THE ENCORE PROJECT: SUPPORTING EDUCATIONAL DESIGN WITH NATURAL LANGUAGE PROCESSING

Filippo Chiarello

School of Engineering, Department of Energy, Systems, Land and Construction Engineering, University of Pisa, Lab – Business Engineering for Data Science

Massimo Panarotto, Johan Malmqvist

Chalmers University of Technology Dept. of Industrial and Materials Science

ABSTRACT

ENCORE (ENriching Circular use of OER for Education) is an Erasmus+ project. This paper introduces the ENCORE approach, a novel strategy developed to address the challenges of the digital era in engineering education. ENCORE harnesses the potential of OER combined with the advancements in Natural Language Processing (NLP) and artificial intelligence, linking educational content with relevant skills in the European Skills, Competencies, Qualifications, and Occupations (ESCO) framework. The approach is designed to enhance the reuse of OERs, integrating them with data-driven tools and pedagogical guidelines to facilitate effective course design and outcome evaluation. This paper particularly explores the integration of the ENCORE approach within the CDIO (Conceive-Design-Implement-Operate) framework, highlighting how it can assist educators in incorporating green digital and entrepreneurial skills into engineering courses. The core elements of ENCORE, including a specialized database and ENCORE enablers, are outlined, emphasizing their role in crafting relevant learning paths and enriching engineering education. The paper concludes with a forward-looking perspective on the future development of ENCORE, focusing on its testing and refinement in engineering education through upcoming staff development activities and pilot testing, underscoring the importance of its alignment with the CDIO framework in enhancing the learning experience in the rapidly evolving landscape of engineering education.

KEYWORDS

Open Educational Resources, Natural Language Processing, Engineering Education, Digital Skills Development, Sustainable Educational Practices, Standards: 2, 6, 9

INTRODUCTION

In the contemporary digital era, educational methodologies and institutional models are experiencing profound transformations. The shift from traditional classroom learning to digital platforms enables learning anytime, anywhere (Kaplan, 2022). This evolution demands new competencies in anticipation, adaptability, and decision-making under uncertain conditions. However, the educational sector (also engineering education) often lags in adapting to these social and technological shifts.

The plethora of information and emerging technologies necessitates pedagogical strategies grounded in robust scientific and technical knowledge. The pressing challenges of the 21st century, such as climate change and digitalization, underscore the need for updated teaching and learning methods. The COVID-19 pandemic has further accentuated this urgency, presenting a crossroads for educational reform (Williamson, Eynon, & Potter, 2020). A paradigm shift from mere knowledge dissemination to flexible, proactive methods employing automation and Artificial Intelligence (AI) is essential. This requires educators to rethink their practices and engage actively in designing and debating educational interventions that foster technologically advanced, sustainable, and resilient societies (Goodyear, 2015).

In the European context, policies and initiatives aim at digital transformation while ensuring inclusive and sustainable growth. The ”A strategy for smart, sustainable and inclusive growth” policy underscores the balance between technological progress, sustainability, and social inclusion. The Digital Education Action Plan (2021-2027) and the updated Digital Competence Framework reflect the EU’s commitment to adapting education to the digital age, particularly in the wake of the pandemic and climate change (Commission, 2020). The European Skills Agenda, part of the European Digital Strategy, focuses on skill development and utilization in response to digital transformation and the transition to a climate-neutral Europe.

The abundance of open educational resources (OERs) presents significant opportunities for education, allowing educators to focus on value-added activities (Wiley, Bliss, & McEwen, 2014). However, this shift calls for a transition from open educational resources to open educational practices. The EU supports this through various initiatives, including the Joint Research Centre’s ”Opening up Education: A Support Framework for Higher Education Institutions” and the Open Education Europa platform (dos Santos, Punie, Muñoz, et al., 2016). However, the global movement towards OER usage, especially during the COVID-19 emergency remote teaching, has highlighted challenges in OER quality and usage (dos Santos et al., 2016).

