The 16th International Scientific Conference
eLearning and Software for Education
Bucharest, April 23-24, 2020
10.12753/2066-026X-20-178

ANALYSIS AND MODELING THE DOMAIN OF OPEN EDUCATIONAL RESOURCES FROM LEARNING ANALYTICS PERSPECTIVE

Malinka IVANOVA
Technical University of Sofia, Bulgaria
m_ivanova@tu-sofia.bg

Gabriela GROSSECK
West University of Timisoara, Romania
gabriela.grosseck@e-uvt.ro

Carmen HOLOTESCU
IoanSlavici University, Timisoara, Romania
carmen.holotescu@islavici.ro

***Abstract**: Open Educational Resources (OERs) contributes to ways for knowledge sharing and reusing among given scientific and academic societies. The students and teachers have rapid and unlimited access to content, prepared for different purposes and digital formats. Many projects have been started to extend and improve the existing digital collections. A few analytical tools are developed for automatic identification and monitoring of newly created and added OERs and for tracing the changes in existing OERs. The researchers have applied learning analytics techniques for discovery, analysis and reflection on given characteristics of emerged online content.*

The aim of the paper is to review, analyze and model the domain of the OERs focusing on the following research questions: What is the current state of research in the area of OERs?, What are typical features of OERs? What kind of tools are used for OERs development, tracing and analysis?, How learning analytics contributes to OERs creation, improvement and reusing?

The applied methodology consists of several procedures related to: discovery of appropriate scientific literature in abstract and citation databases and usage of search engines, indexing and delivering the full text or metadata concerning available scientific production; summarization and analysis of returned results through bibliometric and machine learning methods and techniques; development of analytical models that outline the current achievements and predictive models that show the future trends.

The findings point out the educational context of application of OERs, the technological approaches for automatic identification and analysis of OERs, the tools for OERs creation from pedagogical and technical point of view, the attainments of learning analytics for further development and enhancement of OERs.

Modelling the domain of OERs leads to better understanding how the OERs are created and utilized, what are the main areas of their usage, what are the future directions of evolvement. The models could be used by teachers and researches for outlining the current findings and for decision making about OERs application and topics for research identification.

***Keywords:** Open educational resources; learning analytics; machine learning; modelling.*

I. INTRODUCTION

Many projects at national and international level have been started in order to research the needs and requirements for open content and also to develop methodologies and tools for Open

Educational Resources (OER) creation. Digital and open education is connected to increase students' awareness, possibilities for lifelong learning and also to a collaborative process of continuous improvement the educational quality. OER initiatives are recognized as innovative approach for proposition of boundless generation and sharing of knowledge.

This extensive production of learning materials requires learning analytics algorithms and tools to track the emerged content, to analyze it and to improve/remove the re-mixed learning materials. *Resource Inspection, Selection, and Enhancement Framework* (RISE) is just one approach presented by Bodily et al. that automatically compare the students' use of resources and their grades with the aim similar and different resources among the utilized learning content in one course to be identified [1]. The result is related to continuously course content optimization as ineffective resources are improved or removed. On the other hand, Piedra and Caro report the development of a platform (called Serendipity) that consists of two analytical tools: a search engine for datasets extraction from digital repositories of OpenCourseWare (OCW) and OERs and a map that gives graphical information about projects related to open courses and open resources [2]. Another analytical tool for tracing the OERs in support of educators is proposed by Avila et al. [3]. It is based on the principles of Universal Design Learning and facilitates material/content construction and evaluation the quality of OERs. The tool is implemented as a module from a Learning Management System. It can be seen that analytical tools and frameworks for discovery and improvement of OERs elaborate in parallel with the continuously creation of open content and its usage for formal, non-formal or informal learning.

