

Article

Educational Warehouse: Modular, Private and Secure Cloudable Architecture System for Educational Data Storage, Analysis and Access

Daniel Amo ^{1,*}, Pablo Gómez ², Luis Hernández-Ibáñez ³ and David Fonseca ^{4,*}¹ Group of Research GRETEL, La Salle, Engineering Department, Ramon Llull University, 08022 Barcelona, Spain² Intern at Group of Research GRETEL, La Salle, Engineering Department Ramon Llull University, 08022 Barcelona, Spain; pablo.gp@salle.url.edu³ Group of Research VideLAB, Civil Engineering Department, Universidade da Coruña, 15071 A Coruña, Spain; luis.hernandez@udc.es⁴ Group of Research GRETEL, La Salle, Architecture Department Ramon Llull University, 08022 Barcelona, Spain

* Correspondence: daniel.amo@salle.url.edu (D.A.); david.fonseca@salle.url.edu (D.F.)

Abstract: Data in the educational context are becoming increasingly important in decision-making and teaching-learning processes. Similar to the industrial context, educational institutions are adopting data-processing technologies at all levels. To achieve representative results, the processes of extraction, transformation and uploading of educational data should be ubiquitous because, without useful data, either internal or external, it is difficult to perform a proper analysis and to obtain unbiased educational results. It should be noted that the source and type of data are heterogeneous and that the analytical processes can be so diverse that it opens up a practical problem of management and access to the data generated. At the same time, ensuring the privacy, identity, confidentiality and security of students and their data is a “sine qua non” condition for complying with the legal issues involved while achieving the required ethical premises. This work proposes a modular and scalable data system architecture that solves the complexity of data management and access. On the one hand, it allows educational institutions to collect any data generated in both the teaching-learning and management processes. On the other hand, it will enable external access to this data under appropriate privacy and security conditions.

Keywords: cloud computing; modularity; educational data analytics; learning analytics; academic analytics; data privacy and security; digital identity; API; data storage; open data

Citation: Amo, D.; Gómez, P.; Hernández-Ibáñez, L.; Fonseca, D. Educational Warehouse: Modular, Private and Secure Cloudable Architecture System for Educational Data Storage, Analysis and Access. *Appl. Sci.* **2021**, *11*, 806. <https://doi.org/10.3390/app11020806>

Received: 30 November 2020

Accepted: 12 January 2021

Published: 16 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Data Revolution: Industry, Society and Education

The industrial revolution 4.0, where big data is considered the core, fosters great advances in the technification, digitalization and datafication of the business sector. As Marciano et al. said, “We are just at the beginning of this co-evolutionary path leading toward an epochal revolution. This process is affecting all sectors and all countries.” [1]. Some technologies such as big data, artificial intelligence and machine learning stand out in this revolution, all becoming increasingly present in society in various forms, such as platforms or mobile apps due to the adoption of cloud computing by the business sector. Two main reasons support this adoption, one economical and the other technological. Gong et al. set forth the economic perspective of cloud computing as an “economic pattern as the main reason why so many companies jump into the hot pool of cloud computing”. He also described its main technological features as “service-oriented, loose coupling, strong

fault-tolerant, business model and eas[y to] use” [2] (p. 1). The ease of use of cloud computing combined with virtualization has made it possible to reduce costs on data access and storage, and this has been a huge game changer in the business sector. As stated by Marston et al., “one of the significant opportunities of cloud computing lies in its potential to help ... upfront investments that have stymied past efforts Small businesses can exploit high-end applications like ERP software or business analytics that were hitherto unavailable to them.” [3–5] (pp. 181–182).

The union of these technologies has directly impacted society, with the Internet of Things being a clear example of the ease of everyday tasks using thermostats, surveillance cameras and buttons to dispense toilet paper. This trend is expected to follow an exponential growth in terms of technical change and socioeconomic impact [6,7]. Natural language processor technologies based on machine learning or deep learning are examples providing correction and translation services online. Artificial intelligence-based technologies capable of recreating non-existing human faces are sometimes used to create fake videos promoting legal and moral debate. This can be seen in the works of authors such as Perc et al. regarding the juristic challenges of artificial intelligence [8], Bechmann and Kim regarding research ethics in big data [9], Dixon-Román and Parisi regarding the growth of data capitalism and ethics in artificial intelligence [10] or Coghlan et al. regarding artificial intelligence in an education context [11].

Many solution stacks are based on a cloud computing centralized structure to offer the same product to everyone. Despite that fact, these technologies cannot be placed on closed environments such as a mobile app because its development is continuous and the machine learning process requires quality data to improve, providing a big data environment and interceding for it. Machine learning provides an accurate and precise solution if the dataset on which it is based contains lots of quality data. That is why these technologies require lots of real-time data to offer a decent service, with its mistakes included, even in the research context, in the need to balance privacy when creating knowledge for society’s good as stated by Bechmann and Kim [9].

Some technologies, like machine learning, are not based on new concepts. Some authors already acknowledged this concept in the 90s [12], but the amount of data creates a need for those technologies. To sum up, the business sector offers machine learning-based services on cloud computing environments, promoting a society–business interdependent cycle with data lying in its central point. In this scheme, ethics [13], data privacy and trust [14] are essential to a fair balance of power. This is defined as the main statement for data capitalism, a concept related to a series of social disconformities provoked by big technological companies. Some authors such as Bellamy [15], West [16] and Zuboff [17] give the name “Surveillance Capitalism” to this new data-based social era where this interdependency is unbalanced against the citizen. Some examples can be found in fake news [1], data leaking [18], society manipulation and digital attention economy [19] besides a lack of privacy and security.

1.2. Educational Context

Unfortunately, the education sector faces a similar situation. This context reflects the evolution of the society–industry binomial. Its growth rate is lower, but there is a trend to apply these data-based technologies coming from the business sector [20]. To be precise, the educational sector is the new target for companies to increase their profits, as many governmental initiatives such as Spain’s HAZ [21] and UK’s educational space standardization [22] are introducing companies to this sector as a new market [23,24]. It would not be surprising if these companies could offer big data technology applicable to the educational sector.

From this first contact between business towards education, the benefits of applying big data technologies and methodologies will undoubtedly be experienced [25–27], since they are primarily accepted as positive by society. The available data should explain the evolution of the educational context, the teaching process, learning–teaching processes,

academic performance and even behavior, aiming at improving quality and results. Due to this fact, many business technological practices are incorporated into the educational context. These specific practices are big data, machine learning, artificial intelligence and cloud computing. Combining those technologies causes the loss of ownership of the data generated by educational institutions due to the digital learning deployment environments from the educational centers in the control of the business. This migration of educational services is due to the resource administration and budget reduction implied by cloud computing. This new context of outsourced Educational Technology (EdTech) offers automatic processes such as educational data mining, learning analytics, academic analytics, multimodal learning analytics or a new generation of smart student advisors, merged all together as a new digital ruleset. This new EdTech, owned by biased algorithms [28,29], makes it easier to automatize decisions, overriding human criteria and opening contemporary debates in a social and educational context [30].

1.3. Data-Related Issues

The big data technological stack applied to education implies a big revolution. Still, it also enhances distrust among educational institutions by using these analytical technologies, as it provides an unreliable context and a loss of control [31]. Some examples of these analytical technologies are the massive automatic decision-making; massive sensible data collection from students [32]; unauthorized access to data [33]; enormous filtering, analysis and predictive tools against students will [34]; or data transfer without a legally defined relation [35]. There is an unstable situation in collecting, treating and analyzing educational data, metadata and personal data [31].

The enthusiasm for integrating big data processes, data-based decision-making, data processing and even international transfer between countries has, in some cases, led to problems of misuse, filtering and improper access [36]. This educational data revolution raises concerns both ethically and in terms of exposing students' privacy, identity, confidentiality and security of data, personal data and metadata (PICSDPDM) [32,35]. An example of this is the use of learning analytics in educational processes, which has, since its inception, aroused mistrust of data collection and processing processes [37]. We should not rule out the privacy and security weaknesses of the various EdTech companies that facilitate massive data theft, as shown in data breaches. It was experienced among US schools [38], in student loans companies [39] and even public administrations that allowed thousands of parents' and students' data to be left uncovered, as happened in Madrid [40] and Catalunya [41].