Intelligent systems are increasingly developing educational tools to manage the abundance of resources. This evolution necessitates new approaches to skills recognition and certification, highlighting the potential of micro-credentials and open digital badges (Segarra-Faggioni & Romero-Pelaez, 2022b). These innovations, however, require further empirical research and expert insights for effective policymaking and user engagement.

Addressing these challenges, this paper introduces the ENCORE (ENriching Circular use of OER for Education) approach, a novel method to confront the digital era’s educational challenges developed in the realm of the ENCORE project (https://project-encore.eu/). ENCORE leverages OER knowledge, applying Natural Language Processing (NLP) to link information with relevant skills in the European Skills, Competencies, Qualifications, and Occupations (ESCO) framework. It combines data-driven tools with pedagogical guidelines to

enhance the reuse of OERs. The paper also explores the integration of the ENCORE approach with CDIO to help educators to incorporate new content in engineering courses.

THE CORE ELEMENTS OF ENCORE: NATURAL LANGUAGE PROCESSING AND OPEN EDUCATIONAL RESOURCES

This section synthesizes existing research on Natural Language Processing (NLP) applications in Open Educational Resources (OER) and topic extraction from education and labor market documents. Key insights are drawn from the literature, notably the works of Ferreira et al. Ferreira-Mello, André, Pinheiro, Costa, and Romero (2019), highlighting the state-of-the-art in NLP applied to educational resources.

OERs, encompassing a variety of educational materials like textbooks, videos, and lesson plans, are crucial for broad educational access. However, their effectiveness is often hindered by poor quality metadata, impeding resource discoverability (Tavakoli, Faraji, Mol, & Kismihók, 2020). This challenge is exacerbated as OER authors frequently overlook metadata creation due to its time-consuming nature. AI, particularly through NLP techniques, emerges as a solution for enhancing OER metadata by extracting key information from their content (SegarraFaggioni & Romero-Pelaez, 2022a).

NLP, a subset of AI, interprets human language in written form, and is pivotal in converting OERs into analyzable text (Tavakoli et al., 2020). A common NLP methodology in this context is topic modeling, with Latent Dirichlet Allocation (LDA) being widely used for identifying OER topics (Tavakoli et al., 2022). Nevertheless, LDA's unsupervised nature leads to varying topic detail levels (Puccetti, Giordano, Spada, Chiarello, & Fantoni, 2023). Addressing this, a balanced approach in topic identification is essential, avoiding overly abstract or specific terminologies to enhance educational resource recommendation systems.

Furthermore, defining 'education topics' is necessary to overcome LDA limitations (Nadeau & Sekine, 2007). Named Entity Recognition (NER) serves this purpose, classifying educationrelated topics in text. NER employs three main methods: gazetteer-based, rule-based, and machine learning-based (Giordano, Chiarello, Melluso, Fantoni, & Bonaccorsi, 2021; Giordano, Coli, & Martini, 2022).

Advancements in word embedding techniques have significantly impacted NLP. These techniques, including traditional (word2vec, GloVe) and contextual embeddings (BERT), have improved entity recognition efficiency (Mikolov, Chen, Corrado, & Dean, 2013). Parallelly, NLP's role in labor market analysis is expanding, with applications in skill and knowledge extraction, relevant to both labor and education sectors (Spada et al., 2022).

Despite the limited application of NER in OER, the increasing integration of NLP to enhance OER quality and accessibility shows promise Gazzola et al. (2022). This integration aligns with the CDIO (Conceive-Design-Implement-Operate) framework in engineering education, where enhancing educational resources through AI and NLP can significantly contribute to the 'Implement' and 'Operate' stages. These stages focus on realizing and sustaining the educational approach, where AI-assisted metadata improvement and resource discoverability play a crucial role.

