European Commission

Results of the second phase of the pilot project for linking learning outcomes of qualifications with ESCO skills and next steps

MSWG 14-03

European Skills, Competences, Qualifications and Occupations

Introduction

The present document aims at reporting to members of the ESCO MSWG and the EQG AG on the results of the second phase of the pilot project for testing the use of an automated approach for linking learning outcomes of qualifications to ESCO skills and to discuss the next steps with both groups, meeting on 10 June (ESCO MSWG) and 15/16 June (EQF AG).

The document recalls the policy objectives behind the use of an automated approach for linking qualifications to ESCO skills and proposes a way forward for a potential third phase of the project.

Members are invited to provide their input on the content of the note, to comment on the results of the pilot, and to endorse the further development of the linking tool with the aim of becoming permanently available.

1. Background

For ESCO to reach its full potential as a bridge between education and training and the labour market, the ESCO skills terminology could be used to systematically identify and analyse which skills are related to a particular qualification.

This can help make qualifications more transparent across Europe. It can also help the translation, comparison and/or review of qualifications. Furthermore, his identification and analysis of the content and profile of a qualification can be used to indicate matches and/or mismatches with skills needs of occupations and sectors. Linking learning outcomes of qualifications to ESCO skills means that employers can more easily grasp the labour market value of a qualification, in particular in a cross border context. Individuals may see their chances on the labour market improved through better matching based on richer qualifications information.

In February 2020, the Commission presented to the ESCO Member States Working Group (MSWG) and to the EQF Advisory Group the initial results of the pilot project for linking learning outcomes of qualifications to ESCO skills.

The project allowed assessing the required effort for creating links between learning outcomes of qualifications and ESCO skills. To support the linking process, the Commission developed an IT tool using artificial intelligence, in particular **Natural Language Processing (NLP)**, and based on requirements expressed by the participating countries. As a result, the exercise confirmed the methodology of an automated approach with an initial degree of human intervention, as suggested by the 2019 ESCO Qualifications Pillar study[1].

In **April 2020**, the Commission invited the ESCO MSWG and the EQF AG to express their interest in participating in the second phase of the project. The call was also open to organisations testing the usage of ESCO.

The second phase of the pilot had the following objectives:

[1] https://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=8181&furtherPubs=yes

The study was presented during the joint meeting of the EQF AG and the ESCO MSWG on 6 February 2019.

- To improve the technology supporting the automated linking of learning outcomes of qualifications with ESCO skills, through an **enhanced matching algorithm based on artificial intelligence**.
- To test the automated linking and the performance of the matching algorithm in **different EU languages**.
- To test the usability of the hierarchical structure of ESCO skills and introduce rules and matching relations between skills and learning outcomes of qualifications, **allowing users to define exact matches between skills and learning outcomes or identify broader/narrower relations.**
- To make **suggestions on the ESCO skills pillar** that will feed into the ESCO continuous improvement process.

To this end, participants were invited to test a **minimum of 25 qualifications** covering different EQF levels and including at least some vocational and some academic qualifications, and to **share the linking data** with the Commission for further improving the matching algorithm.

Overall, **13 Member States and organisations** participated in the second phase of the pilot as full members or observers:

- 9 EU Member States: Belgium, Czech Republic, France, Italy, Latvia, the Netherlands, Poland, Romania and Slovenia.
- 2 stakeholder organizations: ETUC and EURASHE.
- 1 EU agency: The European Training Foundation (ETF).
- 1 international organisation: The Inter-American Development Bank.

2. Project implementation

The second phase of the pilot started in September 2020 and lasted until January 2021.

A first workshop was organised on 22 June 2020 to present the objectives and main features of the pilot and the improved version of the IT tool to link learning outcomes of qualifications with ESCO skills. During the workshop, participating countries described their interest in the project and presented the qualifications that would be linked with ESCO skills.

The pilot involved **367 qualifications**: 47 from BE-fr, 25 from BE-nl, 24 from FR, 32 from IT, 24 from LV, 19 from the NL, 26 from PL, 25 from RO, 39 from SL, 50 from 5 ACQF countries via the ETF (Cameroon, Cape Verde, Kenya, Mozambique and South Africa) and 42 from the IDB.

The IT tool and its user manual were presented to the participants in a webinar on 6 November 2020, during which the Commission replied to the written questions submitted by the participants during the first two months of the project.

The results of the linking exercise were discussed in a final workshop on 8 February 2021. The workshop's agenda included individual national presentations on the project results, a discussion on common issues identified by participating countries, feedback on the functioning of the IT tool and on the further improvements to be prioritised for future development. To this end, prior to the meeting the Commission circulated a list of questions and points of interest on the perceived quality of the ESCO vocabulary and on the usage of the IT tool.

3. Features of the IT tool supporting automated linking of learning outcomes of qualifications to ESCO skills

The IT tool included the following functionalities:

- A **text-splitting mechanism** to automatically define learning outcomes from the full text of a qualification, offering the possibility to customise the text splitting before initiating the linking activity.
- Suggestions of concepts that may be relevant to a selected Learning Outcome entity based on the **machine learning algorithm**.
- Suggestions of concepts that may be relevant to a selected Learning Outcome coming from the **ESCO API**.
- The option to manually select relevant ESCO skills by browsing **the ESCO skills hierarchy**.
- The option to browse **the ESCO occupation pillar** to select skills linked to occupations that may be relevant to a selected Learning Outcome.

