IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 6, 2019, accepted March 12, 2019, date of publication March 20, 2019, date of current version April 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906343

A Hybrid Infrastructure of Enterprise Architecture and Business Intelligence & Analytics for Knowledge Management in Education

OSWALDO MOSCOSO-ZEA[1], JORGE CASTRO[1], JOEL PAREDES-GUALTOR[1], AND SERGIO LUJÁN-MORA[2]

[1]Facultad de Ciencias de la Ingeniería e Industrias, Universidad Tecnológica Equinoccial, Quito 170509, Ecuador
[2]Departmento de Lenguas y Sistemas Informáticos, Universidad de Alicante, 03690 Alicante, Spain

Corresponding author: Oswaldo Moscoso-Zea (omoscoso@ute.edu.ec)

This work was supported by the Universidad Tecnológica Equinoccial.

ABSTRACT Advances in science and technology, the Internet of Things, and the proliferation of mobile apps are critical factors to the current increase in the amount, structure, and size of information that organizations have to store, process, and analyze. Traditional data storages present technical deficiencies when handling huge volumes of data and are not adequate for process modeling and business intelligence; to cope with these deficiencies, new methods and technologies have been developed under the umbrella of big data. However, there is still the need in higher education institutions (HEIs) of a technological tool that can be used for big data processing and knowledge management (KM). To overcome this issue, it is essential to develop an information infrastructure that allows the capturing of knowledge and facilitates experimentation by having cleaned and consistent data. Thus, this paper presents a hybrid information infrastructure for business intelligence and analytics (BI&A) and KM based on an educational data warehouse (EDW) and an enterprise architecture (EA) repository that allows the digitization of knowledge and empowers the visualization and the analysis of dissimilar organizational components as people, processes, and technology. The proposed infrastructure was created based on research and will serve to run different experiments to analyze educational data and academic processes and for the creation of explicit knowledge using different algorithms and methods of educational data mining, learning analytics, online analytical processing (OLAP), and EA analytics.

INDEX TERMS Big data, business intelligence, data warehouse, educational data mining, knowledge management.

I. INTRODUCTION

Boards of directors in higher education institutions (HEIs) are recognizing the potential and the leading role that Big Data analysis and knowledge management (KM) plays to improve the decision making processes [1]–[4]. The task of analyzing information in HEIs is currently challenging, mainly due to two factors. Firstly, due to a large amount of information generated every day by different applications and gadgets; and secondly, due to the problem of distributed and heterogeneous information systems. The latter problem is known as islands of information or information silos, which is produced by standalone applications dispersed in different departments and units of most HEIs. This problem has its roots in the disorganized and non-centralized growth of most HEIs due to the autonomy that their units have. These factors increase the complexity to store, integrate, process and analyze data in HEIs.

Moreover, this complexity upsurges due to the emergence and implementation of new technological devices and software as wearables, apps and web applications which increases the volume of data. One of the solutions to tackle this complexity and to perform a sound analysis of this huge amount of information known as Big Data is the development of an information infrastructure, which integrates processed and cleaned data loaded from different internal and external data sources [5]. This infrastructure for HEIs allows carrying

The associate editor coordinating the review of this manuscript and approving it for publication was Saqib Saeed.

out experiments with high quality data to obtain reliable, believable and accessible information.

One of the reasons for the implementation of this novel education infrastructure is due to the unique business model in universities, which greatly differs from traditional industries as commercial enterprises. The access to information in HEIs is not always easy to obtain mainly due to the hierarchical organizational structure and the levels of power in the academic departments. The proposed infrastructure enables decision support and is primarily based on an educational data warehouse (EDW) to handle KM. KM is a discipline that promotes the creation, use, distribution and transfer of knowledge in organizations [6]. An important strategy for continuous improvement in HEIs is the implementation of a KM system as a differentiating factor that allows providing efficient educational services to students in their academic and administrative processes.

Business Intelligence and Analytics (BI&A) is in the core of the proposed infrastructure. BI&A refers to the technological process of multidimensional analysis of data and Big Data realized with a plethora of tools and techniques to improve the decision making process [7]. The main reasons to use and implement the proposed infrastructure is that this analysis cannot be done with operational data sources since they are not adequate for BI&A, Big Data analysis nor for knowledge creation. The main objective of operational data sources is to support the daily operations of an organization and are built to provision only the storage and handling of structured data. These data sources are regularly busy serving specific software applications and thus, specific business processes. In addition, if they were used for analysis, this would affect the performance of those applications because intensive calculations of billions of records with different data formats and structure are required in some cases. Consequently, there is a growing need for a data repository in HEIs that can be used exclusively for BI&A and KM to improve decision making. This data repository needs to be populated with high quality data and must serve specifically for this purpose [8].

We sense that one of the shortages of the data analytical process in HEIs is that the implementation of an information infrastructure which stores clean and high-quality data is not always a priority. We have seen that much attention has been paid lately in research to the study of learning analytics and educational data mining (EDM) [9], [10]. However, very little attention is given to studying the infrastructure and processes that ensure and support that the results of analytics are of high quality and can contribute to making reliable decisions [11]. To cope with this problem, we propose a hybrid solution to support data and organizational analytics. One of the main components of the solution is an EDW. An EDW is one of the core technologies of BI&A for the analysis of Big Data and knowledge creation in HEIs. Data analysis in an EDW can be done using online analytical processing (OLAP) and EDM.

Nevertheless, an EDW cannot store organizational knowledge which resides in models (people and processes). For example the existing relations between businesses processes, employees, applications, and technology. This issue can be solved by the introduction of another tool that allows the modeling, storing and analytics of the organizational components mentioned before. Thus, in order to have a clear picture of these components, the novelty of our proposed approach is the incorporation of an enterprise architecture (EA) tool [12]. According to Gartner [13] ''Enterprise Architecture is the process of translating business vision and strategy into effective enterprise change by creating, communicating and improving the key requirements, principles, and models that describe the enterprise's future state and enable its evolution''. In this paper, we propose a hybrid information infrastructure grounded on BI&A and EA which was implemented in a case study performed on a private university. As far as we know, this is the first approach that tries to integrate these two technologies to create new types of analytics and therefore, first-hand explicit knowledge. Therefore the research question to answer with this work is:

How to implement an infrastructure for BI&A and KM in HEIs oriented to excellence and quality for the satisfaction of administrative staff and the academic community.

After introducing the topic of this paper in this section, the rest of the paper is structured as follows: Section 2 describes the research made to decide on the main components of the proposed infrastructure and explains them in detail; Section 3 presents the hybrid infrastructure and lists the steps for the implementation, moreover, it presents the new types of analytics that this solution can provide; finally, in Section 4 conclusions are drawn and the outlook of the work is described.

II. THEORETICAL BACKGROUND

Previous empirical research indicates the lack of a comprehensive knowledge infrastructure implementation [14]. Hence, this investigation is focused on the development of technological infrastructure for BI&A and KM as a means of value creation for HEIs. After the research done, we propose to hybridize components from different fields to guarantee the analysis of different organizational dimensions and information. Moreover, this infrastructure supports the transformation of information into useful knowledge using BI, machine learning and data mining. Thus, the fields of study which are part of this work are presented in this section. These fields are BI&A, EA, KM, and EDM. Also, this section describes a KM framework, which guides the implementation of the proposed infrastructure [15].