Facing this devastating picture, we focus on PICSDPDM-compliant solutions, such as encryption [32,35] or visualization [42] plugins for Learning Management Systems (LMS) or even ethics principles in educational data analytics [33]. From this point of view, the legal system is far from regulating every technological issue, as its growth rate is far greater than the legal evolution, as in the Blockchain's case, an example of technology already used but far away from compliance of regulations such as General Data Protection Regulation (GDPR), as shown by the work of Amo et al. [33] and of Lafarre and Van der Elst [43]. Despite these differences, the legal system is working towards standardization. Legal frameworks as GDPRs created a base, while in the EU, it assures the chance to deploy the technologies mentioned before (big data, machine learning, learning analytics and cloud computing). Due to massive espionage laws enacted in the USA, there is a collection of doubts pointing out its unreliable status, as shown in the "Schrems I" and "Schrems II" cases [44–49]. These two resolutions automatically invalidated the US-EU Privacy Shield [50], hence reinforcing the already existing GDPR legal framework in Europe [51] or Data Privacy Law in California [52] or even forcing the development of legal frameworks worldwide to assure data exchange such as the Data Protection & Privacy in Australia [53]. There is a general awareness about privacy for everyone (including students).

Still, there is a long way to solving it in a legal context, as current laws are oriented towards correcting instead of preventing. We believe in the previous statement and agree that the use of technology may lead to a balance.

1.4. Balance between Punishment and Prevention

As previously stated, there is a complicated ongoing situation in data management in the educational sector, specifically in the student–university relationship (within the university level) and in the entity–administration–university relationship (interuniversity status). A solution is required that will allow data access and management across entities, considering that entry could be public and private. Data’s nature, fragile and sensible, and the need to protect it, ensuring or even sharing its open core, affect its refinement, visualization, access and exchange. This requirement was reason we create a technological structure that respects the given concerns (data analysis, access management and privacy-friendliness for this sector).

Regarding legal matters, we have stated that, legality, as of today, this (solution/proposal/approach/structure mentioned above) is very far from regulating emerging technologies. We believe that it is more corrective than preventive. There are still problems concerning trust and loss of control in managing educational data that are not avoided due to legal loopholes. This situation generates an asymmetry of power that tilts the balance in favor of the technology companies’ profits, leaving users who use their services at their mercy and without many desirable privacy settings. An example of this is the analysis by Norwegian Consumer Councils. It highlights the dark patterns that those technologies such as Google, Facebook or Microsoft use to reduce the privacy options on their devices [54] and to even force the users to accept being tracked continuously [55]. The guide published by the Norwegian Data Inspectorate is another example in an educational context of how to monitor and profile students when using Google tools [56].

We are aware that there are as many decrees and regulations (from now on, laws) as there are different jurisdictions, some of them more prone to protecting the citizens’ PICSDPMs, including educational roles, and some less.

To reach a balance between correction and legal prevention and to avoid power asymmetries, we believe that it is necessary to have a technological stack that automates every jurisdiction’s legal framework by default and design. In this approach, the educational data’s control must be the first point to solve, either through pseudo-anonymization, anonymization or encryption procedures. We affirm and demonstrate that the previous sentence is made possible using developments made by some of the authors of this paper, such as the Protected Users plug-in [32] for the Moodle LMS. The Protected Users plug-in allows any user to adopt a second identity and to remain anonymous for whatever reason required, such as situations of harassment and cyber bullying, or gender violence, or for any other cause that requires him or her to remain anonymous. Regarding GDPR, students have the right to anonymize their identity in any of the courses they are enrolled due to those mentioned and other undesired causes. The policies and data privacy plugins for LMS solve general issues raised by the GDPR but usually do not allow students to be anonymous in any enrolled course. Ensuring PICSDPDM of educational roles is possible with a technology stack that automates laws. Therefore, we are faced with a need for acceptable practices in EdTech and a data management architecture that facilitates them.

1.5. A Proposal for a Modular and Scalable Architecture

Technology is always present. However, law and ethics are not found implemented together in some educational technology solutions. Therefore, we believe that the triad formed by laws, ethics and technology must always be present in order to face the problems already mentioned, now and in the future. Regulations are inevitable, but laws should not be seen as the new ethics. EdTech provides classrooms with a set of digital tools that may enhance learning but neither diminishes its possibilities nor raises privacy data problems and concerns. LegalTech offers legal protection as a default [8,57–59],

which is understood in this paper as the technology to help the data privacy officer and legal department enforce the law. Hence, somehow LegalTech and EdTech should unite as a holistic and integral solution to ensure educational data privacy and security.

- From our point of view, the educational sector requires a technological stack that
- allows storing, refining and analyzing educational data following the law and ethics in a way that respects all academic roles. Institutions have to be capable of applying and automating the law's rigor in their educational solutions by default and design. Nonetheless, they should also be agile and flexible in applying their moral and ethical principles, as stated in their educational project's mission, vision and scope, without breaking the legal regulation.
 - gives the educational centers a choice between a local deployment or a cloud computing scheme but prioritizing local deployment. We propose that approach in detail in our Local Educational Data Analytics (LEDA) framework's principles [60]. We expose the need to improve first local ad hoc solutions as an additional solution and to acknowledge the perks of using the aforementioned technological stack (big data, machine learning, artificial intelligence and cloud computing) on education yet refusing its malpractices. The seven principles of the LEDA Framework are (1) legality; (2) transparency, information and expiration; (3) data control; (4) anonymous transactions; (5) responsibility in the code; (6) interoperability; and (7) local first, where we advocate for this principle to be considered by every institution to increase the control over educational data.

There is a need for a proposal that can draw a relationship between technologies and law in equal terms, considering a complementary ethic approach, to solve the needs for storing, analyzing and sharing educational data over a privacy- and security-compliant environment. Hence, we thought of a technological solution that can be adapted to present and future problems, incorporating premises related to privacy and security of data protection, legal protection and ethics by default and by design.

We introduce such a proposal in this document. The second section presents a replicable solution in detail and how this architecture aims to protect educational data. The third section describes some case uses for this solution deployed at La Salle, Universitat Ramon Llull (La Salle–URL). Finally, we discuss and expose ongoing projects and future lines.

2. Materials and Methods

2.1. Educational Warehouse

We chose the term “educational warehouse” for the fusion of EdTech and LegalTech. It can be defined as a modular platform for educational data analytics, with a “local first” approach architecture but scalable and cloudable if needed. Hence, we found it necessary for the solution to have the following technical features:

- to be modular;
- to permit both a rigid and hybrid scenario (local or cloud computing) but with a “local first” approach;
- to use a decentralized scheme, totally or partially;
- to be scalable in terms of volume of data, processing capacity, and public and private access;
- to be technically adaptable to any educational context;
- to be private, permissioned and temporalized; and
- to automate law compliance to treat safety and privacy as transversal axes of the solution by default and by design.

The goal of our proposal is to build a context of confidence and absolute control over data. For this reason and taking the technical features into account, it acts as a digital entity that should offer the following:

- total control over data, both modular and local, once they get into the architecture. Our modular architecture performs as a data bunker with private, permissioned access. A modular architecture permits the distribution of responsibilities, privatizing access to known users and creating a taxonomy of roles and temporary permissions for acting in different parts of the architecture to limit the power of access. Moreover, the “local first” principle put in the first instance places this solution on the opposite side of cloud computing, hence ensuring an airtight space. In short, it provides local control of who or what has access to data, what for and for how long.
- a set of acceptable practices. Our modular architecture facilitates the integration of good practices further than required by law and in compliance with each institution’s morals and ethics. We will later describe a series of acceptable practices associated with each module of this architecture.

Therefore, we propose an architecture that facilitates regularizing and defining data management processes to ensure adequate privacy and security levels through technological automation of law and ethics.

2.1.1. Basic System Architecture

Educational institutions are continually generating data. Some of these data result from academic management processes such as administrative procedures, qualifications or certifications. In contrast, others result from the teaching–learning processes, such as feedback related to student’s tasks or comments about their academic behavior. The analytic approaches to educational data must fulfil the requirements of both kinds of data. The origin of this data can be either internal or external depending on the EdTech services that are applied [61]. For example, in times of evaluation, there is a need to know the student’s status and other data coming from different academic, learning and behavioral interactions; this data can be hosted both internal platforms and in third-party services available in the company’s cloud.

In the same way, educational institutions can benefit from data generated by other institutions, that is, the case of research activities, where quality data are fundamental for the correct generation of results. Educational institutions generate a series of anonymous data that can even be considered open data. Those datasets, once depersonalized, are of great value for research. Moreover, institutions are obliged to share data with public administration and government services. The possibility of automating those obligations, liberating tasks now done by machines, facilitates data exportation or even the interoperable connection between the institution and the administration.

