Journal of Network and Innovative Computing.
ISSN 2160-2174 Volume 10 (2022) pp. 036–042

A Data Warehousing System for ICT Competencies Assessment

Vladimir Dobrynin[1], Michele Mastroianni[2] and Olga Sheveleva[3]

[1]Dubna State University,
Dubna, Moscow Oblast, Russia
i@vdobrynin.ru

[2]Università degli Studi di Salerno, Dipartimento di Informatica,
Via Giovanni Paolo II, 132 - 84084 Fisciano (SA), Italy
mmastroianni@unisa.it

[3]Dubna State University,
Dubna, Moscow Oblast, Russia
shoe.asp19@uni-dubna.ru

Abstract: **The development of the digital economy is one of the priority areas for most countries, so the requirements for ICT specialists are changing and increasing, as a consequence of these transformation. In this perspective, the strengthening, measuring and assessment of digital competences are becoming crucial to ensure quality and security of the digital products and services implemented. This paper proposes MMACK, a new Meta-model for assessing ICT competencies and knowledge. This work also explores the feasibility of the implementation of such a model using Data Warehousing software.**
Keywords: Competence model, Cybersecurity, Data Warehousing, OLAP, Hypercube.

I. Introduction

At the end of the 20th century and the beginning of the 21st century, the educational system is being modified in Europe and the United States. As a result, the National Skills Standard Board is being approved in the United States [24], and the Bologna Declaration on the Formation of a Single European Educational Space comes into force in Europe [1]. This declaration was also signed by Russia in 2003 [8], [23]. One of the main principles of the Bologna Declaration and the National Skills Standard Board is a competence-based approach to assessing learning outcomes. At the initial stages, the competence-based practice-oriented approach was considered as the prerogative of secondary vocational educational institutions; now it has actively penetrated into higher education as well. The peculiarity of the competence-based approach is that in the learning process not only ready-made knowledge is acquired, but the conditions for the origin of this knowledge are traced. The methodology is based on learning through activity. It is also important that each competence at each level should have a clear assessment system with clear results. Both the learning process and the assessment process should be as transparent and understandable as possible to all participants.

The competence-based approach is more complex than just a knowledge test: it is structured in a multi-component form. This approach is aimed at "overcoming the main drawback of the existing system of professional training: the gap between the theoretical and practical aspects of the professional activity formed in the course of training" [19]. NPEC work group defines competence as "combination of skills, abilities and knowledge needed to perform a specific task [5], [11]. This paper is in prosecution of a preliminary work [4], and aims to design a system for the implementation of a new methodology for assessing digital competencies in the ICT field, and in particular in the Cybersecurity-related topics.

The rapid growth of cybercrime has been confirmed by many studies [2]. CybersecurityVentures.com experts predicted that cybercrime will cost the global economy $ 6.1 trillion a year by 2021 and reach $ 10.5 trillion in annual losses by 2025 [2]. The recent Covid-19 pandemic and the consequently mandatory lockdown (and home-working activities) of business and public sector employees made the situation worse. Interpol reports an increasing rate in Cybercrime activities and new emerging threats related to the pandemic [3]. One solution to this problem can be the development of security competences for the conscious use of digital devices and the Internet. As a conclusion, the development of digital competences in general and security competences in particular is more of a necessity in the modern world than a super-

[1]Commission, E.U.: The Bologna process and the european higher education area, https://ec.europa.eu/education/policies/higher-education/bologna-process-and-european-higher-education-area_en

[2]Morgan, S.: Cybercrime to cost the world $10.5 trillion annually by 2025, https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021

[3]Covid19: Cybercrime analysis report. Interpol, august 2020, https://www.interpol.int/content/download/15526/file/COVID19CybercrimeAnalysis Report-August2020.pdf

fluity. The development of competences is a complex system: the development of educational material, a methodology for checking the formation of competences, etc. In the following chapters, a methodology for assessing the formation of competences and its implementation is proposed.