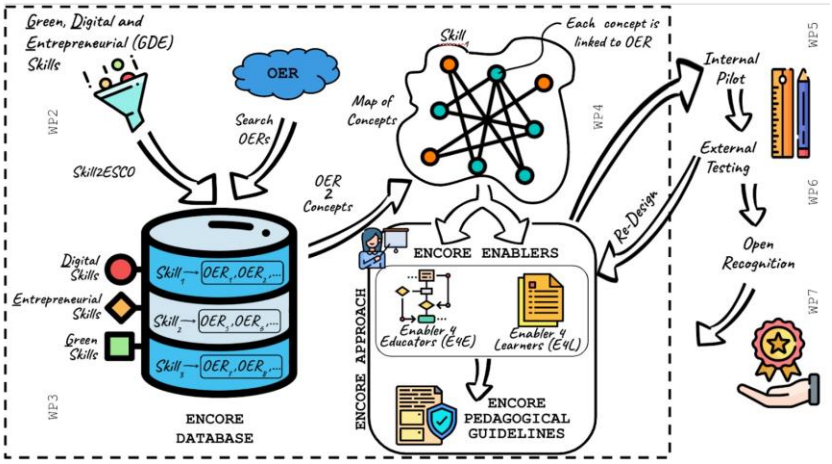

Figure 1. Diagrammatic representation of the ENCORE ecosystem, illustrating its components and the flow of information (from the ENCORE website)

THE ENCORE METHOD

In light of the aforementioned context, this section delineates the conceptual framework and methodological underpinnings of our proposed solution, ENCORE. The primary aim of ENCORE is to enhance teaching methodologies and facilitate the recognition of skills pivotal to addressing contemporary macro-trends such as digitalization, environmental change, and the economic resurgence post-COVID. Our objectives are multifaceted: (1) Equip students and professionals with the competencies necessary to navigate economic, ecological, and technological shifts; (2) Address the skill gap between industry requirements and academic offerings; (3) Assist educators in keeping pace with the dynamic skill landscape; (4) Address the inertia in educational sector reforms; (5) Aid educators in adapting to evolving educational contexts with effective pedagogical strategies; (6) Assist students in authenticating skills acquired through unconventional pedagogical methods. Figure 1 synthetise the ENCORE method in graphical form.

Our strategy is to augment the teaching and recognition of skills impacted by these macrotrends, utilizing the wealth of information embedded in Open Educational Resources (OER)[2]. This will involve extracting knowledge from OER texts using Natural Language Processing (NLP), a field at the intersection of computer science, linguistics, and AI, which enables computers to comprehend and analyze human language. This methodology allows for the extraction of valuable insights from OER. Within this framework, our approach integrates data-driven tools for instructional design with pedagogical guidelines, steering educators towards effectively crafting courses that achieve desired learning outcomes and evaluate those outcomes . Our focus is on learning outcomes associated with skills necessary to confront challenges in digitalization, environmental change, and post-COVID economic recovery. This is further supported by leveraging another textual knowledge source, the ESCO (European framework of Skills, Competences, Qualifications and Occupations). Through NLP, OER content is scrutinized, and skills are identified and aligned with those listed in ESCO, leading to the creation of a specialized database categorizing OER and corresponding skills.

[2] https://www.oercommons.org/

Building on this knowledge base, we will develop the ENCORE methodology, which amalgamates data-driven tools and approaches for educational design with pedagogical principles. This will empower educators to effectively utilize ENCORE's technological resources in crafting courses that not only reach but also assess intended learning outcomes using open resources. A key ambition of the ENCORE strategy is to harness and repurpose OER, as well as to devise novel tools for both teaching and learning. Furthermore, the initiative advocates for an open methodology in recognizing learning outcomes, aligning with the European Union's ongoing micro-credentialing processes.

The ENCORE project adopts a technological ecosystem perspective, characterized by independent yet cohesive components. Central to this ecosystem is a database linking Open Educational Resources (OERs) to General Digitalization and Entrepreneurship (GDE) skills, alongside ENCORE enablers crafting relevant learning paths. This section outlines the ecosystem's key elements: the database and enablers, their architecture, and the pedagogical evaluation process for validation.