In the same time, the University Libraries take this role of knowledge distributorof OERs, assisting all educational participants in their activities related to interacting with OERs. At this moment, only one part of university libraries proposes to their students and educators a wide variety of services for fast access, search, re-mix and evaluation of OERs. For example, the Illinois University Library not only gives extended information, tips, and guides regarding different aspects of OERs, but also the Openly Available Sources Integrated Search (OASIS) tool is integrated on its web site [4]. Another example comes from University of Washington Libraries which facilitate the research, knowledge discovery and open content development through extensive catalogues, classifications, and guidance, concerning available open services and digital repositories. Also, it proposes information about existing authoring tools and ToolKits for open contend creation [5]. Moreover, the library of the University of Queensland informs about creative commons licenses and copyright law at production, re-usage and distribution of OERs, points out the collections of open textbooks, ebooks, open journals, search engines for open image archives, audio and video collections, open software and applications [6]. Also, a list with Massive Open Online courses (MOOCs), MOOC platforms and MOOC search engines is published. A check list for evaluation of OERs is designed for identification the most suitable open content by students and researchers for a given purpose.

Recently, the William and Flora Hewlett Foundation have noticed the need from better technical infrastructure for OERs that includes: tools for authoring and automated feedback, tools for metadata management, analytical tools, tools for import/export in different platforms and formats [7]. Barker and Campbell in their report [8] examine the importance of technologies for OERs dissemination and that includes: digital repositories for specific and general content and web sites for sharing different content types.

Therefore, the aim of this paper is to outline the achievements in the area of OERs in order to analyze the current state and to point out the future trends. For this purpose analytical and predictive models are developed, based on statistical and machine learning techniques. The paper is organized as follows: the second section explains the research methodology, the third section outlines the OERs landscape discussed in the scientific publications, the fourth section proposes several analytical and predictive models regarding the OERs domain and the last one is concluding with some discussions and issues for future studies.

II. METHODOLOGY

The research methodology consists of: a) Search literature procedure that includes formation of query request for scientific papers identification in the largest abstract and citation databases SCOPUS, Web of Science, and Dimensions, as well as in search engines for scientific production like

Google Scholar and Semantic Scholar. The main search query includes: *open educational resources* and the complementary queries add key words and group words such as *learning analytics*, *software*, *tools*, *higher education*, and *scientific production*. Then, found results are filtered according to: year of publication, document type and source type and additionally sorted on relevance. Finally, the bibliographic data describing the query results are extracted in appropriate formats and prepared for analysis (the experiment is done in December 2019). b) Construction and visualization of bibliometric networks with support of software VOSviewer [9] and performance of scientometrics and bibliometrics analysis through RStudio, Bibliometrix and Biblioshiny applications[10]. SCOPUS and Web of Science allow data extraction to be performed at most, respectively 2000 documents and 500 documents. Downloaded data from Dimensions concern 2500 documents. c) Applying machine learning algorithm for linear regression through Octave software platform and algorithm for time series prediction through framework for neural networks construction Neuroph Studio to find some trending issues.

III. DRAWNING OPEN EDUCATIONAL RESOURCES LANDSCAPE

Following the above described methodology, the bibliometric network is constructed with a dataset from SCOPUS through the software VOSviewer. Firstly, the results are obtained only for the term: *open educational resources*. The chosen type of analysis is *co-occurrences* of author keywords (terms) in scientific papers. Two terms are connected only with one link that could be with different thickness and it depends of how many publications contain these two terms together [11]. The applied counting method is *full counting* according to which each link is characterizes with strength 1 and if the co-occurrences are n, then the total co-occurrence of this term is n. Figure 1 shows the connection and the strength of the term *open educational resources* with other terms and the created visualization is used to outline the OERs landscape. We classify the major connected terms to the term OERs in 8 groups that are different than the automatically generated clusters by VOSViewer: Learning analytics, Pedagogy, Copyright, Learning context, Quality in OERs, Technology, Infrastructure, and Services that are further explained through additional terms as it is presented in Table 1. The number of occurrences (O) and the total link strength (TLS) for each term is also shown (in the following format O/TLS).