II. Related Work

Although the topic is not the most covered road by scholars, there are several research works about using Data Warehousing solutions in Higher Education field as a decision support tools.
An early paper (2003) to take into account is [25], in which is proposed the use of DWH software in strategic decision making for Universities, discussing opportunities, drivers and obstacles. Another early paper to examine is [3], in which are described the advantages in using DWH software in Academic environment in order to offer a better quality of the instruction and to manage human and economic-financial resources. The paper is focused on the experience of Italian Universities at the time, and also presented a complete data mart related to teaching activities.
Another paper [1] deals with an E-governance implementation for higher education system with the use of DWH and data mining techniques, In the paper is proposed a basic architecture and a logical design for an University database. In [14] are presented the main design issues of the implementation of a DWH system for educational data analysis, exploring logical design and OLAP choices. In [15] are described the main design considerations for the implementation of the DWH in an educational scenario, focusing on architecture design choices and Extract, transform and Load (ETL) tools. In [22] is explored how big data technology could be implemented with data warehouse to support decision making process of Higher Education sector, introducing *modern* Data Warehousing, aiming to help reduce difficulties associated with traditional data analysis with a big data technology approach to data warehouse. Finally, some recent experiences may be found in [17] and [26].
Regarding surveys on the topic, an extensive analysis of 34 empirical research papers published from 2008 to 2018 is presented in [16], This paper shows that the Kimball's approach is the most applied methodology for DWH design in education, and the Star schema is the most used approach to implementation. In this paper are also classified the papers analysed in six dimensions (schema proposal, analysis of the user requirements, analysis of the business requirements, effectiveness, implementation, and data analysis), and is highlighted that most papers are focused on schema proposal and implementation. In [10] is performed a survey on the use of DWH software in the field of education, focused on application in Indian universities.
The study of such related work has allowed us to draw up a synthetic table in which the main advantages and issues related to implementing of decision support systems for the Educational field using DWH software are summarised (Table II).

Table 1: Advantages and issues of using DWH in implementation.

Advantages
• If DWH software is already in use, the cost and the effort of this implementation may be very low
• If DWH software is not yet used, the introduction of this king of software may be used by University to implement a number of features to support decision-making processes (e.g. scientific production analysis, cash flow monitoring, ...)
• The data may be automatically extracted from existing Student Career University systems
• Ease of competences model extension (e.g. adding new attributes)
• Opportunity to perform competence analysis in relationship with other student's attributes (curriculum, learning cycle, ...)
Issues
• The use of commercial (i.e.: non open-source) DWH software may be very costly
• The introduction of DWH could lead to a heavy organisational effort for the University

III. The methodology for assessing the level of competence

For the full-fledged development of the competence-based approach, it is necessary to create a transparent, understandable system for assessing competences. In this part we present MMACK, a new *Meta-model for assessing ICT competencies and knowledge* which allows to assess the level of competence formation among students, as well as track a path of their competences development. The source of the research is the result of the authors' many years of work, tested in the educational process in various disciplines at the Dubna State University. The level of competence is one of the principles of organizing the technology of acquiring meta-knowledge by students. Meta-knowledge is the basis for creative independent acquisition of new knowledge and competencies. This approach improves the organizational, scientific and methodological activities of the teaching staff, as well as the structure and content of the educational and methodological complex of disciplines. Confirmation is the topics of defended master's theses, published articles.
In the proposed model the competence level (CL) is defined as a knowledge of the subject area. We believe, the competence level can be measured as a function (F) of three variables: *UNderstanding* (UN), *Have The Skill* (HSK) and *ABility* (AB).

$$CL = F(UN, HSK, AB) \tag{1}$$

In the Formula 1 the three variables are defined as follows. It is to be noted that the HSK and AB are both *Skills*, which are automated components of a person's conscious action, which are generated in the process of its implementation.

- **UN** - the student's ability to recognize the meaning (content) of a text or speech. This ability is based on

Table 2: Samples of tests for assessing competence variables.

Type of test	Samples
UN test	Students need to find its meaning for the corresponding term.
HSK test	Students have to determine what essence is reflected in the selected text and how this essence is highlighted.
AB test	To give students different interpretation of a concept and they have to note the similarities and differences between interpretations of the concept.

vocabulary of words, concepts, definitions, terms, abbreviations and etc. Generally, a teacher relies on the previous acquired basic knowledge or makes an analogy with the processes already studied when explained a new topic or giving a lecture in a discipline. Additionally, a student has a feeling or adequate perception of the topic, and he grasps what the teacher is explaining.

- **HSK** - the ability of a student to solve typical problems of the discipline in an automatic mode (without the involvement of consciousness, on reflexes). Solving problems without involving reflexes can be achieved by solving typical tasks or through training. Important indicators of skill are - the time spent on solving the problem and the quality of the solution (there are mistakes or not).