Database

The ENCORE database forms the technological ecosystem's core, cataloging OERs and their association with GDE skills. Its design emphasizes robustness and adaptability, accommodating evolving OER and skill characterizations. This flexibility ensures minimal impact on the ecosystem during data schema modifications. Initially, the database will utilize the Dublin Core Metadata Element Set (DCMES), an open-standard metadata schema aligned with the database's goals of simplicity, understandability, international scope, and extensibility Kunze and Baker (2007); Weibel, Kunze, Lagoze, and Wolf (1998).

The database's development includes populating it with high-quality, relevant OERs. This involves assessing existing OER repositories, selecting pertinent resources, and extracting metadata and content using APIs or web scraping. Post-extraction, a quality filter vets each OER, followed by NLP techniques to identify and link relevant GDE skills.

Inter-component communication within the ecosystem is crucial. The ENCORE database features a REST API, facilitating metadata and content retrieval by other tools, notably the ENCORE enablers.

ENCORE Enablers

The ENCORE enablers are designed to assist educators in creating effective learning paths. The Educator-for-Educator (E4E) Enabler aids in constructing these paths, which guide students towards intended learning outcomes (ILOs). This tool displays a proto-conceptual map of course concepts extracted from OERs, helping educators select concepts for inclusion and define progression criteria between them. Gamification techniques enhance the appeal and effectiveness of these paths. The E4E Enabler, an open-source software, incorporates gamification mechanics to motivate students, making learning paths both challenging and engaging.

For students, the defined learning path is transformed into an executable format through the Educator-for-Learner (E4L) Enabler. Utilizing Microsoft's open technologies and notebook software, this tool facilitates interactive learning and assessment. Notebooks combine fixed and editable text cells, allowing students to engage with materials and assessments designed by educators. The learning sequence in the notebook, derived from the E4E output, is further

enriched with gamification elements to boost student engagement. The E4L Enabler will also be released as open-source software, promoting interactive and engaging learning experiences.

POSSIBLE INTEGRATION OF ENCORE AND CDIO APPROACH

The ENCORE approach, integrated within the CDIO framework, provides a novel pathway for educators to instill green digital and entrepreneurial skills in engineering education. This subsection delves into how ENCORE's database of Open Educational Resources (OER), focused on these key skill areas, aligns with the CDIO syllabus' phases - Conceive, Design, Implement, Operate - enriching the curriculum content with practical, sustainable, and innovative perspectives.

Table 1 highlights specific contributions, advantages for educators and students, and potential challenges.

Conceive Phase

In the 'Conceive' phase, where students learn to conceptualize new systems and products, ENCORE plays a crucial role by offering OERs centered around green digital technologies and entrepreneurial thinking. This integration allows students to explore sustainable and innovative approaches from the conception stage of engineering projects. For educators, ENCORE provides a diverse range of resources to demonstrate the importance of sustainability and innovation in the early stages of engineering design.

Design Phase

During the 'Design' phase, students engage in creating detailed plans for engineering systems. ENCORE enriches this phase by providing access to OERs that highlight green design principles and digital innovation strategies. Utilizing NLP, ENCORE can selectively present design concepts that emphasize eco-friendly practices and digital technologies, thereby fostering a mindset geared towards sustainable and technologically advanced engineering solutions.