Based on those mentioned above, we designed a modular architecture that allows for data import, storage, analysis and even external access. This modular characteristic facilitates such scenery, being flexible, adaptable, decentralizable and scalable. It is flexible and adaptable so the different modules can be implemented using the technology required in every case and can be located either inside the institution, in the cloud or in a mixed local–cloud implementation. This arrangement permits adaptation to every institution’s needs or even decentralization. Decentralization may allow the replication of the primary system’s architecture and may modulate it as needed, whereas scalability would take place at a module level in conjunction with cloud computing.

Figure 1 describes the system architecture layers implementing the aforementioned characteristics.

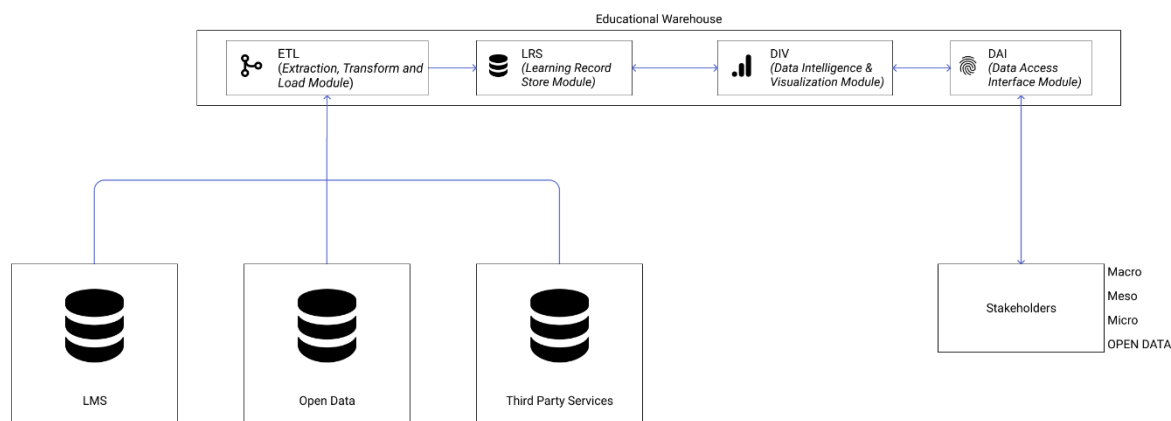


Figure 1. The educational warehouse architecture: data are exchanged from different sources to a collector module where it can be processed and then sent into record storage. When there is a petition to analyze data, the Data Intelligence and Visualization module (DIV) accesses the storage to obtain the required data and stores its results. To make the request, the stakeholders use the data access interface module.

Each educational warehouse integrates the following essential modules: Extract, Transform and Load (ETL), Learning Record Store (LRS), Data Intelligence and Visualization (DIV) and Data Access Interface (DAI). It is possible to customize them internally or to even add new ones according to each institution's requirements and needs. Some modules can be disabled if necessary. For each module, we proposed a series of possible actions as an example of good practice to ensure PICSPDM. As aforementioned in Section 1.5, each module may be affected by the local jurisdiction. Therefore, the technical actions and good practices associated with the ethical-legal framework should be adequate ones as determined by law.

- **ETL:** This is the module that allows importing data from abroad. It contains the software and hardware in charge of importing data from different data sources (management and educational software of the institution, EdTech, LMS, etc.). The content imported to the ETL module can be of an encrypted and anonymized nature. Once stored, it cannot be identified by any entity but the users involved, i.e., using Pretty Good Privacy (PGP) encryption technology [62]. It can also be managed securely by employing secure connections, such as Secure Shell (SSH), Secure Socket Layer (SSL) or Virtual Private Network (VPN). Regarding data management, data registries can be transferred entirely and removed from their origin, just as data from users that have not offered or consented to processing their data can be excluded as well. In the LMS or EdTech, the educational warehouse administrators may not have management competencies. However, they may demand the same good practices to ensure PICSDMPD or other more convenient ones. For instance, in the Moodle LMS, some of the authors managed to encrypt the user table since most of the personal data reside there and used a set of views that allow decryption or encryption at convenience. This measure can be complemented by ensuring students' anonymity through second identities with the Protected Users plugin [32], thus not exporting data that can be used to identify people. The entire database can also be encrypted with a double user control system where not even the administrator can access or decrypt the data. Some of the authors of this paper conceived such a system and named it AuthChecker [35]; simple authentications, private and permissioned Application Program Interfaces (API), and Learning Tools Interoperability (LTI) technologies are recommended to ensure data access by manually specified roles.
- **LRS:** This is the module where data are stored. It contains all that software and hardware needed to store the data in the original format or as the result of transformations and analyses. The private and secure nature of data can be guaranteed by defining temporary and regulated accesses, with users generated by a previous contract with

the institution; by storing encrypted data; by or applying change control protocols (integrations through high-level log systems, definition of non-editable data, etc.), among other practices.

- **DIV:** This module analyses and visualizes information internally. It contains all the software and hardware in charge of performing the institution's analytical approaches and the data visualization tools. Information given to those systems can be pseudo-anonymized or anonymized. Their response, being equally masked, can be integrated into the LRS module as far as no correlation of data with real identities is guaranteed. Modularity allows LRS to be hosted locally while DIV resides in cloud computing, making it necessary to encrypt the roaming data or to even use private connections such as SSH, SSL or VPN.
- **DAI:** This module controls access to data according to legal regulations. It contains all the software and hardware that permits internal solutions and other educational institutions, government and administrations, partners and third-party collaborators to access data under a private and regulated regime. At the same time, it could serve as open data for public benefit. It manages access to the information securely. A set of essential, contractual permissions can be defined automatically in the user creation process. This is the case of accessing one's action history or the acceptance or rejection of data transfer agreements. Each user profile may request an extension of its capabilities under a manual review process; this process may generate a set of legal bonds that will detail the relationship between the user and the system. The reviewers, i.e., the data privacy officer in the case of GDPR, may reject the request. A set of permissions that do not require an extensive legal bond can also be offered using under-request automatic approval procedures, allowing easy access to research data and collaboration for open data consortiums.

Stakeholders play an essential role in the design of the architecture of the system. They somehow help to shape it in terms of data access. From the perspective of analytics, we make a category of stakeholders [63,64] that allows us to embrace the most critical roles in teaching–learning processes, management and research:

- **Macro-level analytics** tries to make analytics (data) accessible between institutions and third parties. For instance, this level's objective is for educational institutions to be able to access statistical data from state exams carried out by students throughout their lives or to facilitate administrations to be able to access statistical data from institutions. It is fed with the information generated by the meso and micro sublevels. The results at this level are reflected in a possible transformation of the institutional system (school, university, educational organization, etc.) and changes in academic models or pedagogical approaches.
- **Meso-level analytics** operates at the institution level through business intelligence. The objective of this level is, among other institutional aspects, the improvement of the different educational processes at the institutional level and strategic business decision-making, for example, to identify those courses that are more effective or functional. Its benefits are linked to the optimization of decision-making at the administrative level, the increment of educational "production" and even improvements in resource allocation.
- **Micro-level analytics:** This level addresses the analysis of the interactions carried out by every student, both independently and as a part of a group. These analytical processes include personal and sensitive student data, such as book loans, geolocations, financial data, social media conversations or even the clickstream of virtual learning environments. The benefits of this level result in a system that can identify students at risk, alert of possible dropouts and even provide students with conclusions and advice that can help them improve. The micro level is intended to be introduced to coordinators and instructors who deliver content to students and evaluate their work.

- Open-data analytics: all educational institutions are likely to generate anonymized data. Making them public in a raw format or even in a processed format evokes a desire for transparency and open knowledge that can be advantageous to society. This open level of data is accessible to any independent person, research group, third party or citizen who requires access to data that educational institutions, especially public ones, can make available without violating student data privacy or security.

Modularity constitutes the alma mater of the educational warehouse architecture. We are not the first to propose an architecture of this kind, but our efforts to define a flexible architecture that stands up for data privacy and security goes much further than the previous existing solutions.

Firstly, it is differentiated from traditional database management in that it adds a data management scheme that includes several layers of privacy and security. It is also possible to distribute the architecture partially between the institution and cloud computing. It also fosters research by allowing data access under open data consortiums or by accessing data with API systems further than merely creating and sending Comma Separated Values (CSV) files or datasheets under request. Although it adds some complexity to data management, it can perform the same operations as those of a traditional architecture database. However, the set of technical actions and good practices in a modular structure define a completely different mode of operation.