- **AB** - the student's ability to compare, compare - identify similarities and differences - highlight, argue, reason. The method of performing an action mastered by the subject, provided by the totality of acquired knowledge and skills; the ability to perform an action according to certain rules, and the action has not yet reached automation. It is formed through exercises and creates the ability to perform an action not only in familiar, but also in changed conditions. AB has two interpretations, in contrast of UN: be able to use certain methods that allow you to make a decision; highlight the meaning and essence of what they are talking about or about what they write. This concept reflects the fact that the student not only understood the material, but also analyzed it: comparison, juxtaposition, highlighting, reasoning. Consequently, one chain of thoughts, generate another chain of thoughts - a chain of cause and a chain of effect in reasoning.

The levels of UN, HSK and AB are determined by a system of testing rules (ST). ST includes: tests (TST), testing rules (W), rules for assessing the test result. As a result of testing, the levels of UN, HSK and AB are determined. It is assumed that the tests to be carried out multiple times. We assume that tests are conducted once time a semester, however, the frequency of tests may vary depending on the education (training, teaching) system. The Figure 1a shows the change of CL depending on time (the figure depicts the CL of 4 students who were measured three times in some intervals).
Moreover, this system allows to track the change in each indicator of CL (UN, HSK, AB) over time. Figure 1b shows the CL model, where the x, y, z axes correspond to the level UN, HSK, AB. Consequently, the cube reflects a set of levels UN, HSK, AB. The scale of levels UN, HSK, AB consists of

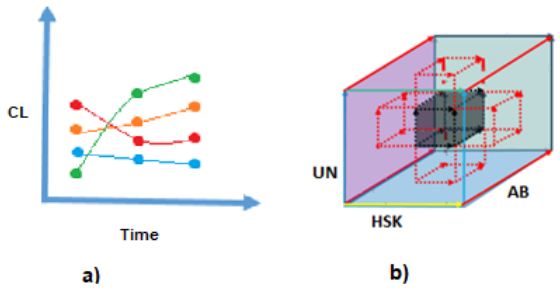

Figure. 1: a) Time-dependent student measurement graph; b) Three-dimensional representation of changes in student knowledge.

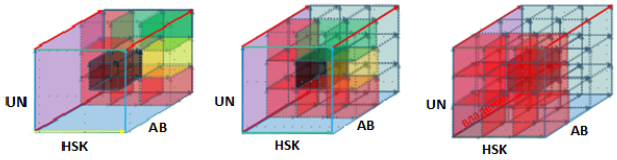

Figure. 2: Displaying levels of UN, HSK, AB.

3 levels: low, average and high. The black cube reflects the medium level UN, HSK, AB.
The scale of levels UN, HSK, AB consists of 3 levels, Therefore, the cube we are considering consists of 27 cubes - spaces where a point with a student's level of knowledge can be (Fig. 2). From the central cube, cubes can be laid left-right, up-down along three axes (UN, HSK, AB). Red, yellow and green are the colors of the cubes that correspond to the low, average and high CL levels.
A sample of three-dimensional representation of students' knowledge over time is presented in the Figure 3. The y-axis is the names of the students and the z-axis is the time axis. This system allows to make a slice for a particular student and see the dynamics of changes in the level of each indicator.
General recommendations for studying, recommendations of

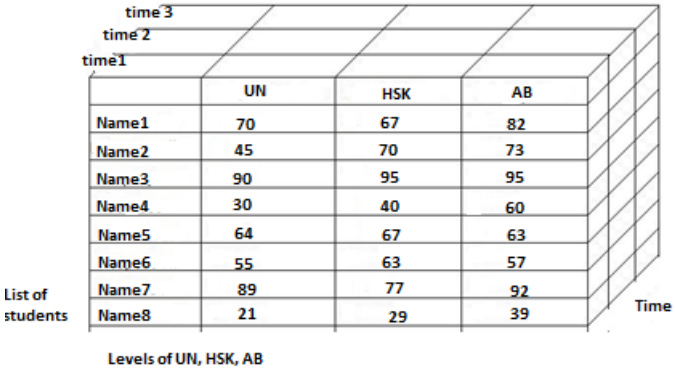

Figure. 3: Three-dimensional representation of students' knowledge over time.