A. BUSINESS INTELLIGENCE & ANALYTICS

BI&A is ''a system comprised of both technical and organizational elements that present historical information to its users for analysis and enables effective decision making and management support, for the overall purpose of increasing organizational performance'' [16, p. 161]. BI&A in HEIs builds upon a set of tools and applications that enable the analysis of data and Big Data to improve data governance and performance [17]. To achieve this objective, stakeholders require having access to all the required information in the organization. A data warehouse is a preferred repository to analyze the business, its requirements, and its trends. The BI&A process that we suggest goes beyond the analysis of data from applications. The infrastructure recommended in this investigation implements technology for capturing the business processes and the services that are offered in the organization across the technological infrastructure and through the applications and information systems.

Markets and Markets project the worldwide BI market will reach $26.88 billion in 2021 [18]. The BI market includes all BI&A platforms, management suites, and advanced analytics solutions [19]. One of the key applications of BI&A is the creation of knowledge to enable informed decision making [20]. The creation of knowledge involves establishing procedures for capturing implicit and explicit knowledge from people, processes, and technology and set up a mechanism for analyzing this knowledge in the quest to find new related information that once again can infer new knowledge. One of the key disciplines that support the initiative to identify the relationships between knowledge dimensions is EA.

B. ENTERPRISE ARCHITECTURE

Advances and innovation in science and technology, as well as pressure for increasing student's admissions and internationalization, have led to a very complex environment in educational organizations. Besides, regulation and deregulation imposed by governments and accreditation institutions cause the implementation of technological changes to be extremely challenging due to a lack of understanding of the consequences and impact that changes can cause [21]. EA is a blueprint that incorporates methods and techniques to have a complete view of the organization using models, frameworks, constraints, principles, and guidelines [21]. This blueprint supports control, governance and diminishes the organizational and the technological complexity to improve the impact of changes in HEIs. Moreover, this blueprint can be used to support different KM activities as measurement, identification, acquisition, development, use, preservation and distribution of knowledge [22].

Additionally, EA supports directors and authorities of HEIs in making decisions while allowing the organization to obtain benefits such as cost savings, process standardization, identification of duplicated functionality of applications, the development of a common understanding of technology and governance of key data resources [23]. EA is essential to overcome some of the main challenges of a broad number of organizations [24]. In the case of HEIs, the main challenges are data integration, the opening of new campuses, implementation of novel technologies, pressure for innovation, attracting prospective students, individual student solutions and outsourcing.

C. KNOWLEDGE MANAGEMENT

Knowledge is considered a capability which can strengthen the positioning of an HEI [25]. Knowledge gives identity and differentiation to an organization. The differentiation comes from the strengths and weaknesses that the different knowledge assets provide to organizations. Knowledge has the same characteristics of a resource as it generates value and is difficult to imitate. The differentiation that knowledge assets give organizations provides a competitive advantage that is hard to replicate [26]. One of HEIs' main objective is to produce knowledge and become knowledge providers to citizens [27]. Therefore, knowledge must be captured and managed appropriately. Knowledge is defined as experience, facts, know-how, processes and beliefs that increase an organizational or individual's capability [28].

KM is the process of identifying, capturing and transferring organizational knowledge to increase the organizational competitiveness [29]. Also, KM supports the exploitation and development of the knowledge assets of an organization with a view to strengthening the organization's capabilities [30]. The reasons for conducting KM in HEIs includes exchanges of academic staff, information overload, the increasing need for expert administrative and academic staff, improvement in decision making, and digitalization of knowledge. To succeed in the implementation of KM, it is necessary to first develop a process for knowledge retention, and sharing and secondly to establish a technological infrastructure for this purpose [31].

The knowledge infrastructure must be capable to manage technological assets, people and processes. The first element to be managed in the infrastructure is technology. Technology is defined as: "The purposeful application of information in the design, production, and utilization of goods and services, and in the organization of human activities" [32]. Moreover, according to [33], technology consists of two core components: 1) a physical component composed of products, tools, equipment, blueprints, techniques, and processes; and 2) the informational component which is composed of the know-how in the following areas: management, marketing, production, quality control, reliability, skilled workforce, and functional areas. In this paper, technology is referred to as objects used by humans (tools, software, hardware, machines). Technology is in the core of data capturing and knowledge creation.

The second element to be managed in this infrastructure is people. According to [34, p. 10] "knowledge resides in the user and not in the collection of information…it is how the user reacts to a collection of information that matters". Thus, it is important to design a mechanism that serves to capture and to transfer implicit knowledge immerse in the academic and administrative staff to a data repository. The high rates of staff turnover in HEIs [35], [36] call for the implementation of strategies to retain the knowledge workers (academic staff). Organizations that failed to retain these workers are spending more regarding finding new staff and providing training [35]. It is important to note that a high performer worker deliver up to 400% more than their counterparts, therefore, the cost of wasted talent represents a serious problem to organizations [37]. Furthermore, turnover is considered a major concern in HEIs around the world [38]. Staff turnover cost US

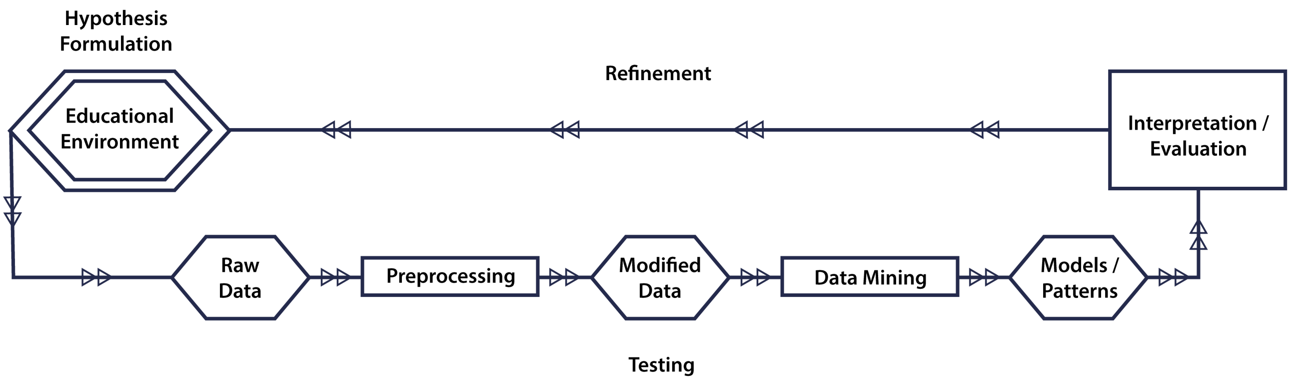


FIGURE 1. Steps in knowledge discovery with EDM.

companies 600 billion in 2018 and is estimated that this number will grow to 680 billion in 2020 [39]. The high rates are indicators that strategies need to be implemented to overcome the loss of knowledge when academic or administrative employees leave the organization.