CDIO Phase	Integration of ENCORE	Benefits	Potential Challenges
Conceive	Providing OERs on green digital technologies and entrepreneurial thinking	Enriches conceptual learning with sustainability and innovation	Aligning resources with specific project concepts
Design	Presenting green design principles and digital innovation strategies	Encourages eco-friendly and technologically advanced solutions	Ensuring the relevance and accuracy of content

Implement	Offering case studies on implementing green technologies and digital tools	Bridges theory and practice in sustainable engineering	Tailoring examples to diverse student projects
Operate	Access to resources on sustainable operation and entrepreneurial management	Prepares students for operational challenges with a sustainability focus	Balancing operational, sustainability, and business aspects

Table 1. Integration of ENCORE in CDIO Phases with a Focus on Green Digital and Entrepreneurial Skills

Implement Phase

In the 'Implement' phase, the focus shifts to the practical implementation of designed systems. Here, ENCORE's repository offers practical examples and case studies demonstrating the application of green technologies and digital tools in real-world engineering projects. This phase benefits significantly as students get to understand the challenges and methodologies involved in implementing sustainable and technologically driven designs.

Operate Phase

The 'Operate' phase revolves around the operation and maintenance of engineering systems. ENCORE's role in this phase is pivotal in providing resources that discuss sustainable operation practices, digital optimization of systems, and entrepreneurial management strategies. These resources prepare students to manage and operate engineering projects with a focus on longterm sustainability, digital efficiency, and innovative business models.

CONCLUSIONS

The integration of the ENCORE approach within the CDIO framework marks a pivotal advancement in engineering education, addressing the pressing challenges of digitalization, environmental change, and economic shifts. This paper has detailed how ENCORE, leveraging OER and NLP, enriches each CDIO phase with green digital and entrepreneurial skills, enhancing the educational landscape for engineering students.

Looking ahead, the next phase of ENCORE's development is crucial. It involves the recruitment and training of higher education lecturers and vocational trainers in staff development activities centred around the ENCORE approach. This phase aims to enrol a diverse group of educators to test and maximize the application of ENCORE across various skill sets and course topics.

The focus will then shift to external development activities where these educators will be equipped to implement ENCORE in their teaching practices. The feedback from these activities is instrumental in refining the ENCORE methodology. Following this, a pilot phase will see the

ENCORE approach applied in real-world educational settings, providing a valuable opportunity to evaluate its effectiveness and impact.

The significance of testing ENCORE in engineering education, mainly through the lens of the CDIO framework, cannot be overstated. This setting offers an ideal testbed for assessing the adaptability and efficacy of the ENCORE approach. It promises to not only enhance the learning of critical skills but also provide insights into its scalability across educational contexts.

In summary, the future trajectory of ENCORE is set to make a substantial impact on engineering education. Through its upcoming phases, it will undergo rigorous testing and refinement, ensuring its role as a transformative tool in equipping future engineers with essential skills in an ever-evolving global landscape.

FINANCIAL SUPPORT ACKNOWLEDGEMENTS

This research has been partly supported by ENCORE Project (ENriching Circular use of OeR for Education) – European Union Erasmus+ Partnerships for Innovation: Alliances (ERASMUSEDU-2021-PI-ALL-INNO) – Project Number: 101055893.

REFERENCES

Commission, E. (2020). *Europe 2020: A Strategy for Smart, Sustainable and Inclusive Growth.* https://faolex.fao.org/docs/pdf/eur202078.pdf. (Accessed: 2023-04-15)

dos Santos, A. I., Punie, Y., Muñoz, J. C., et al. (2016). *Opening up education: A support framework for higher education institutions* (Tech. Rep.). Joint Research Centre (Seville site).

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., & Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(6), e1332.

Gazzola, M., Leal, S., Pedroni, B., Theoto Rocha, F., Pompéia, S., & Aluísio, S. (2022). Text complexity of open educational resources in portuguese: mixing written and spoken registers in a multi-task approach. *Language Resources and Evaluation*, 1–30.

Giordano, V., Chiarello, F., Melluso, N., Fantoni, G., & Bonaccorsi, A. (2021). Text and dynamic network analysis for measuring technological convergence: A case study on defense patent data. *IEEE Transactions on Engineering Management*.

Giordano, V., Coli, E., & Martini, A. (2022). An open data repository for engineering design: Using text mining with open government data. *Computers in Industry*, *142*, 103738.