Secondly and referring to authors who previously proposed similar architectures, our proposal evolves in some ways from those previous ones. Aziz et al. proposed in 2014 [65] a linear architecture to carry out business intelligence over educational data. Their educational data warehouse solution has similarities to our LRS module. Flanagan and Ogata 2017 [66] also proposed a linear typology, centered on learning analytics but not yet implemented unlike ours and constrained to a mode of operation based on unique identifiers, where data entered the system from LMSs and ended in a user dashboard in a closed environment. Our solution is more flexible regarding technical proposals since it can even apply those by Aziz et al.; by Flanagan and Ogata; by other analytical approaches beyond learning analytics; and above all, by using nonlinear open topology. We refer to a linear typology where the proposed architecture has a start in the data input and an end in the user reports. The architecture that we offer, as shown in Figure 2, adapts to any typology where the data output from an educational warehouse can be the input for another educational warehouse. To further exemplify how to organize a topology with an educational warehouse, we use a network architecture simile. Each node is represented by an educational warehouse connected with others in a star, ring, tree or any other required topology, thus creating a multidirectional interconnected network of educational warehouses, emphasizing educational data privacy and security.

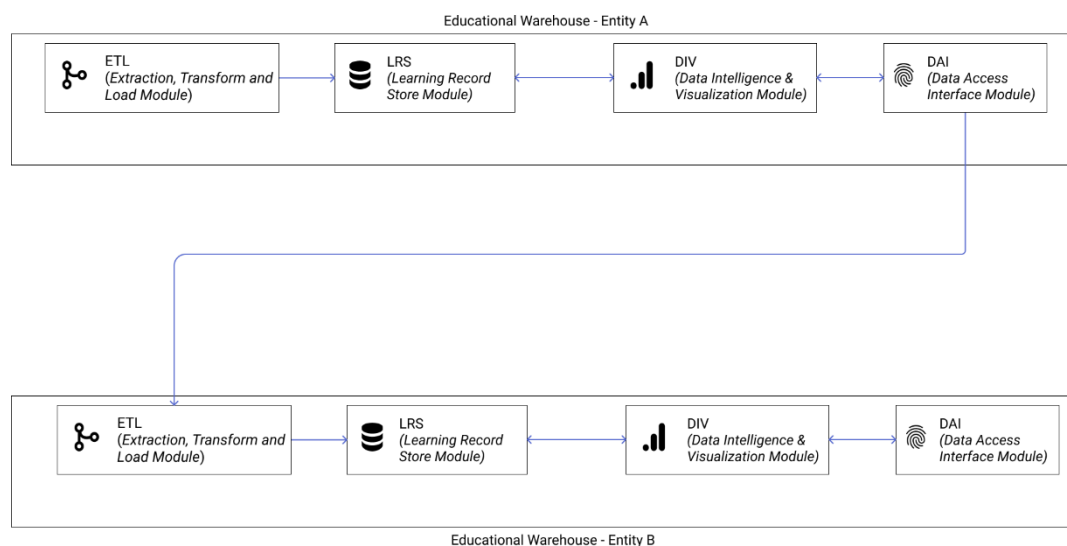


Figure 2. Interinstitutional use of an educational warehouse: data are exchanged between entity A and entity B, where entity A's Data Access Interface (DAI) is the data origin of Entity B's Extract, Transform and Load (ETL).

2.1.2. Use Cases

This proposal is set out to be used for different purposes. In this paper, we present different approaches where an educational warehouse can facilitate the management of educational data. Specifically, we propose three use cases, although the implementation of an educational warehouse is not limited to them.

The first case that we present describes a development carried out at La Salle-URL, where all modules are used and new features are developed. We implemented this structure in a local environment to regulate data access securely, controlling the environment and its access. Since needs are growing, the modules described in the architecture are made in different working environments to facilitate other teams' development.

At a practical level, the educational warehouse state implemented at La Salle-URL is described in Figure 3. Data are extracted from educational tools such as Moodle (LMS), third-party applications (such as Kahoot [67]) and public access (Open Data). Data are processed to reduce their volume and to get a better understanding of their nature. Once this data has been understood and adapted, it is stored in the storage module. Data related to the operation of Moodle and data associated with the validation of an educational method used in La Salle-URL [68] are currently stored. If a request is made to access the data, an access gateway or a client is enabled so data can be accessed securely. This case is presented in detail in the third section.

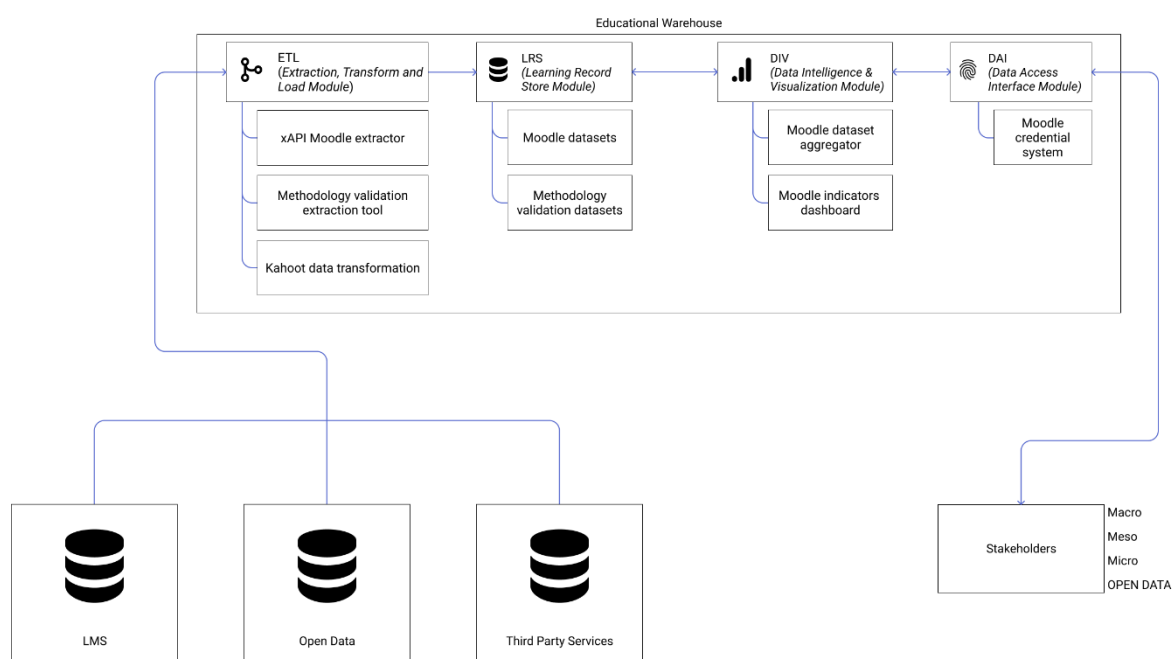


Figure 3. The educational warehouse architecture applied at La Salle-URL: representation of the current modules developed.

A second possible case is based on providing and managing data for research teams, which require information coming from different entities. This information may be regulated under an agreement between entities or based on open data between educational entities. A summary of the required modules can be seen in Figure 4.

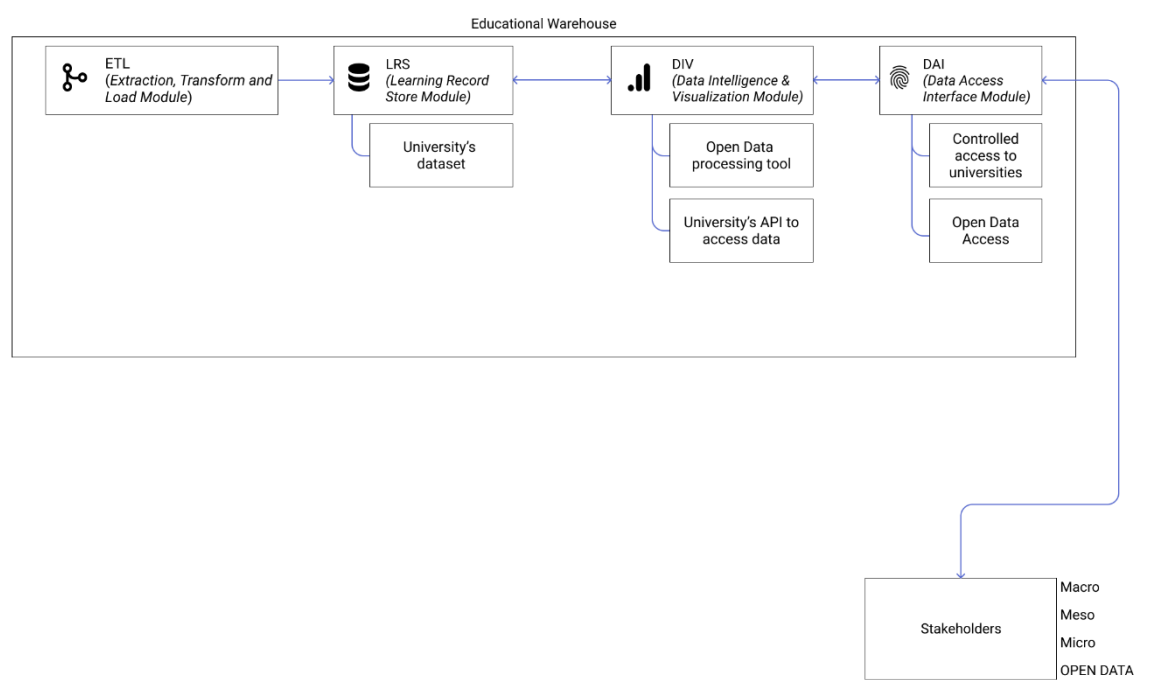


Figure 4. The educational warehouse architecture proposal for research entities working with open data: the modules represented offer the ability to progress towards the study's objective.