In the visualization featured on Figure 1, each term is represented by a bubble. The size of a bubble is directly proportional to the number of publications which contain the term analyzed. Usually, terms co-occurring often are tending to be located close to each other [12].

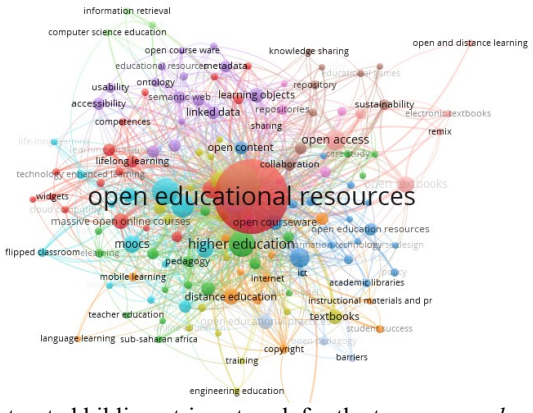


Figure 1. The constructed bibliometric network for the term *open educational resources* (keywords co-occurrence network map, VOSviewer network visualization)

According to these terms extracted from the obtained bibliographic data we can conclude that the scientific discussions connect OERs to: appropriate infrastructure mainly realized through digital repositories, to contemporary technologies like web 2.0, semantic web and cloud computing, to many

services for content authoring and interacting, to suitable pedagogical scenarios. Also, the attention is focused on OERs quality, copyright law and tools for content analysis. OERs are used in different learning context – in distance and blended learning, in engineering education and teachers training, collaborative and self-regulated learning.

Groups (O/TLS)	Additional terms (O/TLS)
Learning analytics (16/88)	content analysis (11/50), information retrieval (7/25), linked data (26/109), metadata (17/100)
Pedagogy (18/85)	instructional design (18/156), connectivism (7/25), competences (10/51), personalization (14/95), social networks (27/144), serious games (5/12), course design (11/44), peer production (8/50), learning design (17/55), curriculum design (12/83), assessment (9/66), teaching (23/158)
Copyright (10/81)	open license (9/40), creative commons (19/92), policy (8/30)
Learning context (17/88)	distance learning (48/301), language learning (8/27), mobile learning (27/152), eLearning (89/365), life-long learning (23/130), MOOC (132/482), teachers training (17/69), collaborative learning (29/121), engineering education (7/25), HE (74/308), blended-learning (22/103), PLE (11/47), self-regulated learning (10/49)
Quality in OERs (11/39)	quality assurance (12/72), quality model (5/16), quality control (1/30), quality criteria (4/11)
Technology (15/46)	Web 2.0 (16/97), cloud computing (12/40), web accessibility (6/23), widgets (6/33), wiki (6/52), youtube (5/23), ontology (18/68), semantic web (22/128), search engines (7/37)
Infrastructure (6/21)	learning objects repository (12/52)
Services for content creation and interacting (12/37)	interactive learning resources (5/20), open textbooks (5/22), open courseware (53/236), open access (46/276), usability (11/34), reuse (21/100), sharing (6/22), learning objects (9/35)

Table 1. The terms connected to the term *open educational resources*

The RStudio is an integrated environment with open source tools and applications for data processing and analysis and Biblioshiny is one of these tools used for quantitative exploration and visualization of bibliographic datasets. The advantages of the applied approach could be summarized as follows: (1) it gives possibilities for working with custom datasets with bibliographic data extracted from SCOPUS, Web of Science and Dimensions, (2) it provides very detailed bibliometric analysis based on co-word analysis, coupling method, analysis of scientific collaboration and co-citation, (3) it presents different aspects of the analyzed datasets in the form of visual graphics.The results after performed analysis on datasets obtained from SCOPUS, Web of science and Dimensions are similar and this is the reason the generated graphics to be presented only from one of the explored data sources. On Figure 2 the scientific production during the years according to SCOPUS is shown and it can be seen that the interest to this topic increase from year to year. Figure 3 gives information about the countries of the most cited authors and among the top ten are: USA, UK, Canada, Australia, Germany, Spain, Greece, Netherlands, France and Finland.