The third element that must be captured in the knowledge repository are the processes. Edwards describe processes as ‘‘the way people, organizations and even technology actually do things’’ [40, p. 297]. The importance of processes in a KM initiative is described in different papers [41], [42]. Modeling of the current and future processes of an organization is an important step in the implementation of knowledge infrastructure. Processes modeled in an EA tool can be easily explained to new staff and they support stakeholders in the process of maintaining implicit knowledge of personal leaving the organization.

D. KNOWLEDGE MANAGEMENT FRAMEWORK

A KM framework is an abstract structure that guides the conception of a KM project. A KM framework supports the creation, capturing, use, distribution, and transfer of explicit and implicit knowledge. One of the frameworks that provide an important explanation of KM is presented by Nonaka and Lewin [43]. To understand this framework it is important to differentiate explicit knowledge from implicit knowledge (also known as tacit knowledge). Explicit knowledge can be formulated, documented and reproduced. On the contrary, implicit knowledge is difficult to document since it is associated with human knowledge. Implicit knowledge is composed of skills that people acquire such as speaking a new language or performing a work activity, therefore, is difficult to transfer to other persons. Nonaka states that the process to generate knowledge is based on the conversion of tacit knowledge to tacit knowledge, explicit knowledge to explicit knowledge and tacit knowledge to explicit knowledge and vice-versa [15]. The ways to transfer knowledge within this approach are for example by sharing experiences between old and new workers, by observing a skilled worker, knowledge obtained in meetings, capturing staff's knowledge in books or information systems and by analyzing reports and generating new knowledge.

The KM frameworks proposed in the literature are classified as prescriptive, descriptive, and hybrids [44]. Prescriptive frameworks are based on a methodology or procedures to give directions on how to engage on KM endeavors (centered on KM tasks). In contrast, descriptive frameworks are formulated to describe KM. This is to say, these frameworks enable the identification of factors that influence the success or failure of KM endeavors. There is also a hybrid approach for KM which includes characteristics from both prescriptive and descriptive KM frameworks (includes tasks to be done and how to implement an initiative in a practical manner). In the study from [45], he compares 160 KM frameworks developed around the globe and identifies that half of the studied frameworks present and hybrid approach, therefore, hybrid KM frameworks are the most used.

E. EDUCATIONAL DATA MINING

EDM uses methods, tools, and algorithms of data mining to investigate educational data from students, teachers and administrative staff of HEIs. EDM is defined as an ‘‘emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to better understand students, and the settings which they learn in’’ [46, p. 1]. Romero and Ventura [47] propose an EDM method which is shown in Figure 1. This figure depicts the knowledge discovery process with EDM. In the first step of the process, the preprocessing of raw data obtained from an educational environment is performed. This raw data is then modified (a new dataset is created) and used with EDM methods or algorithms. After that, a model is defined and the experiments are carried out. The results of experimentation are then evaluated and refined to improve the process [47]. EDM has been used in the last years to improve academic indicators [48].

III. HYBRID KM INFRASTRUCTURE

This section presents the hybrid data infrastructure for KM in HEIs. The objective of the infrastructure is improving data analytics and therefore, decision making. This infrastructure was designed after performing research intended to

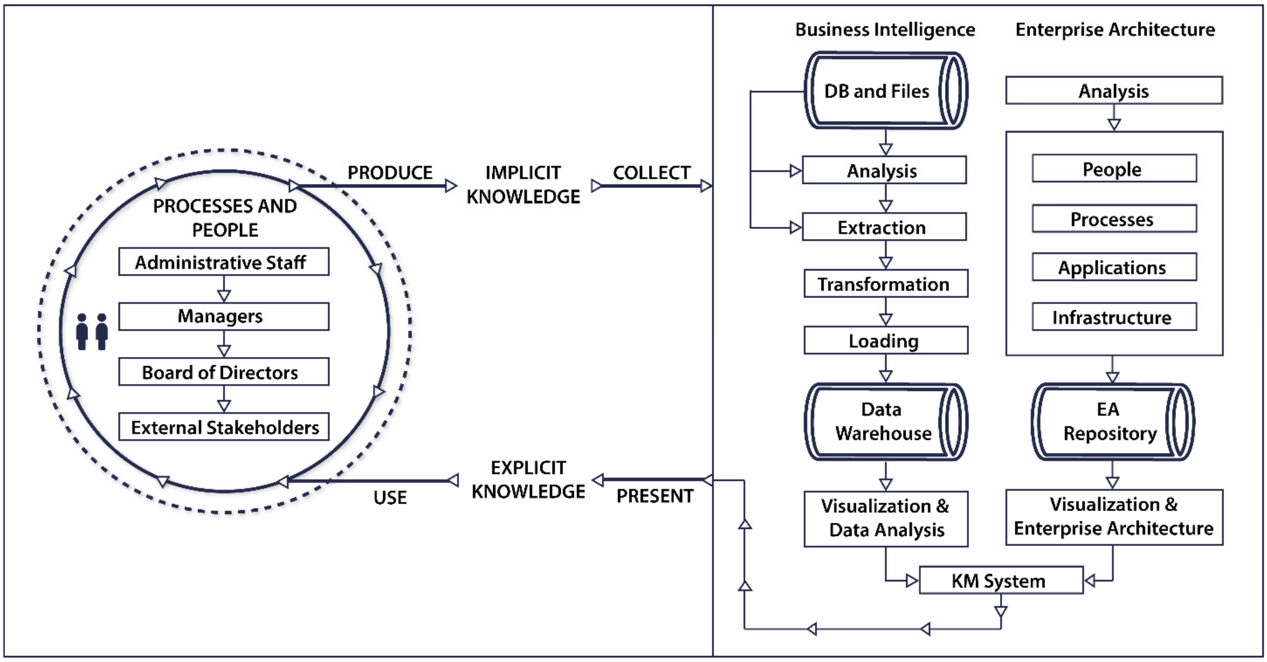


FIGURE 2. **Flow diagram of the knowledge management framework.**

understand the main sources of knowledge acquisition in an organization and the attempts to implement KM systems in the past [29], [49]–[52]. The hybrid approach proposed is materialized from the integration of EA and BI&A. Therefore, this infrastructure suggests a method for collecting, presenting, using and producing knowledge.

A. KNOWLEDGE INFRASTRUCTURE

The novel knowledge infrastructure proposed in this paper is based on a KM framework that we designed to improve decision making in organizations [15]. This KM framework, used as the model for the implementation of the proposed infrastructure, includes EA and BI&A to guide the capturing of all the knowledge dimensions. The KM framework shown in Figure 2 illustrates the knowledge creation process. The component at the right depicts how explicit knowledge is produced by using EA and BI&A. This component takes implicit knowledge as an input. The implicit knowledge is produced by people and the understanding of the collection of processes in different areas of the organization. This understanding begins with the analysis of the architectural models digitized in an EA tool.