This second case shows how an educational data management system allows for creating synergies in research, as it offers the possibility to create a collaborative and easily accessible environment. To comply with good practice requirements, a researcher should only adopt the educational warehouse architecture model and connect the DAI and ETL modules as the output and input gate. We believe that fields such as educational data mining or learning analytics can benefit from the synergies offered by different educational entities' collaboration. They allow the state-of-the-art to remain accessible to those interested or involved.

Finally, we present a third possible case, now aimed at institutions with limitations to implement an educational warehouse at the local level. Given this architecture's modular design, it remains functional as long as the different modules operate and the established privacy principles are respected. This means that its implementation, if desired, can use cloud computing, with the benefits that this technology brings and taking into account the risks involved. In the same way, a module can be fragmented, as it is possible that at a given point there is a volume of data large enough to require the application of big data techniques in cloud computing to store and analyze, such as MapReduce, clustering or similar [69]. Components of the same module can be located in different computing environments, as shown in Figure 5.

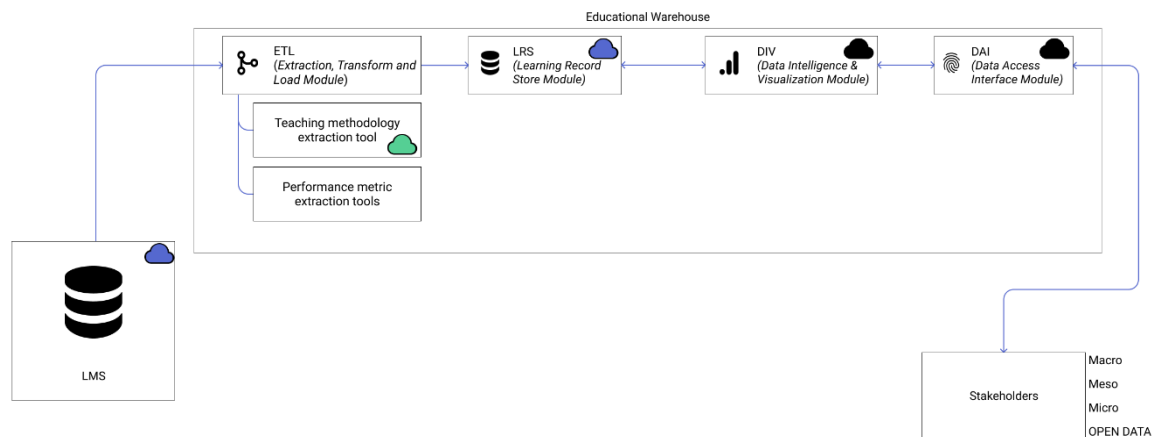


Figure 5. The educational warehouse architecture proposal for a cloud-based deployment: each colored cloud represents a different environment. The lack of a cloud symbol means that it is a local component.

These three use cases, one of them already in execution and the other two drawing possible scenarios, demonstrate a solution that institutions such as universities can use to facilitate the administration, processing and regulation of educational data. The educational warehouse also permits collaboration in joint research and an easy way to regulate access to data, maintaining the balance between technology and the legal system.

3. Results

La Salle–URL implemented an initial version of the educational warehouse which permits the development of different research lines. The implementation of the architecture is progressing and being adopted in several in-house projects. It follows an end-to-end path gradually; initially, the existing data are understood and then stored so researchers can access them. Finally, tools are enabled for analysis and evaluation.

One of the projects is based on the extraction of indicators to validate the viability of the educational methodology that was applied in La Salle, Self Directed Based Learning (SDBL) [68]. Using the educational warehouse structure, data were extracted, transformed and adapted to a subsequent local-level analysis database.

Another ongoing research project aims to extract generalized indicators of LMS user interaction with the educational warehouse to its fullest, applying all ETL, LRS, DIV and DAI modules.

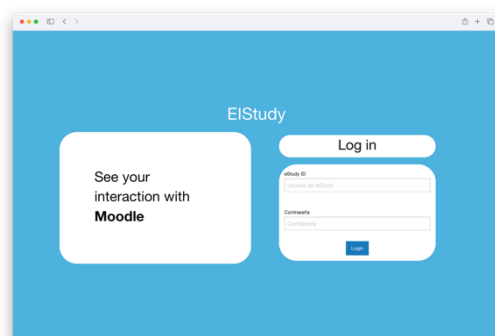
Moodle's Executive Interaction Board

A third project under development, named EISStudy, implements a tool to describe the students' behavior in a Moodle-based LMS. This tool provides the users of the virtual campus at La Salle–URL with a depiction of their interaction with the virtual facility. At the same time, it offers an API for the researchers to access those datasets. All functionalities of EISStudy follow a series of rules implemented in the modules of the educational warehouse architecture.

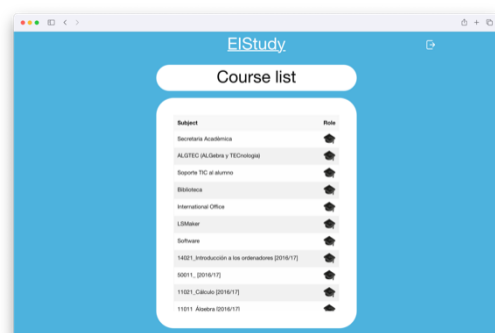
- ETL and LRS modules: Initially, the loaded data are based on the Moodle reports. Thanks to Moodle's relational model, information can be extracted that describes user's interactions with the platform using different levels of detail.
- DIV module: This solution offers a micro-level analysis using students' highly detailed, sensitive, personal and behavioral data. The data are anonymized to preserve the student's identities while presenting related information.
- DAI module: This interface's development to access data considers the LMS user roles (see Figure 6). The user is offered a set of indicators depending on whether they are a teacher or a student (see Figure 7). A teacher had access to viewing the data of

the students in the course. A student has access to only their data but not that of other classmates in the course, thus avoiding privacy conflicts.

Thanks to this tool, it is possible for teachers to detect all interactions with resources related to the subjects they teach and to adapt their teaching process. Students are also allowed, in this case, to be aware of their own interaction with the platform, helping them to understand what kind of situation they may face passing a subject or asking the teacher for help based on the information displayed.



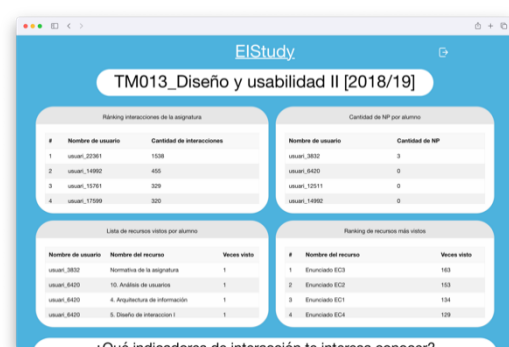
(a)



Subject	Role
Secretaría Académica	
ALISTEC (ALGebra y TECnologías)	
Seguimiento TIC al alumno	
Biblioteca	
International Office	
Libreros	
Software	
14201_Introducción a las ordenadores [2016/17]	
80011_ [2016/17]	
11021_Cálculo [2016/17]	
11011_Análisis [2016/17]	

(b)

Figure 6. (a) EISStudy's client login form: the user must log in using their credentials to get data related to his or her profile. If there is a failure to log in, no data will be sent to the user. If it works (b), it will enable the petition that sends details on the courses the user is enrolled in.



