 21, no. 9, pp. 1–9, 2020, doi: 10.14201/eks.21554.

[16] D. Amo, R. Torres, X. Canaleta, J. Herrero-Martín, C. Rodríguez-Merino, and D. Fonseca, "Seven principles to foster privacy and security in educational tools: Local Educational Data Analytics," in *TEEM'20: Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 2020, pp. 730–737. doi: 10.1145/3434780.3436637.

[17] H. Drachsler and W. Greller, "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*, Edinburgh, United Kingdom, 2016, pp. 89–98. doi: 10.1145/2883851.2883893.

[18] F. J. García-Peñalvo and A. Corell, "La COVID-19: ¿enzima de transformación digital de la docencia o reflejo de una crisis metodológica y competencial en la educación superior?," *Campus Virtuales*, vol. 9, no. 2, pp. 83–98, 2020.

[19] F. J. García-Peñalvo, A. Corell, R. Rivero-Ortega, M. J. Rodríguez-Conde, and N. Rodríguez-García, "Impact of the COVID-19 on Higher Education: An Experience-Based Approach," in *Information Technology Trends for a Global and Interdisciplinary Research Community*, IGI Global, 2021, pp. 1–18. doi: 10.4018/978-1-7998-4156-2.ch001.

[20] F. J. García-Peñalvo, A. Corell, V. Abella-García, and M. Grande, "Online assessment in higher education in the time of COVID-19," *Education in the Knowledge Society*, vol. 21, pp. 12–26, 2020, doi: 10.14201/eks.23013.

[21] D. Amo, M. Alier, F. J. García-Peñalvo, D. Fonseca, and M. J. Casany, "GDPR security and confidentiality compliance in LMS' a problem analysis and engineering solution proposal," in *TEEM'19: Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*, León, 2019, pp. 253–259. doi: 10.1145/3362789.3362823.

[22] Y. Jang, R. Katz, and K. Dalkir, "Are Higher-Education Institutions Ready for Learning Analytics? Governance, Ethics, Confidentiality and Privacy," *Journal of Leadership, Accountability and Ethics*, vol. 18, no. 1, pp. 137–148, 2021.

[23] D. Amo, M. Alier, F. García-Peñalvo, D. Fonseca, and M. J. Casañ, "Privacidad, seguridad y legalidad en soluciones educativas basadas en Blockchain: Una Revisión Sistemática de la Literatura," *RIED. Revista Iberoamericana de Educación a Distancia*, vol. 23, no. 2, pp. 213–236, 2020, doi: 10.5944/ried.23.2.26388.

[24] J. Seanosky, D. Jacques, and V. Kumar, "Security and Privacy in Bigdata Learning Analytics," in *Proceedings of the 3rd International Symposium on Big Data and Cloud Computing Challenges (ISBCC–16')*, Cham, Switzerland, 2016, pp. 43–55. doi: 10.1007/978-3-319-30348-2_4.

[25] F. Jose Garcia-Penalvo, "Digital Identity as Researchers. The Evidence and Transparency of Scientific Production," *Education in the Knowledge Society*, vol. 19, no. 2, pp. 7–28, 2018, doi: 10.14201/eks201819272.

[26] M. Alier, M. J. Casany, C. Severance, and D. Amo, "Learner Privacy, a pending assignment," in *ACM International Conference Proceeding Series*, 2020, pp. 725–729. doi: 10.1145/3434780.3436635.

[27] B. Herold, "InBloom to Shut Down Amid Growing Data-Privacy Concerns," *Education Week*, 2014. [Online]. Available: http://blogs.edweek.org/edweek/DigitalEducation/2014/04/inbloom_to_shut_down_amid_growing_data_privacy_concerns.html

[28] M. Grothaus, "Pearson data breach: details of hundreds of thousands of U.S. students hacked," *Fast Company*, 2019. [Online]. Available: https://www.fastcompany.com/90384759/pearson-data-breach-details-of-hundreds-of-thousands-of-u-s-students-hacked

[29] L. Amoore, "Why 'Ditch the algorithm' is the future of political protest," *The Guardian*, 2020. [Online]. Available: https://www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics

[30] A. Pardo and G. Siemens, "Ethical and privacy principles for learning analytics," *British Journal of Educational Technology*, vol. 45, no. 3, pp. 438–450, May 2014, doi: 10.1111/bjet.12152.

[31] S. Slade and P. Prinsloo, "Learning Analytics: Ethical Issues and Dilemmas," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1510–1529, 2013, doi: 10.1177/0002764213479366.