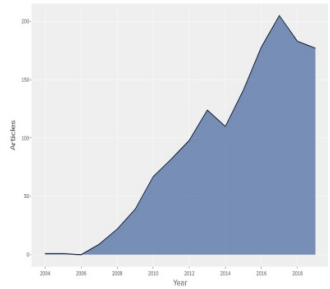


Figure 2. The dependency between number of articles and years (generated by Biblioshiny)

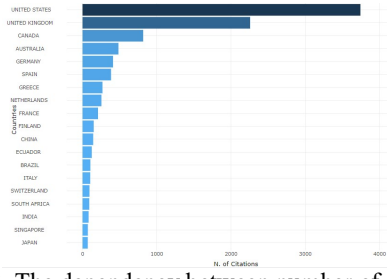
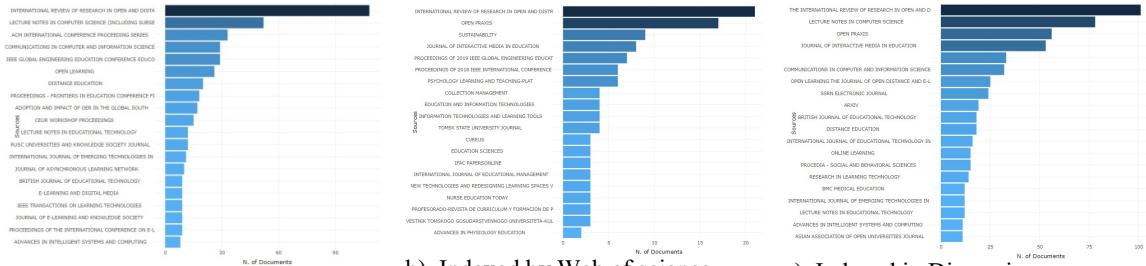


Figure 3. The dependency between number of citations and countries (generated by Biblioshiny)

The journals and conference proceedings that publish scientific papers with topics related to the OERs are shown on Figure 4. Among the top three sources according to SCOPUS are: *International Review of Research in Open and Distance Learning* (IRRODL), *Lecture Notes in Computer Science*, *ACM International Conference Proceedings*. According to Web of Science we have on Top 3 publishing sources - *International Review of Research in Open and Distributed Learning*, *Open Praxis*, and *Sustainability*. As far is concerning Dimensions we spot *International Review of Research in Open and Distributed Learning*, *Lecture Notes in Computer Science* and *Open Praxis*. All these journals are internationally recognized (with high impact factor), open access, peer-reviewed and with topics related to open, distance and distributed learning. One notable is *Sustainability*, an interdisciplinary journal that publishes papers in human sustainability from social, economical, environmental and cultural perspective. Also, it can be seen that among the most relevant sources for publishing is the *Lecture Notes in Computer Science* which includes technically oriented scientific articles that leads to the conclusion for the increased interest to the OERs from the side of computer scientists.



a). Indexed in SCOPUS b). Indexed by Web of science c). Indexed in Dimensions

Figure 4. The most relevant publishing sources (generated by Biblioshiny)

The frequency of the used terms in the scientific production during the years in logarithmic scale is presented on Figure 5. With higher frequency are terms like: higher education, eLearning, MOOCs, open access, online learning, learning objects, open textbooks, open educational practices, blended learning. The terms used in the context of OERs during the past year are: textbooks, open pedagogy, academic libraries, serious games, engineering education and STEM.