The knowledge discovery process inside the right box of the framework has the following steps: analysis of existing databases and files, extraction of useful information, transformation, formatting, and loading of data. This process known as Extraction, Transformation, and Loading (ETL) prepares data into a customizable format, cleans data with errors and eliminates duplicates, among other tasks. The purpose of this step is to load valuable data into the target database to improve the data analytics process. An EDW is a database composed of dimensions and fact tables which is independent of the transactional and operational environment of the organization. This multidimensional repository is used because it allows intensive calculations, processing, and analysis in millions of rows of data. Therefore, due to its features, it is considered the best target database [53]. The populated EDW is explored by data analysts using BI&A and data mining tools to visualize organizational knowledge.

Alternatively, the knowledge digitalization can be done in an EA tool. An EA tool supports the creation of architectures to translate implicit knowledge into models which describe organizational structures (people), business processes, applications, and technological infrastructure. The digitalization process includes an analysis of the different units and departments of the organization. Interviews with the staff must be realized by a team of business architects to document and model the different activities they perform.

EA and BI&A are the main methodologies for the creation of explicit knowledge. This type of knowledge must be presented easily and understandably. Consequently, the framework suggests the presentation of knowledge by using a KM system. The results of the BI&A and the EA process can be visualized and analyzed in the KM system. The output of the component at the right is explicit knowledge in the form of reports and dashboards that are presented to be used by people in all levels of the organization and can support in the design or redesign of new and existing processes. The explicit knowledge is the main input of the left box of the framework. Explicit knowledge can support and enhance decision making activities and can increase knowledge levels of the organization members. It supports the transfer of knowledge

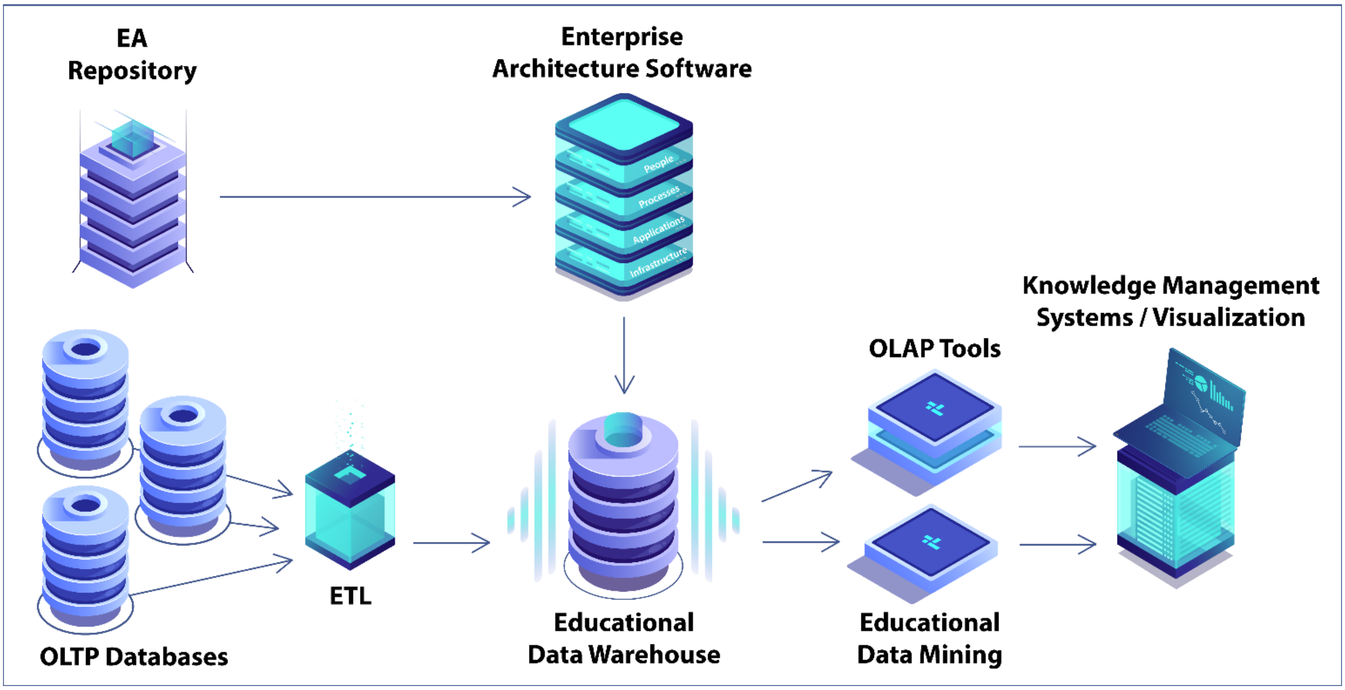


FIGURE 3. Proposed BI&A infrastructure.

to new employees as well. As seen in the framework depicted in Figure 2, the KM process is a cycle in which knowledge is produced on a daily basis.

A literature review that we performed in the past, warns researchers the little importance it has been given to the EDW topic [11]. Therefore, we believe it is necessary to give guidelines to researchers and practitioners on how to implement an EDW plus how to integrate it with an EA tool to implement a KM initiative in an HEI. In this section, an information infrastructure based upon the proposed KM framework is presented. This infrastructure highlights the main technological components that must be implemented in an HEI to analyze organizational information, produce and manage knowledge. The main components of our proposal — EA Software, EDW, OLAP tools, EDM software, KM system and visualization tools— are shown in Figure 3 and are described in the following sections.

B. IMPLEMENTATION OF THE INFRASTRUCTURE

The implementation of the proposed knowledge infrastructure in an HEI requires five steps that are shown in Figure 3: 1) Implementation of EA software, 2) Implementation of an EDW, 3) Implementation of OLAP tools, 4) Implementation of EDM tools and 5) Implementation of a KM system and visualization tools. The details of these steps are described in this section.

C. IMPLEMENTATION OF ENTERPRISE ARCHITECTURE

The first step to implement the proposed knowledge infrastructure requires the selection of an EA software tool. An EA tool supports the modeling of the organizational components of an HEI as people, academic departments, processes, applications, and technical infrastructure. With the processes of all the different dimensions of the organization modeled, a data analysist can have a clear picture of the organization. This visibility makes it easier to identify the pertinent applications and the data that could be loaded into the EDW. In order to choose the right EA tool, it is recommended to perform an evaluation and selection of software tools. In a previous work, we presented the initiatives for performing an evaluation of EA tools and after investigating the literature [54], we chose the three most adequate methodologies for this purpose. Moreover, the steps to initiate an EA project were described [54], [55]. Once the EA tool has been carefully chosen by performing different analysis (obtaining a long-list, hands-on evaluation, obtaining a short-list, final evaluation), digitization of knowledge should be carried out. The digitization of knowledge requires a stakeholder analysis and a team of specialized people to perform interviews before the start of the project. These interviews are essential for capturing the know-how of the academic and administrative staff and the activities they perform. It is necessary during or after the interviews to model the applications, the business processes and the different procedures and daily activities that the staff performs. The main output of this phase is the identification of the relevant and requested data repositories and information that will be used in the ETL to the EDW.