Table 3: Comparison between Inmon and Kimball approaches (Based on [15]).

	Inmon	Kimball
Initial time	High	Low
Initial cost	High	Low
Maintenance	Simple	Expensive
Skills required	Specialist team	Generalist team
Data requirements	Enterprise Wide	Single business areas

additional materials, and also refine the education program can be made based on the results analyze of the CL for a specific group of students (Fig. 3). Those data can be used for a deeper analysis for each student: to determine the dynamics of particular student development in a certain subject area and, on the basis of it, make recommendations for further studying.

IV. Architectural choices

A Data Warehousing may be defined as *a data-driven decision support system that supports the decision-making process in a strategic sense and, in addition, operational decision-making* [13]. The most used different methodologies proposed in designing a DWH are Inmon [9] and Kimball [12]. The main difference between these methodologies is the approach used to implement the data marts: Inmon proposes a top-down approach whereas Kimball proposes a bottom-up approach.
Inmon's methodology provides unique dimensional view of data across data marts, building a centralized Data Warehousing, and all data marts will be loaded from that central DWH (Fig. 4a). The top-down approach is flexible to support change management as it looks at the organization as a whole. On the other hand, Kimball presents a different vision of data warehouses. He considers that the data warehouse can be seen as a set of consistent data marts and based on shared conformed dimensions (Fig. 4b). The main difference from both methodologies is that from the Inmon's point of view, an enterprise predefined DWH is needed, and in a further step the data marts will be extracted, while in the Kimball's approach the data marts are created directly involving the business units before the creation of an enterprise DWH, which is an aggregation of all data marts. [15]. A summary of the main pros and cons of both approaches is shown in Table 3.
Due to the experimental nature of this work, we choose the Kimball approach, which ensures shorter startup time and lower cost. Moreover, in this way the experimental system is suitable to support the work of independent teams working on different aspects.
Significant components in the implementation of DWH projects are the Extract, Transform and Load (ETL) tools, which have in charge the process of extracting data from the Enterprise database and other sources and and importing them into the DW. Those tools operate in three steps [18]:

- extracting data from various sources sources (SQL DB, XML files,...) and converting into a single format appropriate for the transformation and loading process
- converting and normalising data (measure units, different character sets,...)
- loading data in data marts

The design of ETL tools mainly depends on the particular software used for DWH system, and on the quality of data to be processed. In our project, we assume that the data used for competencies assessment originate from standardised format for test data results, and a standardised ETL tool will be enough to ensure the correctness of processing. In future plans, we think that also other kind of data will be used for assessment, and on that occasion we will plan to design ad-hoc ETL tools.

V. Design of the proposed system

In DWH systems, the information is delivered to users using Online Analytical Processing (OLAP), which is a technique that enables multi-dimensional and multi-level analysis on a large volume of data, providing aggregated data visualizations with different perspectives [21]. The OLAP processing is based on multidimensional analysis of data in the form of "Facts"; a fact is a measure (or a set of measures) which represents a fact about the managed entity or system. In Figure 5 is shown an example of the hypercube representation of the sales of a firm: The measurement of fact are the units sold and the amount of the revenue, and the dimensions of hypercube are Time, Products and Branch related to a particular Sales fact.
It is clear that the OLAP hypercube structure fits in perfectly with the representation of our competences model as shown in Figure 3. We assume also that the results of competencies assessment test (in a structure similar to Figure 2) have been loaded to the Student Career DB of the University.
For the implementation, a Kimball two-layers architecture is chosen [7]. The raw data, originating from the Student Career DB, are subjected to a filtering phase in a Data Staging area using ETL (Extract, Transform and Load) Tools and loaded to the appropriate Data Marts, and then the users can process the data of own interest via Data Marts (Fig. 6).
First of all, we choose to design the data relationship using a conceptual model to provide a higher level of abstraction in describing the warehousing process and the structure of the information needed in all its aspects. The model used Dimensional Fact Model (DFM) by Golfarelli et al [6], due to its simplicity and ease of implementation. In our preliminary design, we have one Fact, three variables and three dimensions of interest:

- Fact: Competences_Test
- Variables: UN, HSK and AB (Equation 1)
- Dimensions: Time, Students, Course

The variable Time is hierarchically structured in two alternative ways, Day-Month-Year and Day-Semester-Year, in order to enable queries both based on *academic* time or in natural time. Also the dimension Student is structured in a simple hierarchy: we use a generic Student_Career as an example, but many different structures may be designed using the data stored in Student Career DB of the University. The DFM diagram for Competences tests is drawn in Figure 7.