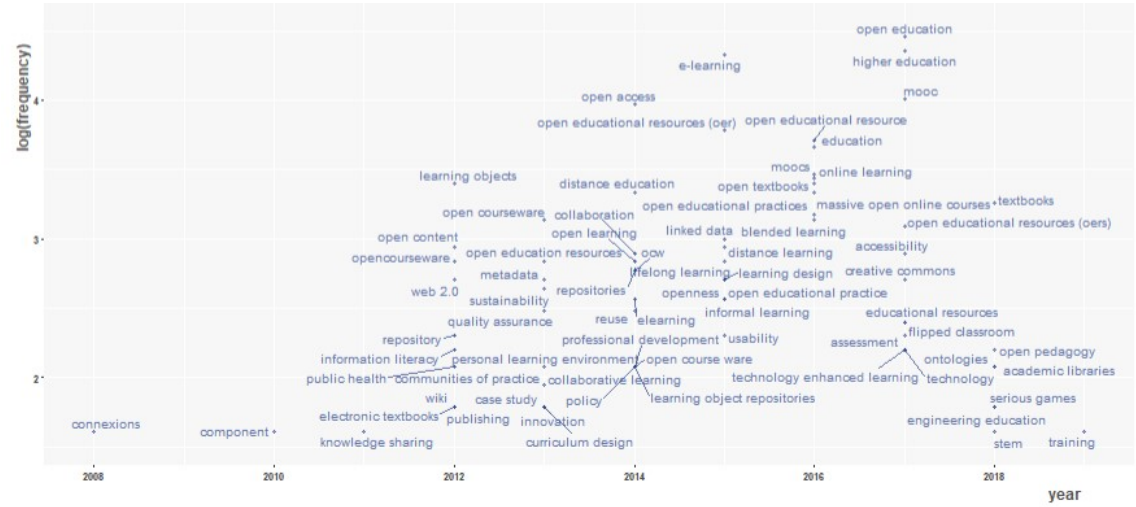


Figure 5. The dependency between frequency of the used terms and years (visualization, made by Biblioshiny)

It is interesting to understand whether authors from different countries collaborate among each other or the OERs initiatives, projects and publications are isolated events. For this purpose, the country collaboration map is visualized on Figure 6 and it can be seen that the research in the domain of OERs is based on authors' collaboration, good practices exchange and performance of collective

work. Examples for Bulgaria and Romania are provided on Figure 7. Bulgarian authors are in close connection to authors from USA, Spain, Germany, Australia, Italy, Romania, France, North Macedonia and Hungary, while Romanian authors collaborate with researchers from Spain, Germany, Italy, Bulgaria, Portugal, Slovenia, North Macedonia, Hungary, and Malta.