4. Main findings and results from the second phase of the project

The linking exercise covered 367 qualifications in 9 languages, with an average percentage of **mapped text of 44%** (text that was linked with at least one skill or knowledge concept). This number varies across participants based on the number of qualifications effectively mapped: participants who completed (or almost completed) the linking exercise showed a percentage of mapped text higher than 90%, while participants who mapped only few qualifications had the lowest score.

In total **3724 learning outcomes** were mapped to at least one ESCO concept, generating **11998 links**. The number of established links per learning outcome varies across linking projects: based on the data obtained, it can be expected that **1 to 3 ESCO skills or knowledge concepts are mapped to each learning outcome entity**.

In terms of mapping relations[2], data showed a prevalence of the relation type "close match", suggesting that the selected ESCO concept does not fully cover the scope of a learning outcome. The average percentage of exact matches was 20%, however this number varies across projects (ranging from around 60% to 0%). These statistics should be interpreted, taking into account three aspects:

- "close match" is the default relation type in the linking tool (which implies a link will be automatically defined as close match if no further action is taken by the reviewer);
- the linking tool did not provide for the option of linking to broader levels of the skills hierarchy (skill groups); and

[2] The tool allowed four types of mapping relations: "exact match", "narrower than", "broader than" and "close match".

- the interpretation of relation types might have been different between participants.

The performance of the mapping algorithm showed promising results. On average, an ESCO concept selected by participants was in **30% of cases the first concept** suggested by the algorithm, **among the first 3 concepts in 48% of cases**, among the **top 5 concepts in 56% of cases** and among the **top 10 concepts in 66% of cases**.

Based on these analytics and on the comments provided by the participants during the final workshop, the following conclusions can be drawn:

- Linking learning outcomes to skills is done best in the **native language** as this allows the algorithm to learn more about the specifics of a language, thus producing better matches.
- The **structure of learning outcomes** plays a very important role for the correct functioning of the algorithm: on average, 10-15 words are found in a learning outcome, which contributes to returning a high number of concepts, sometimes not (or partially) related to the meaning of the text.
- **Splitting sentences** in a correct way is very important for the functioning of the tool. In this regard, the variety of approaches used in the description of learning outcomes (numbering, punctuation, flat text) represents a challenge.
- It is important to further specify the definitions of the different types of matching relations given the **relevance of exact matches** for training the AI algorithm.
- The current dataset can already be used for the evaluation/comparison of different candidate algorithms. However, **more data are needed to refine the model** and obtain results that are more accurate.
- Sometimes relevant information is contained in the **structure or in the description of a qualification** (e.g. unit title). These data would be useful to support the understanding of learning outcomes descriptions by the algorithm.
- **Human intervention is an important component** of this exercise and a review of the results provided by the machine learning algorithm is needed at this stage of development of the technology. This has an impact on the effort needed to link learning outcomes of qualifications and ESCO skills. **Future improvements to the AI model will reduce the time needed to perform this exercise**.
- Learning outcomes descriptions are generally broader than ESCO skills and knowledge concepts, leading to links with multiple skills. A vector-approach to linking could be used to indicate those concepts that, if combined, have sufficient similarity to a learning outcome description.
- The ESCO skills taxonomy is very granular and a certain level of abstraction should be introduced in order to link qualifications from higher levels (EQF 5-7). Overall, **the linking provided better results for VET qualifications compared to qualifications from higher education**.
- The **ESCO skills hierarchy supports** the search of relevant skills within the ESCO dataset.

5. Perspectives of AI usage and further improvements of the matching algorithm

To continuously improve natural language processing algorithms, large amounts of data are needed. Especially verified or validated data, in this case the links between learning outcomes and skills, is crucial. This validated data allows the Commission to define a test data set to compare different (versions of) algorithms and quantify their relative performance in an objective way.

During this second phase of the pilot, the basic search algorithm as it is available in the ESCO API and a first standard natural language processing algorithm were available. Thanks to the work of the participants, it was possible to collect sufficient data to build a test data set to develop new and improved versions of the algorithm and objectively measure progress. This will be the focus of a third phase of the pilot.

New versions of the algorithm have been developed and will be integrated in a next phase of the pilot. The participants will get results from all the integrated algorithms and based on the selections that the participants will make, it will be possible to objectively compare the performance of the different algorithms.

This comparative data will lead to a choice of the best performing algorithm available at this time, which will then be used in the final standalone service.

6. Next steps

Based on the results described above, the Commission believes that the ESCO skills taxonomy is a valuable tool to identify which skills are linked to a particular qualification and support the overall goal of promoting transparency of information on qualifications.

At the same time, the Commission believes that further investing in using AI to link learning outcomes of qualifications to ESCO skills is a precondition for developing an AI infrastructure that can promote skills-based job matching, reduce skills mismatches, support re- and up-skilling based on the individual's skillset and bridge the terminological distance between the world of education and training and the labour market.

The next release of ESCO (ESCO v.1.1) will contribute to this exercise by adding a revised terminology on transversal skills, addressing duplications and quality issues and adding new terms related among others to the green, digital and scientific domain. At the same time, improving the IT tool's AI algorithm and user interface is crucial in order to address the issues raised by the participants in the second phase.

To this end, additional functionalities for the IT tool's user interface were developed in the first half of 2021 and the Commission will develop an improved version of the algorithm by end-September 2021. The objective is to make the tool a standalone service available to Member States and public and private organisations willing to establish links between ESCO skills and qualifications, learning opportunities in Europass, digitally signed credentials and other types of learning (such as micro credentials).

The Commission will therefore submit to the ESCO MSWG and the EQF AG a further call for expression of interest to participate in further testing for the further development and improvement of the tool.