In this investigation, we used an EA software tool to model the different organizational dimensions which exist in an HEI. Mainly these dimensions cover the people (students, researchers, academic and administrative staff), the business processes (admission, registration, evaluation, graduation,

TABLE 1. Overview of the procedures for analyzing the EA.
Source: [57, pp. 126–128]

Object under Investigation	Description of Procedure	Typical Questions
Dependency	Directly or indirectly (i.e. cross-level) linked elements in the EA are selected. Relationships and their impact are shown.	What other elements are affected when we replace infrastructure component X?
Coverage	The coverage of departments (e.g. units in a process-product matrix) by application systems is analyzed.	What redundancies or gaps exist in the IT support for process X and/or product Y and/or organizational unit Z?
Interfaces	The interfaces between the application systems are analyzed in terms of their type, number, complexity, frequency/currency, performance, stability, and availability.	Does the support for process X contain gaps and cases of heterogeneity? Are common steps in product processing also handled in a cross-product manner?
Heterogeneity	The heterogeneity of ones IT assets in defined areas of deployment is analyzed.	How many development lines (technologies) are there per deployment area (e.g. a unit in the process-product matrix)? How many infrastructure components are there per cell in the set of infrastructure standards?
Complexity	An analysis is run to determine how many components there are in the EA and how many relationships they have.	How many application systems exist? How many interfaces do they have? How many infrastructure components and platforms exist? How many interfaces exist among them or to the applications environment?
Conformity	Adherence to standards and ascertainment of the degree of variance (e.g. as a percentage of the application systems or infrastructure components)	Has adherence to existing standards (e.g. the set of infrastructure standards) been secured? Have the defined reference architecture models been implemented? What percentage of all components is out of compliance with the standards?
	Compliance rules	Has compliance with legal provisions, market standards and norms (e.g. Sarbanes-Oxley and Solvency II) been secured?
Costs	Reporting on accumulated production, operation, and maintenance costs	What costs are associated at all levels of the EA with the IT support of product X?
Benefits	Benefits calculation, e.g. as a percentage of contributions to the achievement of enterprise goals or via defined KPIs	What contribution to the support of enterprise goals is made by application system X?

research, student welfare, among others) and the information technology (IT) infrastructure (servers, personal computers, laptops, mobile devices, software applications, and the computer network). The EA tool was integrated into the EDW proposed in this work to encounter the relations between the organizational dimensions and to discover insights that could not be seen by using only an EDW. The data from the EA repository are loaded and integrated with the data from the data warehouse into a single repository for a different and new type of analytics. EA analysis is concerned with the application of property assessment criteria on EA models in order to infer new knowledge [56]. EA tools can perform diverse types of analysis. In Table 1 an overview of the data analysis procedures that can be performed using EA is shown. These procedures are the following: dependency, coverage, interfaces, heterogeneity, complexity, conformity, costs, and benefits.

There are different challenges when implementing EA projects [57]. These challenges are mainly because most HEIs are not aware of EA and therefore, an architecture of their business process, applications and infrastructure does not exist. Consequently, the alignment of business and IT becomes challenging. According to the Harvey Nash/KPMG 2017 CIO Survey, EA has become the fastest-growing, in-demand skillset in technology [58]. HEIs using EA could acquire a competitive advantage in a short period by implementing agile EA [59], which is the gradual implementation of EA in the areas or departments of more interest for the organization to obtain value at an early stage.

D. IMPLEMENTATION OF EDUCATIONAL DATA WAREHOUSE

Different EDW design methodologies [53], [60]–[62] can be found in the literature. After performing several analysis in a systematic mapping, the use of Kimball's methodology is suggested in an educational scenario [11], [63]. This suggestion is valid for HEIs where its departments and units are not integrated and operate as information silos. Based on the results of the systematic mapping the following steps are proposed for the EDW design: 1) Diagnosis and requirements analysis, 2) Data source and data supply analysis, 3) EDWdesign and multidimensional modeling, 4) ETL process and 5) Application system, reporting, dashboard, and OLAP development.

1) The diagnosis defines the current situation of the organization, the scope of the EDW project and identification of stakeholders. Different management frameworks can be used for this purpose, for example, The Open Group Architecture Framework (TOGAF), European Foundation for Quality Management (EFQM) excellence model, business process models (BPM) [64], among others.
2) The data source and data supply analysis (Information need analysis) is recommended to be done performing interviews with stakeholders. It comprises three studies. Firstly, an analysis of the information supply of the organization. The information supply analysis includes a study of all relevant data sources that will be significant to the scope of the project (an EA tool might be used). Secondly, an analysis of the objective information needs. The objective information need analysis requires the study of the business models, processes and strategies of the organization. Finally, subjective information needs analysis must be performed which includes eliciting explicit requirements from the stakeholders.
3) The EDW design and multidimensional modeling include the modeling of the conceptual, logical and physical model. The EDW design needs to consider all the requirements gathered in the first phase. There are different methodologies that can be used to design a data warehouse [53], [60]–[62]. In a systematic mapping that we performed on the past on data warehouses in education [11], it was found out that Kimball is the most used design methodology and the star schema is the most implemented schema in educational institutions. But, based on the requirements of the project, practitioners and researchers can choose which methodology or schema suits a particular scenario. It is strongly recommended to use CASE software to convert the models into the EDW schema.
4) The ETL process should be executed in this phase. ETL supports the loading of data from operational data sources into the EDW. This step includes the extraction process to intermediate files or directly to the target database. In both cases, a data cleansing process needs to be done. It is recommended that the loading process is automated so that the repository stays updated.
5) The implementation of the application system to obtain reports and dashboards is the final phase of development of the EDW. In this study, the analysis of data from the EDW can be performed using a data base management system and the KM system that is suggested to be implemented as a future step.

E. IMPLEMENTATION OF ON-LINE ANALYTICAL PROCESSING TOOLS

The OLAP tools need to be selected and installed. These tools support stakeholders in the data analysis and the decision-making process. Moreover, these tools help to perform roll up, drill down, slicing and dicing operations with the data. OLAP has been considered for many years as the cornerstone of BI. There is a new vision for OLAP operations in the future. The idea is to develop a new type of OLAP analysis by defining intentional operators which try to understand the intentions of the user with respect to the information needs. Intentional operators allow users to define their analytical goals over the cubes as intentions and map these intentions with knowledge discovery algorithms to infer new knowledge and to obtain new types of analysis [65].

F. IMPLEMENTATION OF EDUCATIONAL DATA MINING TOOLS

It is suggested the implementation of EDM software which can be used for predicting trends and defining patterns with the data. EDM explores the organizational context and is key in the study and improvement of academic indicators in HEIs as dropout rate, graduation rate, restructuration processes, and organizational management [66].

Diverse tools can be used for this purpose which are equipped with plenty of machine learning algorithms. A list of tools is shown in Table 2.

G. IMPLEMENTATION OF KNOWLEDGE MANAGEMENT SYSTEMS AND VISUALIZATION SOFTWARE

The implementation of KM systems supports the analysis of information from all the organizational dimensions in an organization. HEIs can obtain benefits of these systems. The infrastructure proposed in this paper is a hybrid that integrates an EDW with EA tools. The KM system must be designed with strong visualization capabilities. It is recommended as well to include additional tools and plug-ins which allow the presentation of reports and dashboards. Visualizations tools provide a graphical vision of the analyzed information. These tools are used to create graphical displays and interfaces for software applications. They are used in different industries, including transportation, telecommunications, manufacturing, and education, to display information in formats designed to be easily understood.