(a)



(b)

Figure 7. (a) EIShudy’s visualization tool. Once the user creates a request to the educational warehouse modules, it offers data related to the specified content and its role. If the user is a teacher, he or she will be able to see “individual” values and “group” values, whereas students will only see “individual” values since they are not allowed to see data from other people in that context (b).

4. Discussion

Sentences such as “data are the new oil” or concepts such as “data capitalism” [16,70] highlight the importance of data as drivers of our knowledge-based society. From time immemorial, data has been a source of power and knowledge. That is why the first digitization of information and subsequent automation were so important in the context of businesses and in the beginning of the conception of computers [71]. In the business world, extracting or discovering patterns from data is useful to increase productivity and as a competitive advantage. Business intelligence [72] was born from these first computerized data analyses, which evolved as a series of systematized processes to exploit data both analytically and visually in search of relevant information.

Al Gore’s “Information highways” [73], the digital interconnection of all continents [74] and the technological evolution of telecommunications have allowed business intelligence since the 50s [72] to take one significant step forward. These early technological revolutions in telecommunication together with the adoption of Internet-web technology at all levels systematically generate enormous amounts of data, thus pushing the emergence of big data technologies [1,9]. Big data is emerging in society and consolidating a fourth revolution where the Internet of Things functions as a gateway for the interconnection of virtual and physical space, blurring both realities. There is little or no differences between real and digital identity.

Regarding previous innovations, entities in the educational context are taking part in this revolution by following the same evolutionary pattern. First, Internet-web technology is being adopted when virtual learning environments or learning management systems become relevant, and then, analytical approaches such as educational data mining [75] or

learning analytics [76] and their intersection [77,78] are adopted. Finally, big data and artificial intelligence technologies [79], such as facial recognition or brainwave analysis to know students' attention or emotions are implemented. The Internet of Things is an integration, and we already see functional solutions that combine big data, machine learning, artificial intelligence and cloud computing [6].

However, the adoption of Internet-web technology as a means of communication has also set out exponentially, reflecting what Tim Berners Lee stated in his 1996 Request For Comments (RFC) [80]: "The Referer field allows reading patterns to be studied and reverse links drawn. Although it can be beneficial, its power can be abused if user details are not separated from the information contained in the Referer." Nowadays, lots of data types are collected to improve services as commodities for society and the educational context, risking the privacy and security of people's data and, consequently, both their real and digital identity. This has been the case in recent misuses, data leaks and improper access due to the high value of data in data capitalism platforms [70] where privacy and security are at risk because of daily surveillance [15]. In terms of the educational context, maintaining the privacy, security and digital identity of students and their data is even more critical. In many cases, data related to minors are also involved [81]. Although data analysis could bring benefits, there is an intense fragility in managing privacy and security of students' data in interconnected virtual environments and in exposing their real identity to different hazards, such as cyber-bullying or virtual harassment [32].

We need to become aware of this positive–negative dual context and develop solutions that address the different social issues beyond legality. Legality taken as the new ethic allows us to evolve obscure corners of humanity. It is the case of the mass failing grades of British university students deprived of access to the desired studies (A-levels) [30] or mass deportations of students for alleged copies in linguistic examinations that were evaluated automatically by algorithms full of legal rules (ETS-TOEIC) [82]. The limits of legality and technology, such as in biased results or surveillance technologies, are set by ethics.

As authors and citizens, we are aware of the data problems in educational terms, but at the same time, we know the benefits of processing this data. We believe that legality and ethics can be automated in technology. We must avoid using the law as the new ethic and complement the triad of legality, ethics and technology as the basis for real solutions. In this sense, it is possible to make the laws, corrective at first, preventive.

Cloud computing is a double-edged sword. On the one hand, it facilitates the collection and computation of massive data. On the other hand, it facilitates the exposure of students. We echo this and propose a framework of 7 principles called LEDA to minimize this exposure, prioritizing local technologies and leaving cloud computing as a last resource [60].

In this work, we take a new step and propose a system architecture solution that complements the LEDA framework and aims to achieve a law/ethics balance in the educational context mediated by technologies. The system architecture that we propose is modular. We name it educational warehouse, as it is intended to collect, analyze and manage access to educational data of an institution.

Modularization allows educational institutions to adapt the system to their peculiarities. We define four basic modules of the system: one to import the data, one to store raw and processed data, one to perform the analysis, and another one to allow regulated access from the outside. Without imposing any specific storage technology, software or hardware, we leave total freedom to institutions to integrate, for example, free or proprietary software or storage solutions in an open format such as xAPI [83] and/or to complement it with relational or documentary databases, hosting servers within the same institution or using cloud hosting at those points it deems relevant; enabling or disabling external access; and scaling modularity by replicating the system as islands within the same insti-

tution or interconnecting institutions or third parties such as other universities, educational services, administrations and governments, or even individuals if the educational institution deems it appropriate to publish its data in the open.

As a part of the new pedagogical change that the NCA project [84] is drawing up at the La Salle institution, we implement this solution to provide technical support while complying with the law and enforcing Lasallian ethics. We carried out further developments that we presented in the Results section, where the educational warehouse has allowed us to modulate and continue to develop new interconnected solutions to facilitate and enhance teaching and learning processes.

Our system architecture proposal fosters new research lines to identify new interoperability options between tools and institutions, to identify new storage formats to facilitate the analysis of data, to define machine learning models from specific educational data and LMS, to establish levels of data access in relation to legality or to define good data analysis practices preserving privacy and security of students' data.

Author Contributions: Conceptualization, D.A.; methodology, D.A.; solution analysis, D.A. and P.G.; development, P.G.; investigation, D.A.; resources, D.F. and L.H.-I.; writing—original draft, all authors; writing—review and editing, all authors; visualization, D.A. and P.G.; project administration D.A. and L.H.-I.; funding acquisition, D.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was co-funded by the support of the Secretaria d'Universitats i Recerca of the Department of Business and Knowledge of the Generalitat de Catalunya with the help of 2017 SGR 934.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Marciano, A.; Nicita, A.; Ramello, G.B. Puzzles in the big data revolution: An introduction. *Eur. J. Law Econ.* **2020**, *1*–6, doi:10.1007/s10657-020-09672-4.
2. Gong, C.; Liu, J.; Zhang, Q.; Chen, H.; Gong, Z. The characteristics of cloud computing. In Proceedings of the International Conference on Parallel Processing Workshops, San Diego, CA, USA, 13–16 September 2010; pp. 275–279.
3. Pallis, G. Cloud computing: The new frontier of internet computing. *IEEE Internet Comput.* **2010**, *14*, 70–73, doi:10.1109/MIC.2010.113.
4. Chang, V.; Bacigalupo, D.; Wills, G.; De Roure, D. A categorisation of cloud computing business models. In Proceedings of the CCGrid 2010—10th IEEE/ACM International Conference on Cluster, Cloud, and Grid Computing, Melbourne, VIC, Australia, 17–20 May 2010; pp. 509–512.
5. Marston, S.; Li, Z.; Bandyopadhyay, S.; Zhang, J.; Ghalsasi, A. Cloud computing—The business perspective. *Decis. Support Syst.* **2011**, *51*, 176–189.
6. Qiu, T.; Chi, J.; Zhou, X.; Ning, Z.; Atiquzzaman, M.; Wu, D.O. Edge Computing in Industrial Internet of Things: Architecture, Advances and Challenges. *IEEE Commun. Surv. Tutor.* **2020**, doi:10.1109/comst.2020.3009103.
7. Morrar, R.; Arman, H.; Mousa, S. The fourth industrial revolution (Industry 4.0): A social innovation perspective. *Technol. Innov. Manag. Rev.* **2017**, *7*, 12–20.
8. Perc, M.; Ozer, M.; Hojnik, J. Social and juristic challenges of artificial intelligence. *Palgrave Commun.* **2019**, *5*, 1–7, doi:10.1057/s41599-019-0278-x.
9. Bechmann, A.; Kim, J.Y. Big Data. In *Handbook of Research Ethics and Scientific Integrity*; Iphofen, R., Ed.; Springer International Publishing: Cham, Switzerland, 2020; pp. 427–444. ISBN 978-3-030-16759-2.
10. Dixon-Román, E.; Parisi, L. Data capitalism and the counter futures of ethics in artificial intelligence. *Commun. Public* **2020**, doi:10.1177/2057047320972029.
11. Coghlan, S.; Miller, T.; Paterson, J. Good proctor or “Big Brother”? AI Ethics and Online Exam Supervision Technologies. *arXiv* **2020**, arXiv:2011.07647.
12. Michie, D.; Spiegelhalter, D.J.; Taylor, C.C. Machine learning. *Neural Stat. Classif.* **1994**, *13*, 1–298.
13. Murphy, B.; Rocchi, M. Ethics and Cloud Computing. In *Data Privacy and Trust in Cloud Computing*; Palgrave Macmillan: Cham, Switzerland, 2021; pp. 105–128.