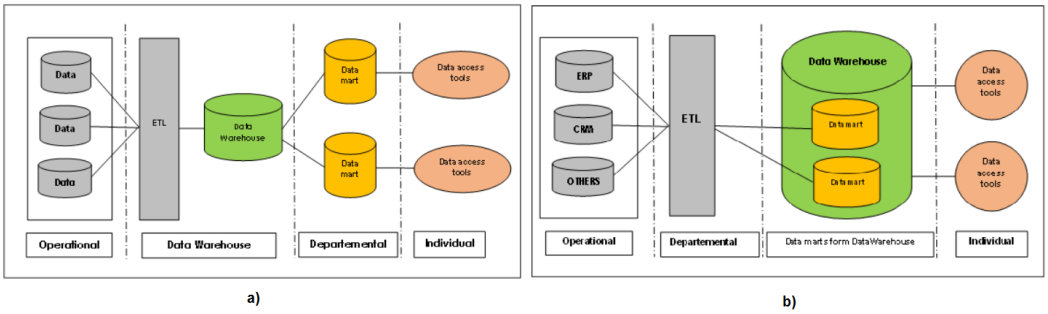

Figure. 4: The two approaches to DWH design: a) Inmon b) Kimball.

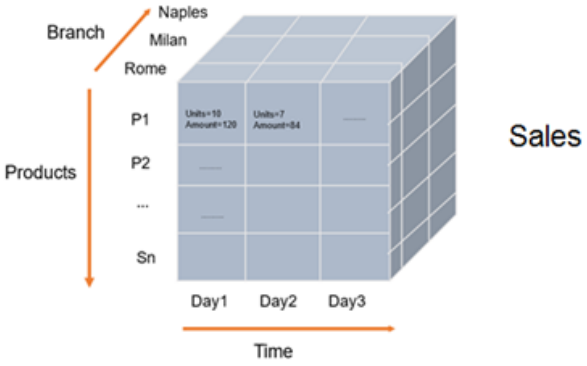

Figure. 5: The hypercube structure of Facts.

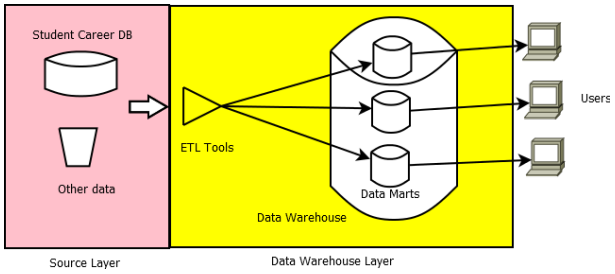

Figure. 6: Basic architecture of DWH system.

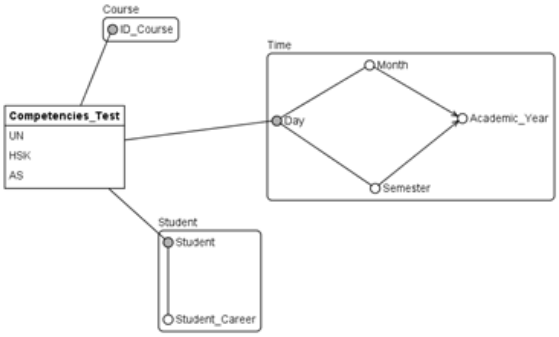

Figure. 7: DFM conceptual diagram of Competences test.

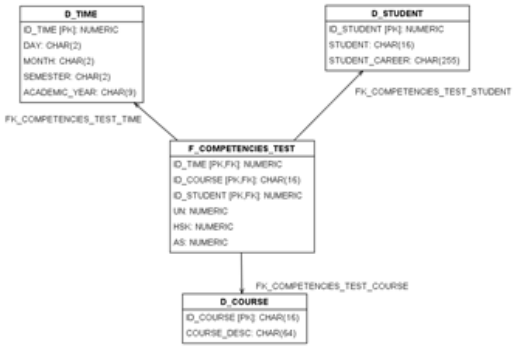

Figure. 8: The STAR schema of Competences_Test.