There are some methodologies for the implementation of KM systems [67], [68]. These methodologies suggest how to develop a project to cover the capturing of the knowledge

TABLE 2. Data mining tools in education.

Tool	Goal
EP Rules	To discover prediction rules to provide feedback for courseware actors
Microsoft SQL Server Analysis Services	Provides online analytical processing (OLAP) and data mining functions for Business Intelligence applications
EDM workbench	Tool to discover registration data information
Inq-ITS System	Track student progress using data mining algorithms and reports
Moodle Datamining Tool	Data mining tool for Moodle developed in Java, based on Keel
RapidMiner	A free platform for predictive analysis using machine learning and analytical techniques
R	Language and programming environment for statistical and graphic analysis
Weka	Collection of automatic learning algorithms for data mining tasks
KEEL	A software tool used in different knowledge discovery tasks
SNAPP	It allows evaluating patterns of student behavior with the design of learning activities
AHA System	Recommend the best links that the student can visit in the next step
Knime	Analytics platform to try to discover patterns in the data and predict trends
GISMO	Interactive graphic monitoring that provides visualization of student activities
SPSS	Predicting the future reliably for better decision making
Sequential Mining Tool	Mining that helps teachers discover important student information
KEA	Software that uses an algorithm to extract keywords from text documents
DB Miner	Mining to discover different levels of knowledge in large relational databases

dimensions. Our approach presents a more technical view on the development of a knowledge infrastructure. In this phase, we suggest creating the knowledge system as a web platform integrating the EDW repository and the EA repository to obtain the organizational collected data to infer new knowledge from both repositories and visualize reports and dashboards.

IV. CONCLUSIONS

This paper presents a hybrid information and knowledge infrastructure for improving decision making in HEIs. This infrastructure was designed based on empirical research intended to improve HEI's management. After analyzing the different possibilities we have proposed a method for the implementation of the infrastructure which answers the research question of this work. The infrastructure.is based on the integration of two data repositories: an EDW and an EA repository. By using BI&A and EDM, different experiments can be carried out to improve academic indicators. The results of these experiments give visibility to academic directors for timely decision making.

The systematic mapping that we performed on the topic of EDW [11] shows that researchers and HEIs do not give adequate importance to the implementation of an EDW. For that reason, this novel approach can guide in the implementation of an infrastructure that can manage all the knowledge generated in the organization and that can be a source of opportunities to improve organizational processes and a valid data source for applying different type of analytics with OLAP tools, dashboard and reporting tools and EA tools to assure quality outcomes in the BI&A process.

The implementation of this infrastructure in HEIs is challenging; mainly due to the fact that in companies it is unexpected to have a 180 degree turn in strategic direction, which is very usual in HEIs when there is a change of authorities. The proposed infrastructure is being implemented in a HEI with promising results. Nevertheless, it is important to establish mechanisms for the evaluation of the improvements made in knowledge processing.

Another fact is that in an HEI the staff has different responsibilities and the board of directors tends to change every four years which make the data analytics process very different to a business organization. For example, in Spain, the steering committee of public universities are elected every four years. The elected Rector and Vice-Rectors normally designate their governing team or academic directors, which in turn designate their trusted assistants. Likewise, this is repeated at the faculty level. These changes in the direction of an HEI are radically different from how a company works. Knowledge could be the cornerstone to overcome all the challenges to maintain the strategic direction over time. Therefore, the hybrid infrastructure design must consider these facts and be very flexible to ease disruptive changes in the strategic direction and management.

The proposed knowledge information infrastructure facilitates the traceability of the students, administrative and academic staff to improve decision making and to improve organizational processes. Besides, it is a source to have a knowledge base in the organization and a tool to produce explicit knowledge. Moreover, the use of EA, as a part of the solution, can support and ease the problems derived from staff turnover which is the main cause of the loss of implicit knowledge and the increase in operating costs.

As future work, we propose to research more mechanisms to obtain the implicit knowledge from the staff of the organization and to automate the ETL process to have the repository updated all the time. Moreover, the hybrid infrastructure presented in this work can help improve the efficiency of a HEI. The models captured in the EA tools can be a source for new software development when a labor-intensive process is identified and can be automated.

Finally, we propose to combine the resulting EA models with model-driven software development. The models can enable the development of applications based on all the required business processes, actors, roles and objects for a given department in a HEI. Using all the capabilities provided by the EA modeling languages, we can design applications as industry-specific as needed with techniques such as profiling or stereotyping. This proposed initiative can serve as a solution for the creation of new knowledge capabilities and will contribute in the desired technological business alignment of any HEI.

REFERENCES

[1] J. Bichsel. *Analytics in Higher Education Benefits, Barriers, Progress and Recommendations*. EDUCAUSE Center for Applied Research. Accessed: Aug. 2012. [Online]. Available: https://library.educause.edu/~/media/files/library/2012/6/ers1207.pdf?la=en

[2] L. V. Morris, S.-S. Wu, and C. L. Finnegan, "Predicting retention in online general education courses," *Amer. J. Distance Edu.*, vol. 19, no. 1, pp. 23–36, 2005.

[3] P. Baepler and C. J. Murdoch, "Academic analytics and data mining in higher education," *Int. J. Scholarsh. Teach. Learn.*, vol. 4, no. 2, pp. 1–9, 2010.

[4] O. Moscoso-Zea and S. Luján-Mora, "Knowledge management in higher education institutions for the generation of organizational knowledge," in *Proc. 12th Iberian Conf. Inf. Syst. Technol. (CISTI)*, 2017, pp. 1–7.

[5] C. L. Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA, USA: MIT Press, 2007.

[6] H. M. Campbell, "The role of organizational knowledge management strategies in the quest for business intelligence," in *Proc. IEEE Int. Eng. Manage. Conf.*, Sep. 2006, pp. 231–236.

[7] H.-C. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quart.*, vol. 36, no. 4, pp. 1165–1188, 2012.

[8] S. Luján-Mora and J. Trujillo, "A comprehensive method for data warehouse design," in *Proc. 5th Int. Workshop Design Manage. Data Warehouses (DMDW)*, 2003, pp. 1–14.

[9] O. Moscoso-Zea and S. Luján-Mora, "Educational data mining: An holistic view," in *Proc. 11th Iberian Conf. Inf. Syst. Technol. (CISTI)*, Jun. 2016, pp. 1–6.

[10] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 135–146, 2007.

[11] O. Moscoso-Zea, J. Paredes-Gualtor, and S. Luján-Mora, "A holistic view of data warehousing in education," *IEEE Access*, vol. 6, pp. 64659–64673, Oct. 2018.