14. Data Privacy and Trust in Cloud Computing. Available online: <https://library.oapen.org/handle/20.500.12657/42922> (accessed on 30 November 2020).
15. Bellamy Foster, J. Surveillance Capitalism. *Mon. Rev.* **2014**, *66*, 1–31.
16. West, S.M. Data Capitalism: Redefining the Logics of Surveillance and Privacy. *Bus. Soc.* **2019**, *58*, 20–41, doi:10.1177/0007650317718185.
17. Zuboff, S. Surveillance Capitalism and the Challenge of Collective Action. In *Proceedings of the New Labor Forum*; SAGE Publications: Los Angeles, CA, USA, 2019; Volume 28, pp. 10–29.
18. Onik, M.M.H.; Kim, C.S.; Yang, J. Personal Data Privacy Challenges of the Fourth Industrial Revolution. In *Proceedings of the International Conference on Advanced Communication Technology, ICACT*, PyeongChang, Korea, 17–20 February 2019; Volume 2019, pp. 635–638.
19. Zuboff, S. Big other: Surveillance capitalism and the prospects of an information civilization. *J. Inf. Technol.* **2015**, *30*, 75–89, doi:10.1057/jit.2015.5.
20. Shahroom, A.A.; Hussin, N. Industrial Revolution 4.0 and Education. *Int. J. Acad. Res. Bus. Soc. Sci.* **2018**, *8*, 314–319, doi:10.6007/ijarbss/v8-i9/4593.
21. Celaa Saluda la Alianza de Grandes Empresas para Innovar en la Educación. Available online: <https://www.lavanguardia.com/vida/20201029/4968492825/celaa-saluda-la-alianza-de-grandes-empresas-para-innovar-en-la-educacion.html> (accessed on 30 November 2020).
22. Realising the Potential of Technology in Education. Available online: <https://www.gov.uk/government/publications/realising-the-potential-of-technology-in-education> (accessed on 30 November 2020).
23. Ainley, P. The new ‘market-state’ and education. *J. Educ. Policy* **2004**, *19*, 497–514.
24. Mathur, K. State, Market and Governance of Higher Education. *Gov. Manag. High. Educ. India* **2020**, 25–43.
25. Nazarenko, M.A.; Khronusova, T.V. Big data in modern higher education. Benefits and criticism. In *Proceedings of the 2017 International Conference “Quality Management, Transport and Information Security, Information Technologies” (IT and QM and IS)*, St. Petersburg, Russia, 24–30 September 2017; pp. 676–679.
26. Picciano, A. Big Data and Learning Analytics in Blended Learning Environments: Benefits and Concerns. *Int. J. Interact. Multimed. Artif. Intell.* **2014**, *2*, 35, doi:10.9781/ijimai.2014.275.
27. Ekong, E.E.; Adiat, Q.E.; Ejemeyovwi, J.O.; Alalade, A.M. Harnessing big data technology to benefit effective delivery and performance maximization in pedagogy. *Int. J. Civ. Eng. Technol.* **2019**, *10*, 2170–2178.
28. Cavazos, J.G.; Jonathon Phillips, P.; Castillo, C.D.; O’Toole, A.J. Accuracy comparison across face recognition algorithms: Where are we on measuring race bias? *arXiv* **2019**, arXiv:1912.07398.
29. Schrag, F. *Review of Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*; Broadway Books: New York, NY, USA, 2017; Volume 24.
30. Lessons from the A-Levels Fiasco: Putting Culture and Values at the Heart of Policymaking. Available online: <https://blogs.lse.ac.uk/politicsandpolicy/culture-values-policymaking/> (accessed on 10 September 2020).
31. Amo Filva, D. Privacidad y Gestión de la Identidad en Procesos de Análisis de Aprendizaje. Programa de Doctorado Formación en la Sociedad del Conocimiento. Available online: <https://repositorio.grial.eu/handle/grial/1951> (accessed on 31 June 2020).
32. Amo, D.; Alier, M.; García-Peñalvo, F.J.; Fonseca, D.; Casañ, M.J. Protected Users: A Moodle Plugin to Improve Confidentiality and Privacy Support through User Aliases. *Sustainability* **2020**, *12*, 2548, doi:10.3390/su12062548.
33. Amo, D.; Alier, M.; García-Peñalvo, F.; Fonseca, D.; Casañ, M.J. Privacidad, seguridad y legalidad en soluciones educativas basadas en Blockchain: Una Revisión Sistemática de la Literatura. *RIED. Rev. Iberoam. Educ. Distancia* **2020**, *23*, 213, doi:10.5944/ried.23.2.26388.
34. Amo, D.; Alier, M.; Fonseca, D.; García-Peñalvo, F.-J.; Casañ, M.J.; Navarro, J. Evaluation of the importance of ethics, privacy and security in Learning Analytics studies, under the LAK conferences. In *Proceedings of the CINAIC*, Madrid, Spain, 9–11 October 2019.
35. Amo, D.; Alier, M.; García-Peñalvo, F.J.; Fonseca, D.; Casany, M.J. GDPR security and confidentiality compliance in LMS’ a problem analysis and engineering solution proposal. In *Proceedings of the ACM International Conference Proceeding Series*; Conde, M.Á., Rodríguez, F.J., Fernández, C., García-Peñalvo, F.J., Eds.; ACM: New York, NY, USA, 2019; pp. 253–259.
36. Pearson Data Breach: Details of Hundreds of Thousands of U.S. Students Hacked. *Fast Co.* Available online: <https://www.fast-company.com/90384759/pearson-data-breach-details-of-hundreds-of-thousands-of-u-s-students-hacked> (accessed on 20 February 2019).
37. Drachsler, H.; Greller, W. 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, Scotland, 25–29 April 2016; pp. 89–98.
38. US Schools Have Lost 24.5 Million Records in Breaches Since 2005. Available online: <https://www.techrepublic.com/article/us-schools-have-lost-24-5-million-records-in-breaches-since-2005/> (accessed on 3 July 2020).
39. Student Loans Company criticised over data breaches. *BBC News*. Available online: <https://www.bbc.com/news/education-27587152> (accessed on 3 July 2020).
40. Una Brecha de Datos deja Expuesta Información Personal de Millones de Padres y Alumnos de Madrid. *elplural.com*. Available online: https://www.elplural.com/autonomias/fallo-informatico-deja-expuestos-datos-personales-millones-padres-alumnos-madrid_252050102 (accessed on 21 November 2020).