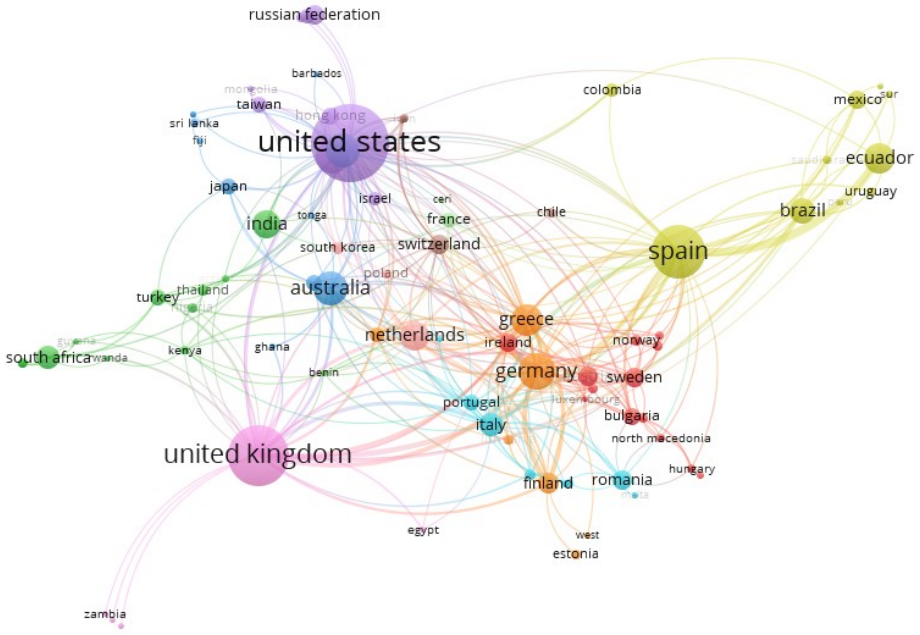
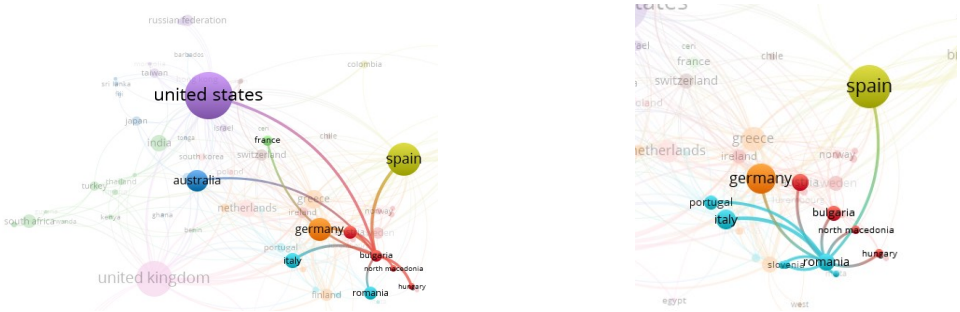


Figure 6. Country collaboration map (VOSviewer visulisation network)



a). Bulgaria b). Romania

Figure 7. Collaboration maps (made through VOSviewer)

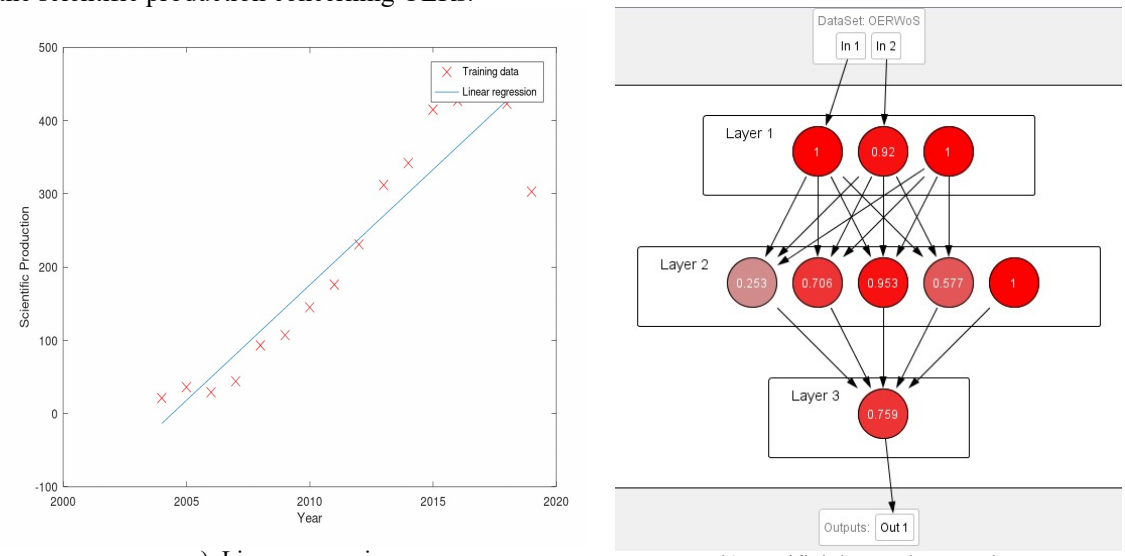
IV. TRENDS IN OPEN EDUCATIONAL RESOURCES