[32] F. J. García-Peñalvo, "Avoiding the dark side of digital transformation in teaching. An institutional reference framework for eLearning in higher education," *Sustainability*, vol. 13, no. 4, p. art. 2023, 2021, doi: 10.3390/su13042023.

[33] EP and the CEU, "Regulation (EU) 2016/679 GDPR," *Official Journal of the European Union*, pp. 88–88, 2016. [Online]. Available: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679

[34] E. and the CEU, "Recital 60 - Information obligation | General Data Protection Regulation (GDPR)." [Online]. Available: https://gdpr-info.eu/

recitals/no-60/

[35] A. García Holgado, S. Marcos Pablos, and F. J. García Peñalvo, "Guidelines for performing systematic research projects reviews," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 2, pp. 136–144, 2020, doi: 10.9781/ijimai.2020.05.005.

[36] F. J. García-Peñalvo, "Developing robust state-of-the-art reports: Systematic Literature Reviews," *Education in the Knowledge Society*, vol. 23, 2022.

[37] D. Moher *et al*., "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *PLoS Medicine*, vol. 6, no. 7, pp. e1000097–e1000097, Jul. 2009, doi: 10.1371/journal.pmed.1000097.

[38] A. Rossi and M. Palmirani, "DaPIS: An Ontology-Based Data Protection Icon Set," *Knowledge of the Law in the Big Data Age*, vol. 317, no. 978-1-61499-984–3, pp. 181–195, 2019, doi: 10.3233/FAIA190020.

[39] A. Rossi and M. Palmirani, "Can visual design provide legal transparency? The challenges for successful implementation of icons for data protection," *Design Issues*, vol. 36, no. 3, pp. 82–96, 2020, doi: 10.1162/desi_a_00605.

[40] M. Palmirani, A. Rossi, M. Martoni, and M. Hagan, "A methodological framework to design a machine-readable privacy icon set," in *Jusletter IT*, 2018, no. February.

[41] A. Rossi and G. Lenzini, "Which Properties Has an Icon? A Critical Discussion on Data Protection Iconography," in *Socio-Technical Aspects in Security and Trust*, Cham, 2021, pp. 211–229. doi: 10.1007/978-3-030-55958-8_12.

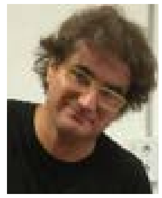
[42] A. Rossi and M. Palmirani, "A Visualization Approach for Adaptive Consent in the European Data Protection Framework," in *2017 Conference for E-Democracy and Open Government (CeDEM)*, May 2017, pp. 159–170. doi: 10.1109/CeDEM.2017.23.

[43] S. de Jong and D. Spagnuelo, "Iconified representations of privacy policies: A GDPR perspective," in *Advances in Intelligent Systems and Computing*, 2020, vol. 1160, pp. 796–806. doi: 10.1007/978-3-030-45691-7_75.

[44] Z. Efroni, J. Metzger, L. Mischau, and M. Schirmbeck, "Privacy icons: A risk-based approach to visualisation of data processing," *European Data Protection Law Review*, vol. 5, no. 3, pp. 352–366, 2019, doi: 10.21552/edpl/2019/3/9.

Daniel Amo-Filva

Ph.D. in Education Sciences from the University of Salamanca (2020), with two master's degrees in education and educational technology, a University Master's Degree in Teacher Training for Compulsory Secondary Education, and a Baccalaureate, Professional Training and Language Teaching (UNIR 2016) and University Master's Degree in Education and ICT, specialization in Research (UOC 2014). With 15 years of participating in and leading different technological projects in the tech-industry, he currently focuses his professional career on university teaching in the Department of Computer Engineering at La Salle, Ramon Lull University, and research group HER (Human-Environment Research). Within HER he coordinates the Technology-Enhanced Learning (TEL) research line focused on the design, implementation, evaluation, and improvement of the impact of educational projects in any academic fields mediated by technology. The use of ethical and analytical methods for educational data treatment are the fundamentals to the evaluation and improvement of such procedures. He actively participates in scientific congress committees and conferences to disseminate to society the knowledge resulting from his professional and personal research. He is the author of the two books "Learning Analytics: the narrative of learning through data" (UOC OuterEdu) and "Learning analytics: 30 experiences in the classroom with data"; the blog eduliticas.com where he disseminates about Learning Analytics; the podcast connecta.danielamo.info where tackles technology, privacy, and humanistic aspects of society; and danielamo.info a personal space where he shows all his past and present projects. Research publications are available at Google Scholar (https://scholar.google.com/citations?user=RNHbv9oAAAAJ&hl=es&oi=ao), and ORCID (https://orcid.org/0000-0002-4929-0438).