41. El Web de la Generalitat Exposa més de 5.000 Dades que Inclouen Correus i Contrasenyes. *elMón*. Available online: <https://elmon.cat/politica/web-generalitat-exposa-5000-dades-inclouen-correus-contrasenyes-187168/> (accessed on 30 November 2020).
42. Amo, D.; Alier, M.; Casan, M.J.; Casañ, M.J. The student's progress snapshot a hybrid text and visual learning analytics dashboard. *Int. J. Eng. Educ.* **2018**, *34*, 990–1000.
43. Lafarre, A.; Van der Elst, C. Legal tech and blockchain for corporate governance and shareholders. In *Research Handbook in Data Science and Law*; Edward Elgar Publishing: Cheltenham, UK, 2018; ISBN 9781788111294.
44. Tracol, X. "Schrems II": The return of the Privacy Shield. *Comput. Law Secur. Rev.* **2020**, *39*, 105484, doi:10.1016/j.clsr.2020.105484.
45. Schrems, M. The Privacy Shield is a Soft Update of the Safe Harbor. *Eur. Data Prot. Law Rev.* **2017**, *2*, 148–150, doi:10.21552/edpl/2016/2/4.
46. Tzanou, M. Schrems I and Schrems II: Assessing the Case for the Extraterritoriality of EU Fundamental Rights. *Data Prot. Beyond Borders Transatl. Perspect. Extraterritoriality Sover. Hart Publ. Forthcom.* **2020**, 1–21. Available online: <https://ssrn.com/abstract=3710539> (accessed on 15 January 2021).
47. Kuner, C. Reality and Illusion in EU Data Transfer Regulation Post Schrems. *SSRN Electron. J.* **2018**, *18*, 881–918, doi:10.2139/ssrn.2732346.
48. Greenleaf, G.; Systems, I. International Data Privacy Agreements after the GDPR and Schrems. *Priv. Laws Bus. Int'l Rep.* **2016**, *139*, 12–15.
49. Bräutigam, T. The Land of Confusion: International Data Transfers between Schrems and the GDPR. *Data Prot. Priv. Eur. Regul. Digit. Age* **2016**, 143–177. Available online: <https://ssrn.com/abstract=2920181> (accessed on 15 January 2021).
50. Weiss, M.A.; Archick, K.U.S. EU data privacy: From safe harbor to privacy shield. *Eur. Union Chall. Prospect.* **2016**, 113–135.
51. Regulation (EU) 2016/679 GDPR. Available online: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (accessed on 27 June, 2019).
52. Rothstein, M.A.; Tovino, S.A. California Takes the Lead on Data Privacy Law. *Hastings Cent. Rep.* **2019**, *49*, 4–5, doi:10.1002/hast.1042.
53. Data Protection & Privacy. Available online: <https://practiceguides.chambers.com/practice-guides/data-protection-privacy-2020/australia> (accessed on 30 November 2020).
54. New Analysis Shows How Facebook and Google Push Users into Sharing Personal Data. Available online: <https://www.forbrukerradet.no/side/facebook-and-google-manipulate-users-into-sharing-personal-data/> (accessed on 30 November 2020).
55. New Study: Google Manipulates Users into Constant Tracking. Available online: <https://www.forbrukerradet.no/side/google-manipulates-users-into-constant-tracking/> (accessed on 30 November 2020).
56. Inspectorate, N.D. *Use of Google Chromebook and G Suite for Education (and other Cloud Services) in Elementary School*; 2020.
57. Buchholtz, G. Artificial Intelligence and Legal Tech: Challenges to the Rule of Law. In *Regulating Artificial Intelligence*; Wischmeyer, T., Rademacher, T., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 175–198. ISBN 978-3-030-32361-5.
58. Hildebrandt, M. 10. 'Legal by Design' or 'Legal Protection by Design'? In *Law for Computer Scientists and Other Folk*; Publisher: Oxford University Press, Oxford, UK, 2019; pp. 251–283.
59. Hildebrandt, M.; Tieleman, L. Data protection by design and technology neutral law. *Comput. Law Secur. Rev.* **2013**, *29*, 509–521, doi:10.1016/j.clsr.2013.07.004.
60. Amo, D.; Torres, R.; Canaleta, X.; Herrero-Martín, J.; Rodríguez, C.; Fonseca, D. Seven principles to foster privacy and security in educational tools: Local Educational Data Analytics. In the Technological Ecosystems for Enhancing Multiculturality, TEEM'20, Online Conference, 21–23 October 2020. Available online: <https://2020.teemconference.eu/papers-in-video/> (accessed on 23 October 2020).
61. Amo, D.; Fonseca, D.; Alier, M.; García-Peñalvo, F.J.; Casañ, M.J.; Alsina, M. Personal Data Broker: A Solution to Assure Data Privacy in EdTech. In Proceedings of the HCI: International Conference on Human-Computer Interaction. Learning and Collaboration Technologies, Orlando, FL, USA, 26–31 July 2019; Designing Learning Experiences; Zaphiris, P., Ioannou, A., Eds.; 2019; Volume 11590 LNCS, pp. 3–14.
62. Zimmermann, P.R.; Zimmermann, P.R. *The Official PGP User's Guide*; MIT Press: Cambridge, UK, 1995; Volume 5.
63. Prinsloo, P. Stuck in the middle? Making sense of the impact of micro, meso and macro institutional, structural and organisational factors on implementing learning analytics. In Proceedings of the Exploring the Micro, Meso and Macro Proceedings of the European Distance and E-Learning Network 2018 Annual Conference, Genoa, Italy, 17–20 June 2018; pp. 17–20.
64. Buckingham Shum, S. *Learning Analytics: Policy Briefing*; 2012.
65. Aziz, A.A.; Jusoh, J.A.; Hassan, H.; Rizhan Wan Idris, W.M.; Md Zulkifli, A.P.; Mohamed Yusof, S.A. A framework for educational data warehouse (EDW) architecture using business intelligence (BI) technologies. *J. Theor. Appl. Inf. Technol.* **2014**, *69*, 50–58.
66. Flanagan, B.; Ogata, H. Integration of learning analytics research and production systems while protecting privacy. In Proceedings of the ICCE 2017—25th International Conference on Computers in Education: Technology and Innovation: Computer-Based Educational Systems for the 21st Century, Workshop Proceedings, Christchurch, New Zealand, 4–8 December, 2017; pp. 355–360.
67. Kahoot Kahoot! Learning Games & Make Learning Awesome! Available online: <http://kahoot.it> (access on 20 June 2018).
68. Alsina, M.; Canaleta, X.; Cubeles, A.; Torres, R. Self Directed Based Learning, una Metodología de Aprendizaje para Programas no Presenciales. In Actas del V Congreso Internacional sobre Aprendizaje, Innovación y Cooperación, CINAIC 2019, Madrid,

- España, 9–11 October 2019; pp. 325–330.
69. L’Heureux, A.; Grolinger, K.; Elyamany, H.F.; Capretz, M.A.M. Machine Learning with Big Data: Challenges and Approaches. *IEEE Access* **2017**, *5*, 7776–7797, doi:10.1109/ACCESS.2017.2696365.
 70. Marciano, A.; Nicita, A.; Ramello, G.B. Big data and big techs: Understanding the value of information in platform capitalism. *Eur. J. Law Econ.* **2020**, *50*, pp. 345–358. doi:10.1007/s10657-020-09675-1
 71. Luhn, H.P. The Automatic Creation of Literature Abstracts. *IBM J. Res. Dev.* **1958**, *2*, 159–165, doi:10.1147/rd.22.0159.
 72. Luhn, H.P. A Business Intelligence System. *IBM J. Res. Dev.* **1958**, *2*, 314–319, doi:10.1147/rd.24.0314.
 73. Gore, A. The Information Highway. *Exec. Educ.* **1994**, *16*, 30–31.
 74. Wiggins, R. Al Gore and the creation of the Internet. *First Monday* **2000**, *5*, doi:10.5210/fm.v5i10.799.
 75. Romero, C.; Ventura, S. Data mining in education. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2013**, *3*, 12–27, doi:10.1002/widm.1075.
 76. Siemens, G. Learning Analytics: The Emergence of a Discipline. *Am. Behav. Sci.* **2013**, *57*, 1380–1400, doi:10.1177/0002764213498851.
 77. Duval, E. Learning Analytics and Educational Data Mining. Available online: <https://erikduval.wordpress.com/2012/01/30/learning-analytics-and-educational-data-mining/> (accessed on 24 July 2018).
 78. Siemens, G.; d Baker, R.S.J. Learning analytics and educational data mining: Towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, BC, Canada, 29 April–2 May 2012; pp. 252–254.
 79. Merceron, A.; Blikstein, P.; Siemens, G. Learning Analytics: From Big Data to Meaningful Data. *J. Learn. Anal.* **2016**, *2*, 4–8, doi:10.18608/jla.2015.23.2.
 80. Berners-Lee, T.; Fielding, R.; Frystyk, H. Hypertext Transfer Protocol--HTTP/1.0 1996. Available online: <http://www.hjp.at/doc/rfc/rfc1945.html> (accessed on 20 July 2020).
 81. Slade, S.; Prinsloo, P. Learning Analytics: Ethical Issues and Dilemmas. *Am. Behav. Sci.* **2013**, *57*, 1510–1529, doi:10.1177/0002764213479366.
 82. Sonnad, N. *A Flawed Algorithm Led the UK to Deport Thousands of Students*.
 83. Bakharia, A.; Kitto, K.; Pardo, A.; Gašević, D.; Dawson, S. Recipe for Success—Lessons learnt from using xAPI within the connected learning analytics toolkit. In Proceedings of the ACM International Conference Proceeding Series, Edinburgh, Scotland, 25–29 April 2016; pp. 378–382.
 84. Herrero-Martín, J.; Canaleta, X.; Fonseca, D.; Merino, C.R.; Kinnear, L.; Amo, D. Designing a Multi-Scale and Multi-Dimensional Assessment for a New National Educational Context. In Proceedings of the Technological Ecosystems for Enhancing Multiculturality, TEEM’20, Online Conference, 21–23 October, 2020. Available online: <https://youtu.be/Qu7xAO3Om28> (accessed on 23 October 2020).