Predictive modeling is defined as data analysis for making decisions or solving problems [13]. The scope of this work is empirically gathered datasets to be processed not only with statistical methods like these used from VOSviewer and Biblioshiny, but also machine learning algorithms to be applied to predict future events. Thus, the first predictive model is created to show the tendency at generated scientific production in the OERs domain and it points out the increased interest from the scientific society (Figure 8a). The used dataset contains bibliographic data extracted from SCOPUS database and Octave software platform [14] is utilized for its building. The applied machine learning method is linear regression that is supervised algorithm and it is described through following formulas [15]: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where x is variable predictor, y is variable response, β_1 shows the line

slope, β_0 is the line intercept, and r is the correlation coefficient,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^{\,2} - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = r\frac{s_y}{s_x};\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x};\ r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right).$$

The second predictive model is constructed with the dataset extracted from Web of Science and the data are normalized to fit the interval [0, 1]. The applied method of supervised machine learning is based on artificial neural networks (ANN) used in the context of number series prediction [16]. Figure 8b presents the created model through Neuroph Studio[17] in the form of multi layer perceptron that learns through backpropagation learning rule. The constructed ANN is optimized to work with two inputs, one output and four hidden neurons. The trained data presents 70% of the dataset. The learning rate is 0.1, the momentum is 0.5 and the obtained total mean square error is 0.0170075871045588. The applied activation function is chosen to be sigmoid. The model can predict the next values in a number sequence and in this case the predicted values show the future number of the scientific production concerning OERs.



a). Linear regression b). Artificial neural network

Figure 8. Predictive models in the domain OERs

V. CONCLUSIONS

The paper presents an approach for learning analytics application for the purposes of OERs domain analysis. It is based on datasets extraction from abstract and citation databases SCOPUS, Web of Science and Dimensions, bibliographic data processing and visualization of constructed bibliometric networks with VOSviewer and Biblioshiny. The findings include extensive information about the context of OERs usage, trending topics, the scientific production, the most relevant sources – journals and conferences, the countries of actively involved authors, as well as the collaboration among them.

The proposed approach in this paper is also used for development of predictive models based on supervised machine learning methods like linear regression and artificial neural networks. They are capable to predict the future evolvement of scientific production in the area of OERs, as well as any other topics in this field.

This study about the domain of OERs could be used by researchers and educators for understanding the current research and as an opportunity to be involved in the recently discussed

topics. It could be applied as a starting point for further research of a given issue, stepping on the outlined in this work "big picture" and on the created prognosis. The researchers could take into account the proposed research approach that includes statistical and machine learning techniques and that describes conceptually the OERs domain. The educators could take this overview picture to propose novelty and more effective solutions for implementation of OERs in educational process.

In the context of the critical situation brought by the COVID-19 outbreak worldwide, when a large number of countries suspended their courses in schools and universities, moving to online learning, the OERs, MOOCs and Open Educational Practices (OEPs) proved their usefulness, being adopted (or continuing previous experiences) by many educational institutions and teachers. They bring innovation, flexibility, openness and collaboration in international communities of learning and practice, as revealed by the exchange of best practices in the international webinar "How to Keep Students Learning during Schools Disruption in COVID-19 Situation", organized by the Smart Learning Institute of Beijing Normal University in March 13 [18].

Teachers should become curators and facilitators of quality OERs and MOOCs, integrating them in their courses.

FutureLearn (https://www.futurelearn.com), the most important European MOOC platform, support teachers with two new sections "How to Teach Online Courses" and "Disease Outbreak Prevention Courses".

Searching on the classcentral.com directory or directly on the big MOOC platforms edx.org, coursera.org and udacity.com, teachers could find MOOCs to improve their knowledge in their academic and research areas, and skills as Open Educators; also they could guide their students to follow different (parts) of such massive courses, supporting them to become autonomous Open Learners, participating in distributed learning communities.