At this stage, is required to produce the STAR schema [20] which defines the tables, the attributes, and the relationships, according with the data structure represented using the DFM model previously designed. The STAR schema has been automatically obtained using a CASE tool, exploiting its ability to forward-engineer the conceptual model and generate the schema for the target database system you need to work with. In this research We use the experimental CASE tool BIModeler (Business Intelligence Modeler) [4].
In Figure 8 is drawn the STAR schema, in which the table beginning with "F_" is the fact and the tables beginning with "D_" are the three dimensions (Course, Student and Time).

VI. Conclusions and future work

In this paper has been presented a new competence metamodel MMACK, the future works for competence model will to develop tests which allow to estimate competence variables (UN, HSK, AB). The tests will be focused on measurement the level of Cybersecurity-related competences. In this research work has also been discussed a possible implementation of the competence model using Data Warehousing software, and have been presented a basic architecture and the conceptual and logical design. The future work for implementation of the system will be the choice of the DWH software to be used; in this way a proof-of concept of a working system will led us to conduct the first practical experiments.

[4]Cazzella, S.: Business intelligence modeler, alpha version - b115 http://www.bimodeler.com/

References

[1] Bhanti, P., Kaushal, U., Pandey, A., et al.: E-governance in higher education: Concept and role of data warehousing techniques. International Journal of Computer Applications **18**(1), 875–895 (2011)

[2] Conteh, N., Royer, M.: The rise in cybercrime and the dynamics of exploiting the human vulnerability factor. International Journal of Computer (IJC) **20**(1), 1–12 (Jan 2016)

[3] Dell'Aquila, C., Di Tria, F., Lefons, E., Tangorra, F.: An academic data warehouse. In: Proceedings of the 7th WSEAS International Conference on Applied Informatics And Communications. pp. 24–26. Citeseer (2007)

[4] Dobrynin, V., Mastroianni, M., Sheveleva, O.: A new structured model for ict competencies assessment through data warehousing software. In: International Conference on Innovations in Bio-Inspired Computing and Applications. pp. 435–446. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-96299-9_42

[5] Eremeev, M.: Model of professional competence for the assessment of professional risks (in russian). Science vector of Togliatti State University. Series: Pedagogy, psychology. pp. 14–21 (2022). https://doi.org/10.18323/2221-5662-2022-1-14-21

[6] Golfarelli, M., Maio, D., Rizzi, S.: The dimensional fact model: A conceptual model for data warehouses. International Journal of Cooperative Information Systems **07**(02n03), 215–247 (1998). https://doi.org/10.1142/S0218843098000118

[7] Golfarelli, M., Rizzi, S., Turricchia, E.: Modern software engineering methodologies meet data warehouse design: 4wd. In: Cuzzocrea, A., Dayal, U. (eds.) Data Warehousing and Knowledge Discovery. pp. 66–79. Springer Berlin Heidelberg, Berlin, Heidelberg (2011)

[8] Grigorevskya, L.B., Ivashchenkob, G.A., et al.: The bologna process in russia: history, contradictions and prospects. Issues of social-economic development of Siberia pp. 97–103 (2021). https://doi.org/10.18324/2224-1833-2021-1-97-103

[9] Inmon, W.H.: Building the data warehouse. Wiley (1993)

[10] Jayashree, G., Priya, C.: Comprehensive guide to implementation of data warehouse in education. In: Peng, S.L., Son, L.H., Suseendran, G., Balaganesh, D. (eds.) Intelligent Computing and Innovation on Data Science. pp. 1–8. Springer Singapore, Singapore (2020)

[11] Jones, E.A., Voorhees, R.A.: Defining and Assessing Learning: Exploring Competency-Based Initiatives. Report of the National Postsecondary Education Cooperative Working Group on Competency-Based Initiatives in Postsecondary Education. U.S. Department of Education, National Center for Education Statistics (2002)

[12] Kimball, R., Ross, M.: The data warehouse toolkit: the definitive guide to dimensional modeling. John Wiley & Sons (2013)

[13] Linstedt, D., Olschimke, M.: Building a scalable data warehouse with data vault 2.0. Morgan Kaufmann (2015)

[14] Mirabedini, S.: The role of data warehousing in educational data analysis. Journal of Novel Applied Sciences **3**(5), 1439–1445 (2014)