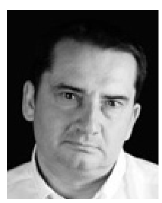


Marc Alier

Marc Alier (1971) received an engineering degree in computer science (1996) and a PhD in Sustainability (2009) in the Polytechnical University of Catalonia (UPC). He is an associate professor at UPC and deputy director at ICE http://www.ice.upc.edu. The last 25 years have worked in research and development related to the e-learning industry. Has participated in the development of several LMS, content authoring tools and interoperability standards. Since 2001, has taught software engineering, project management, information systems, and computing ethics at UPC's School of Informatics. Has been director of several master's programs. Has authored more than 120 papers in journals and conference proceedings. Since 2007 produces several podcasts about technology, science, and its impact on society as a means of dissemination of his professional and personal research.

David Fonseca

Full Professor (2017) by La Salle Ramon Llull University, currently he is the coordinator of the Group of Research on Technology Enhanced Learning (GRETEL), a recognized research group of Generalitat de Catalunya (from 2014), and coordinator of the Graphic Representation Area in the Architecture Department of La Salle (where he is a teacher and academic tutor). Technical Engineer in Telecommunications (URL – 1998), Master in GIS (Universitat de Girona, 2003), Audiovisual Communication Degree (UOC, 2006), Master in Advanced Studies (URL-2007), Official Master in Information and Knowledge Society (UOC, 2008), PhD in Multimedia by URL (2011), also, he is Autodesk Approved and Certified Instructor from 1998. With extensive experience in project manager (from 2000 to act, he has coordinated more than 50 local, national, and international projects, both technological transfer and research funded projects), he has directed 7 PhD thesis and more than 10 other final degree and master projects. Currently he is serving as program or scientific committee in more than 15 indexed journals and conferences, as well as organizing workshops, special issues and invited sessions in different scientific forums.



Francisco José García Peñalvo

Full Professor in the Department of Computer Science and Automation at the University of Salamanca (USAL), with three six-year periods of research, one six-year period of transferring and innovation, and four five-year periods of recognized teaching. He received the Gloria Begué award for teaching excellence in 2019. He was also a Distinguished Professor at the School of Humanities and Education at the Tecnológico de Monterrey, Mexico (2016-2018) and is a Researcher of International Impact at the Universidad Nacional San Agustín, Arequipa, Peru. Since 2006 he is the head of the Research Group Recognized by the USAL GRIAL (research GRoup on InterAction and eLearning), a group that is a Consolidated Research Unit of the Junta de Castilla y León Government (UIC 81). Included in the University of de Stanford's World's Top 2% Scientists list (2019, 2020, 2021) https://doi.org/10.17632/btchxktzyw.3. He has supervised 28 Ph.D. thesis. He has been Vice-Dean of Innovation and New Technologies of the Faculty of Sciences of the USAL between 2004 and 2007 and Vice-Rector of Technological Innovation of this University between 2007 and 2009. He is currently the Deputy Director of the Research Institute for Educational Sciences (IUCE), the Rector's Delegate for Digital Learning and Teaching and the Coordinator of the Doctorate Programme in Education in the Knowledge Society at USAL. He is Editor-in-Chief of the journals Education in the Knowledge Society and Journal of the Information Technology Research, and Associate Editor of many journals, with a special mention to the journals IEEE Transactions on Learning Technologies, IEEE Access, Computers in Human Behavior, and Computers in Human Behavior Reports. He has published more than 100 research papers in JCR SCIE/SSCI-indexed journal (55 Q1). For more detailed information on the publications, these are the public links to the profiles in Google Scholar (http://goo.gl/sDwrr0), Publons (https://bit.ly/2u2FN5l), Scopus (https://bit.ly/3IYoog7), and ORCID (http://orcid.org/0000-0001-9987-5584).

María José Casañ

Is a computer science Engineer from Polytechnic University of Catalonia - UPC (1997) and has a Ph.D. in Science (2013) from UPC. From 2004 she has been a researcher and lecturer, teaching at the School of Informatics on UPC. She has also been a course instructor at the Open University of Catalonia UOC. She teaches courses on Software engineering Projects, Databases, Social and Environmental aspects of computing as well as history of computers. She has developed several Open source projects such as the J2MEMicroDB (a database engine for Mobile devices) and a migration of the authoring software JClic to the OLPC X0 platform (http://laptop.org). She has also participated in the development of several LMS projects and authoring tools. Her research interests focus on the social aspects of engineering education, innovation in higher education degrees, educational innovation that contributes to the quality teaching enhancement, sustainability, and techno-ethics.