In conclusion, it can be said that the open content and all connected to it issues will continue to be in the focus of discussions, research and novel implementations. The authors and educators have a huge volume with useful resources, best practices and theoretical frameworks that could be used as bases for further exploration, analysis and inventions.

The future work will be focused on detailed exploration the domain of OERs for the both countries: Bulgaria and Romania using the method of bibliometric networks construction for analysis and machine learning techniques for creation of predictive models.

Reference Text and Citations

[1] Bodily, Robert;Nyland, Rob;Wiley, David, 2017. The RISE Framework: Using Learning Analytics to Automatically Identify Open Educational Resources for Continuous Improvement, *International Review of Research in Open and Distributed Learning*, 18(2), 103-122.

[2] Piedra, Nelson;Caro, Edmundo Tovar, 2014. *Serendipity: a platform to discover and visualize data from OER*, OCWC Global Conference, 23-25 April 2014, Ljubljana, Slovenia, https://www.researchgate.net/publication/280113398_Serendipity_a_platform_to_discover_and_visualize_data_from_OER.

[3] Avila, Cecilia;Baldiris, Silvia;Fabregat, Ramón;Graf, Sabine, 2017. ATCE: an analytics tool to trace the creation and evaluation of inclusive and accessible open educational resources, *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*,March 2017, 183–187, https://doi.org/10.1145/3027385.3027413.

[4] Illinois University Library,https://guides.library.illinois.edu/oer.

[5] University of Washington Libraries, https://guides.lib.uw.edu/oer.

[6] The University of Queensland Library,https://guides.library.uq.edu.au/how-to-find/open-educational-resources.

[7] The William and Flora Hewlett Foundation, 2018. *How can technology advance open educational resources?,*https://hewlett.org/how-can-technology-advance-open-educational-resources/.

[8] Barker, Phil;Campbell, Lorna M., Technology Strategies for Open Educational Resource Dissemination, Chapter 3, *Open Education International Perspectives in Higher Education*, Patrick Blessinger and TJ Bliss (ed.), Open Book Publishers, 51-71.

[9] Perianes-Rodriguez, A.; Waltman, L.; Van Eck, N.J., 2016. Constructing bibliometric networks: A comparison between full and fractional counting,*Journal of Informetrics*, 10(4), 1178-1195.

[10] Aria, M.;Cuccurullo, C., 2017. Bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*, 11(4), 959-975.

[11] Jan van Eck, Nees;Waltman, Ludo, 2017. *VOSviewer Manual*, https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.6.pdf.

[12] Grosseck, G.; Țîru, L.G.; Bran, R.A. 2019. Education for Sustainable Development: Evolution and Perspectives: A Bibliometric Review of Research, 1992–2018.*Sustainability*2019,*11*, 6136.

[13] Kalechofsky, Hal,*A Simple Framework for Building Predictive Models*, http://www.msquared.com/wp-content/uploads/2017/01/A-Simple-Framework-for-Building-Predictive-Models.pdf.
[14] GNU Octave, https://www.gnu.org/software/octave/.
[15] Sridharan, Ramesh, Linear Regression, Chapter 3, *Linear regression,* http://www.mit.edu/~6.s085/notes/lecture3.pdf.
[16] Velasco,Lemuel Clark P.;Granados, Angelie Rose B.;Ortega,Jilly Mae A.;Pagtalunan, Kyla Veronica D. Medium-Term Water Consumption Forecasting using Artificial Neural Networks, https://www.researchgate.net/publication/318529760_Medium-Term_Water_Consumption_Forecasting_using_Artificial_Neural_Networks.
[17] Neuroph Java Neural Network Framework,http://neuroph.sourceforge.net.
[18] Grosseck, G.; Holotescu, C. 2020. Romanian Educational System adapting to the coronavirus outbreak. Presentation for the International Webinar "How to Keep Students Learning during Schools Disruption in COVID-19 Situation", https://www.slideshare.net/cami13/romanian-educational-system-adapting-to-the-coronavirus-outbreak.