[15] Moscoso-Zea, O., Andres-Sampedro, Luján-Mora, S.: Datawarehouse design for educational data mining. In: 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET). pp. 1–6 (2016). https://doi.org/10.1109/ITHET.2016.7760754

[16] Moscoso-Zea, O., Paredes-Gualtor, J., LujáN-Mora, S.: A holistic view of data warehousing in education. IEEE Access **6**, 64659–64673 (2018). https://doi.org/10.1109/ACCESS.2018.2876753

[17] Neamah, A.: Adoption of data warehouse in university management: Wasit university case study. In: Journal of Physics: Conference Series. vol. 1860, p. 012027. IOP Publishing (2021)

[18] Nwokeji, J.C., Aqlan, F., Anugu, A., Olagunju, A.: Big data etl implementation approaches: A systematic literature review (p). In: SEKE. pp. 714–713 (2018)

[19] Nyushchenko, V.: On the readiness of the teaching staff to implement the competence based approach in teaching [in russian]. pp. 264–266 (2012). https://doi.org/10.26439/ciis2021.5581

[20] Ponniah, P.: Data warehousing fundamentals for IT professionals: Second Edition. John Wiley & Sons (2010). https://doi.org/10.18178/ijiet.2019.9.6.1237

[21] Queiroz-Sousa, P.O., Salgado, A.C.: A review on olap technologies applied to information networks. ACM Transactions on Knowledge Discovery from Data (TKDD) **14**(1), 1–25 (2019). https://doi.org/10.1145/3370912

[22] Santoso, L.W., Yulia: Data warehouse with big data technology for higher education. Procedia Computer Science **124**, 93–99 (2017). https://doi.org/10.1016/j.procs.2017.12.134, 4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia

[23] Shulgaty, L., Golik, V.I.: To implementation of the bologna declaration in russia. Scientific bulletin of the Southern Institute of Management pp. 52–55 (2016). https://doi.org/10.31775/2305-3100-2016-2-52-55

[24] Voorhees, R.A.: Competency-based learning models: A necessary future. New Directions for Institutional Research **2001**(110), 5–13 (2001)

[25] Wierschem, D., Mcmillen, J., McBroom, R.: What academia can gain from building a data warehouse. Educause Quarterly **26**(1), 41–46 (2003)

[26] Yu, X.: The application of data warehouse in teaching management in colleges and universities. In: Journal of Physics: Conference Series. vol. 1738, p. 012090. IOP Publishing (2021)

Author Biographies

Dobrynin Vladimir Nikolaevich Education - higher (Kharkiv State University, Faculty of Mechanics and Mathematics, specialty - mathematician). Science degree Candidate of Technical Sciences, Senior Researcher. Professional activity: Management of complex systems (technical, sociotechnical under conditions of uncertainty). Scientific work Modeling of aircraft control systems Search and exploration of deposits of solid minerals. Teaching activities: Since 1994 at the Dubna University in the fields of systems theory, computer analysis of dynamic systems, knowledge management technologies, models and methods for analyzing design solutions). Position: professor, General scientific experience: 60 years

Michele Mastroianni is currently an Assistant Professor of Computer Science at University of Salerno, Italy, where he teaches Cybersecurity and Fundamental of Programming. He holds a MSc degree (1989) in Electrical Engineering and a Ph.D. degree (2006) in Management Engineering at University of Naples Federico II. He taught since 1994 in Computer Science courses at University of Naples Federico II and University of Campania. He is author of several scientific papers in peer-reviewed International Journal and Conferences, in the field of Computer networks, Security and privacy, Simulation of complex system and network architectures, and Information systems management. He holds more than 30 years of work experience in Privacy, Computer Networks and Management of Information Systems, and has been project leader and expert for several local, regional and national technical and research projects.

Sheveleva Olga Evgenevna took her Bachelor in system analysis and management (profile of program: Business Informatics) in the year 2017 at Moscow Region State Funded Institution of Higher Education “Dubna State University”. She took her Master in system analysis and management (profile of program: Cognitive computing and business analytics) in 2019 at the same University; Master’s Thesis:” Development of spatial awareness of students with the use of 3D technologies”.Actually she is a 3rd year Ph.D. student of Moscow Region State Funded Institution of Higher Education “Dubna State University” from 2019 – until now. Her research fields (topic of future thesis): “Development of a new structured model for ICT competencies as a part of Intellectual Digital University of the Future”.