Goodyear, P. (2015). Teaching as design. *Herdsa review of higher education*, *2*(2), 27–50.

Kaplan, A. (2022). *Digital transformation and disruption of higher education*. Cambridge University Press.

Kunze, J., & Baker, T. (2007). *The dublin core metadata element set* (Tech. Rep.).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, *30*(1), 3–26.

Puccetti, G., Giordano, V., Spada, I., Chiarello, F., & Fantoni, G. (2023). Technology identification from patent texts: A novel named entity recognition method. *Technological Forecasting and Social Change*, *186*, 122160.

Segarra-Faggioni, V., & Romero-Pelaez, A. (2022a). Automatic classification of oer for metadata quality assessment. In *2022 international conference on advanced learning technologies (icalt)* (pp. 16–18).

Segarra-Faggioni, V., & Romero-Pelaez, A. (2022b, July). Automatic classification of OER for metadata quality assessment. In *2022 International Conference on Advanced Learning Technologies (ICALT)* (pp. 16–18). (ISSN: 2161-377X) doi: 10.1109/ICALT55010.2022.00011

Spada, I., Chiarello, F., Barandoni, S., Ruggi, G., Martini, A., & Fantoni, G. (2022). Are universities ready to deliver digital skills and competences? a text mining-based case study of marketing courses in italy. *Technological Forecasting and Social Change*, *182*, 121869.

Tavakoli, M., Faraji, A., Mol, S. T., & Kismihók, G. (2020). Oer recommendations to support career development. In *2020 ieee frontiers in education conference (fie)* (pp. 1–5).

Tavakoli, M., Faraji, A., Vrolijk, J., Molavi, M., Mol, S. T., & Kismihók, G. (2022). An ai-based open recommender system for personalized labor market driven education. *Advanced Engineering Informatics*, *52*, 101508.

Weibel, S., Kunze, J., Lagoze, C., & Wolf, M. (1998). *Dublin core metadata for resource discovery* (Tech. Rep.).

Wiley, D., Bliss, T., & McEwen, M. (2014). Open educational resources: A review of the literature. *Handbook of research on educational communications and technology*, 781–789.

Williamson, B., Eynon, R., & Potter, J. (2020). *Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency* (Vol. 45) (No. 2). Taylor & Francis.

BIOGRAPHICAL INFORMATION

Filippo Chiarello, PhD**,** is currently an Assistant Professor at the School of Engineering, University of Pisa. He teaches Design and Innovation at the master's degrees of Engineering Management and Data Science. His research focuses on natural language processing techniques for studying technological innovation and HR-related phenomena. He is responsible for faculty development for his institution and is the coordinator of the European Project ENCORE and DETAILLs.

Massimo Panarotto, PhD, is currently an Associate Professor with the Department of Industrial and Materials Science, Chalmers University of Technology, Gothenburg, Sweden. He received the Ph.D. degree in mechanical engineering from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2015. He is the author of more than 50 publications in the field of systems engineering and engineering design. His main research focus is on developing decision supports for determining the lifecycle value of a new engineering system, whose applications are in product development within the space, aerospace, and automotive industry. Dr. Panarotto is a Member of the Design Society.

Johan Malmqvist, PhD**,** is a full professor in product development. His research addresses methodologies for data-driven product development. Current research focuses on methods and tools for design for additive manufacturing, for the application of machine learning in design, and for mass customization of pharmaceutical products. Another area of interest is engineering education. Malmqvist was one of the co-founders and is still active in the international Conceive-Design-Implement-Operate (CDIO) Initiative. The engineering education model that has been developed by the CDIO Initiative has been adopted by a large number of across the world.

Corresponding author

Filippo Chiarello
University of Pisa
School of Engineering
L.go Lucio Lazzarino,1
56122 Pisa PI, Italy
filippo.chiarello@unipi.it