[12] H. Franken, S. van Dijk, and B. van Gils. *Three Best Practices for Successful Implementation of Enterprise Architecture Using the TOGAF Framework and the ArchiMate Modeling Language*. The Open Group Blog. Accessed: Feb. 8, 2013. [Online]. Available: https://tinyurl.com/ce7yjbr

[13] Gartner Inc. *Gartner IT Glossary*. Mar. 18, 2013. [Online]. Available: http://www.gartner.com/it-glossary/enterprise-architecture-ea/

[14] S. Gudas, "Knowledge-based enterprise framework: A management control view," in *New Research on Knowledge Management Models and Methods*, H. T. Hou, Ed. Rijeka, Croatia: InTech, 2012.

[15] O. Moscoso-Zea, S. Luján-Mora, C. E. Cáceres, and N. Schweimanns, "Knowledge management framework using enterprise architecture and business intelligence," in *Proc. 18th Int. Conf. Enterprise Inf. Syst. (ICEIS)*, 2016, pp. 244–249.

[16] O. Isik, M. Jones, and A. Sidorova, "Business intelligence (BI) success and the role of BI capabilities," *Intell. Syst. Accounting, Finance Manage.*, vol. 84, no. 4, pp. 161–176, 2011.

[17] Gartner Inc. *Gartner Business Intelligence*. Accessed: Jan. 20, 2019. [Online]. Available: http://goo.gl/LmJRG3

[18] Markets and Markets. *Business Intelligence Market Worth 26.88 Billion USD by 2021*. Accessed: Jan. 22, 2017. [Online]. Available: http://www.marketsandmarkets.com/PressReleases/social-business-intelligence-bi.asp

[19] Forbes. *The State Of Business Intelligence*. Accessed: Jun. 8, 2018. [Online]. Available: http://intelligence.market/business7

[20] J. Lloyd, "Identifying key components of business intelligence systems and their role in managerial decision making," *Appl. Inf. Manag. Progr.*, vol. 1277, no. 2, pp. 1–76, 2011.

[21] M.-E. Iacob, H. Jonkers, D. Quartel, H. Franken, and H. van den Berg, *Delivering Enterprise Architecture with TOGAF and Archimate*. Enschede, The Netherlands: BIZZdesign, 2012.

[22] S. Buckl, F. Matthes, and C. M. Schweda, "Future research topics in enterprise architecture management—A knowledge management perspective," in *Proc. Int. Conf. Service-Oriented Comput.*, 2009, pp. 1–11.

[23] E. Niemi, "Enterprise architecture benefits: Perceptions from literature and practice," in *Proc. 7th IBIMA Conf. Int. Inf. Syst. Digit. Age*, 2006, pp. 1–8.

[24] M. Op't Land, E. Proper, M. Waage, J. Cloo, and C. Steghuis, *Enterprise Architecture Creating Value by Informed Governance*. Berlin, Germany: Springer-Verlag, 2009.

[25] C. Curado and N. Bontis, "The knowledge-based view of the firm and its theoretical precursor," *Int. J. Learn. Intellect. Cap.*, vol. 3, no. 4, pp. 1–18, 2006.

[26] B. Wernerfelt, "A resource-based view of the firm," *Strategic Manage. J.*, vol. 5, pp. 171–180, 1984.

[27] P. G. Altbach, "Higher education and the WTO: Globalization run Amok," *Int. High. Edu.*, vol. 23, pp. 2–4, Mar. 2001.

[28] K. Karemente, J. R. Aduwo, E. Mugejjera, and J. Lubega, "Knowledge management frameworks a review of conceptual foundations and a KMF for IT-based organizations," in *Strengthening the Role of ICT in Development*, vol. 5, 2009, pp. 35–57.

[29] M. Alavi and D. E. Leidner, "Knowledge management and knowledge management systems: Conceptual foundations and research issues," *MIS Quart.*, vol. 25, no. 1, pp. 107–136, 2001.

[30] J. Rowley, "From learning organisation to knowledge entrepreneur," *J. Knowl. Manage.*, vol. 4, no. 1, pp. 7–15, 2000.

[31] P. Heisig, "Harmonisation of knowledge management—Comparing 160 KM frameworks around the globe," *J. Knowl. Manage.*, vol. 13, no. 4, pp. 4–31, 2009.

[32] BusinessDictionary. *Technology Definition*. Accessed: Jan. 20, 2019. [Online]. Available: http://goo.gl/a266MR

[33] V. Kumar, U. Kumar, and A. Persaud, "Building technological capability through importing technology: The case of indonesian manufacturing industry," *J. Technol. Transf.*, vol. 24, no. 1, pp. 81–96, 1999.

[34] C. W. Churchman, *The design of Inquiring Systems: Basic Concepts of Systems and Organizations*. New York, NY, USA: Basic Books, 1972.

[35] M. W. Manogharan, T. Thivaharan, and R. A. Rahman, "Academic staff retention in private higher education institute—Case study of private colleges in kuala lumpur," *Int. J. High. Edu.*, vol. 7, no. 3, pp. 52–78, 2018.

[36] O. Figueroa, "The influences impacting staff turnover in higher education," *J. Manag. Sustain.*, vol. 5, no. 4, pp. 86–93, 2015.

[37] Forbes. *Employee Turnover is the Highest It's Been in 10 Years*. Accessed: Feb. 26, 2018. [Online]. Available: https://www.forbes.com/sites/jeffboss/2018/02/26/employee-turnover-is-the-highest-its-been-in-10-years-heres-what-to-do-about-it#undefined

[38] B. J. Erasmus, A. Grobler, and M. Van Niekerk, "Employee retention in a higher education institution: An organisational development perspective," *Progressio*, vol. 37, no. 2, pp. 32–62, 2015.

[39] *2018 Retention Report, Truth & Trends in Turnover*, Work Inst., Franklin, TN, USA, 2018.

[40] J. S. Edwards, "A process view of knowledge management: It ain't want you do, it's the way that you do it," *Electron. J. Knowl. Manage.*, vol. 9, no. 4, pp. 297–306, 2011.

[41] E. Bou and A. Sauquet, "Reflecting on quality practices through knowledge management theory: Uncovering grey zones and new possibilities of process manuals, flowcharts and procedures," *Knowl. Manage. Res., Pract.*, vol. 2, no. 1, pp. 35–47, 2004.

[42] S. Newell, M. Robertson, H. Scarbrough, and J. Swan, *Managing Knowledge Work and Innovation*, 2nd ed. New York, NY, USA: Palgrave Macmillan, 2002.

[43] I. Nonaka, "A dynamic theory of organizational knowledge creation," *Organ. Sci.*, vol. 5, no. 1, pp. 14–37, 1994.

[44] B. Rubenstein-Montano, J. Liebowitz, J. Buchwalter, D. McCaw, B. Newman, and K. Rebeck, "A systems thinking framework for knowledge management," *Decis. Support Syst.*, vol. 31, no. 1, pp. 5–16, 2001.

[45] P. Heisig, "Harmonisation of knowledge management—Comparing 160 KM framework around the globe," *J. Knowl. Manage.*, vol. 13, no. 4, pp. 4–31, 2009.

[46] International Society Educational Data Mining. *Educational Data Mining*. Accessed: Jun. 29, 2016. [Online]. Available: http://www.educationaldatamining.org/

[47] C. Romero and S. Ventura, "Data mining in education," *Wiley Interdiscipl. Rev. Data Mining Knowl. Discovery*, vol. 3, no. 1, pp. 12–27, 2013.

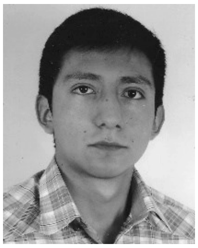
[48] G. Siemens and R. S. J. D. Baker, "Learning analytics and educational data mining," in *The Cambridge Handbook of the Learning Sciences*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2012, p. 252.

[49] M. Alavi and D. E. Leidner, "Knowledge management systems: Issues, challenges, and benefits," *Commun. AIS*, vol. 1, no. 2, p. 7, 1999.

[50] R. Maier and T. Hadrich, "Knowledge management systems," in *Encyclopedia of Knowledge Management*. Alberta, CL, Canada: IGI Global, 2006, pp. 442–450.
[51] A. H. Gold, A. Malhotra, and A. H. Segars, "Knowledge management: An organizational capabilities perspective," *J. Manage. Inf. Syst.*, vol. 18, no. 1, pp. 185–214, 2001.
[52] S. Lee and S. Hong, "An enterprise-wide knowledge management system infrastructure," *Ind. Manage., Data Syst.*, vol. 102, no. 1, pp. 17–25, 2002.
[53] R. Kimball, *The Data Warehouse Toolkit*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.
[54] O. Moscoso-Zea and S. Lujan-Mora, "Suggested methodologies for evaluation and selection of enterprise architecture software for knowledge digitization," *Enfoque UTE*, vol. 8, no. 1, pp. 315–328, 2017.
[55] M. Lankhorst, *Enterprise Architecture at Work Modelling Communication and Analysis*, 3rd ed. Berlin, Germany: Springer-Verlag, 2013.
[56] P. Johnson, L. Nordström, and R. Lagerström, "Formalizing analysis of enterprise architecture," in *Enterprise Interoperability: New Challenges Approaches*. London, U.K.: Springer, 2007, pp. 35–44.
[57] K. D. Niemann, *From Enterprise Architecture to IT Governance, Elements of Effective IT Management*. Berlin, Germany: Friedlander & Sohn, 2005.
[58] KPMG. Harvey Nash/KPMG 2017 CIO Survey Results. Advisory Institute. Accessed: Jan. 21, 2019. [Online]. Available: https://goo.gl/R8WHsv
[59] F. Sandoval, O. Moscoso-Zea, V. Galvez, and P. Tutillo, "Enterprise architecture framework with agile approach based on TOGAF," in *Proc. Int. Conf. Inf. Syst. Comput. Sci. (INCISCOS)*, 2016, pp. 77–81.
[60] W. H. Inmon, *Building the Data Warehouse*, 3rd ed. Hoboken, NJ, USA: Wiley, 2002.
[61] H. Bodo, J. Lechtenbörger, and G. Vossen, "Conceptual data warehouse design," in *Proc. 2nd Int. Workshop Design Manage. Data Warehouses (DMDW)*, 2000, pp. 1–11.
[62] S. Luján-Mora, J. Trujillo, and I.-Y. Song, "A UML profile for multidimensional modeling in data warehouses," *Data & Knowl. Eng.*, vol. 59, no. 3, pp. 725–769, 2006.
[63] O. Moscoso-Zea, A. Sampedro, and S. Luján-Mora, "Datawarehouse design for educational data mining," in *Proc. 15th Inf. Technol. Based High. Edu. Training (ITHET)*, Sep. 2016, pp. 1–6.
[64] A. Garcia-Holgado, F. J. Garcia-Peñalvo, Á. Hernández-Garcia, and F. Llorens-Largo, "Analysis and improvement of knowledge management processes in organizations using the business process model notation," in *New Information and Communication Technologies for Knowledge Management in Organizations*. Cham, Switzerland: Springer, 2015, pp. 93–101.
[65] P. Vassiliadis, P. Marcel, and S. Rizzi, *Beyond Roll-Up's and Drill-Down's: An Intentional Analytics Model to Reinvent OLAP (Long-Version)*. Ithaca, NY, USA: Cornell Univ. Press, 2018, pp. 1–74.
[66] M. Bienkowski, M. Feng, and B. Means. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. U.S. Department of Education's (ED) Office of Educational Technology. Accessed: Jan. 30, 2019. [Online]. Available: https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf
[67] R. Dehghani and R. Ramsin, "Methodologies for developing knowledge management systems: An evaluation framework," *J. Knowl. Manage.*, vol. 19, no. 4, pp. 682–710, 2015.
[68] R. Chalmeta and R. Grangel, "Methodology for the implementation of knowledge management systems," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59, no. 5, pp. 742–755, 2008.

OSWALDO MOSCOSO-ZEA received the Engineering degree in computer science from Pontifical Catholic University, Quito, Ecuador, in 2001, and the M.Sc. degree in information systems from Westfälische-Wilhelms Universität, Münster, Germany, in 2013. He is currently pursuing the Ph.D. degree in informatics with the University of Alicante, Alicante, Spain. Since 2003, he has been an Aggregate Professor with Universidad Tecnológica Equinoccial, Quito, where he has been the Director of the Computer Science Department. He is also a Microsoft Certified Professional in software development and a Subject Matter Expert in data management from CompTIA, Chicago, IL, USA. He has participated in several projects as an IT consultant in different countries. His main research interests include educational data warehousing, big data, enterprise architecture, and knowledge management.

JORGE CASTRO studied computer science. He is currently a Research Assistant with the Computer Science Department, Universidad Tecnológica Equinoccial. In recent years, he has focused on statistical learning and data mining. He has participated in several independent projects as the CEO at Widboo Operating Company, LLC. His main research interests include software development, data storage, and big data. He has been part of the organizing committee of several conferences and international scientific meetings.



JOEL PAREDES-GUALTOR studied computer science and economy with full scholarships. He is currently a Research Assistant with the Computer Science Department, Universidad Tecnológica Equinoccial. He has participated in different research projects and has been part of the organizing committee of several international scientific conferences and meetings. His main research interests include data warehousing, data analytics, model-driven software development, and enterprise architecture. He has received mentions as the best grade point average in the Engineering Faculty and in the Computer Science Department.

SERGIO LUJÁN-MORA was born in Alicante, Spain, in 1974. He received the Ph.D. degree in computer engineering from the Department of Software and Computing Systems, University of Alicante, Alicante, in 2005, and the Computer Science and Engineering degree from the University of Alicante, in 1998, where he is currently a Senior Lecturer with the Department of Software and Computing Systems. He has authored several books and many published articles in various conferences, including ER, UML, and DOLAP, and high-impact journals, including DKE, JCIS, JDBM, JECR, JIS, JWE, IJEE, and UAIS. His main research interests include data warehousing, Web development, Web accessibility, and usability. In recent years, he has focused on e-learning, massive open online courses (MOOCs), open educational resources (OERs), and the accessibility of video games.