



EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR EMPLOYMENT, SOCIAL AFFAIRS AND INCLUSION

Labour Mobility and International Affairs
Labour Mobility, Public Employment Services, ELA

Brussels
EMPL.E.1/FL

MEMBER STATES WORKING GROUP ON ESCO

Plenary meeting

8 December 2022, 9:30, ONLINE

Towards ESCO version 1.2

BACKGROUND

ESCO needs to be frequently updated to reflect emerging trends in the labour market.

Since the early conception of ESCO, the Commission designed a continuous improvement cycle as the precondition for providing a reference terminology that can bridge the communication gap between labour market actors and the education and training domain and support mobility of workers and cross-border matching. This can be summarised in the figure below:

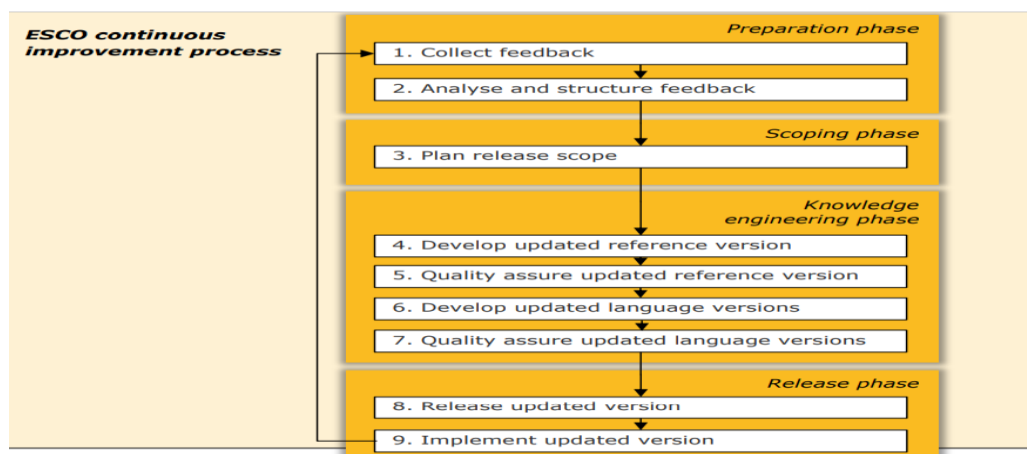


Figure 1. ESCO continuous improvement process

The first update cycle of ESCO (ESCO version 1.1) focused in particular on occupations emerged in the labour market since 2017 and on skills and competences related to the impact of digital technologies and the greening of the economy.

This content update process resulted in 515 new concepts and in the update of 35 ESCO v.1.0 concepts, out of which 109 concepts belong to the domain of digital technologies and 148 are labelled as green occupations, skills and knowledge. A revised thesaurus of transversal skills and list of green and research skills were published as part of this major version, while a list of digital skills was included with the release of the minor version ESCO 1.1.1. Since June 2022, ESCO is also available in Ukrainian.

In the context of the 2023 European Year of Skills, the Commission will publish a new major version of ESCO, version 1.2. This document details the work strands for the preparation of ESCO v1.2, looking in particular at the actions undertaken to improve the quality of the classification and at the creation of new content (occupations, skills and knowledge concepts).

The guiding principles that have been prioritised for the scope of ESCO version 1.2 are indicated with a blue background.



Figure 2. Overview of challenges grouped by guiding principle

QUALITY REVIEW

ESCO still contains inconsistencies and mistakes that need to be resolved in order to improve the quality of the classification. Examples of identified quality issues are:

- Duplicate concepts: occupations and/or skills that are similar in meaning and could be merged.
- Ambiguous skills: skills that can be interpreted in different ways and therefore are difficult to classify, leading to confusion for the end-users. This could be due to broad descriptions, use of more than one action verbs in the concept's title (Preferred Term) or descriptions that are different from the corresponding Preferred Terms.
- Formulation issues: Preferred Terms that are too long, Preferred Terms or descriptions that are formulated inappropriately, have typographic errors, wrong

sentences, action verbs used in terms (preferred or non-preferred) that need revision, etc.

- Inconsistent level of detail:
 - Contextualisation of (transversal) skills: many contextualised skills are not related to one another, making it difficult to easily visualize this contextualisation.
 - Some skills are too detailed and are actually detailed tasks or activities while others are very generic and ambiguous.
 - Some sectors have many occupations while others have only few.
 - Some occupations have many skills while others have only few.

Addressing all quality issues in a single ESCO version is an extremely resource-intensive activity, taking into account the translation effort and the need to ensure that Member States have sufficient time to be consulted on and review the proposed changes.

The Commission is therefore implementing quality improvements in different steps. Changes that do not impact the meaning of a concept and do not require translations (such as improving relations between occupations and skills or allocating skills to different skill groups) are implemented on a continuous basis through the publication of minor versions. Quality improvements affecting the list of skills and occupations (such as addressing duplicate concepts) are implemented only in major versions.

ESCO version 1.2 will be built on the work undertaken in ESCO version 1.1 and address in particular the following challenges:

1. Quality review of the ESCO knowledge pillar.
2. Addressing duplicate skills.
3. Improve the skills-occupations relations.

Quality review of the ESCO knowledge pillar

Knowledge concepts are defined as *the outcome of the assimilation of information through learning. Knowledge is the body of facts, principles, theories and practices that is related to a field of work or study.* The main distinction between a knowledge and a skill is that the first represents a theoretic field rather than an action which, instead, is the way to describe skills' concepts.

Although the ESCO terminological guidelines state that knowledge concepts should be described using nouns, in some cases they present verbs, which may create confusion. A full review of the 2000 knowledge concepts in ESCO is therefore needed to handle such critical cases.

For this purpose, the Commission is applying Artificial Intelligence (AI) algorithms to detect knowledge concepts whose descriptions start with verbs. More specifically, AI models are trained to distinguish verbs from nouns or other grammatical structures and are deployed to specifically search for verbs at the beginning of the concepts' descriptions. Once a verb is detected, the ESCO Secretariat will analyse whether the knowledge concept is instead a disguised skill and adapt the metadata accordingly. Should a concept be considered as a knowledge, the description is modified by replacing the verb with a

constructed sentence. When all changes are completed, the revised descriptions are then translated in all the ESCO languages.

Figure 3 below shows an example of the work that is done to improve the knowledge pillar.

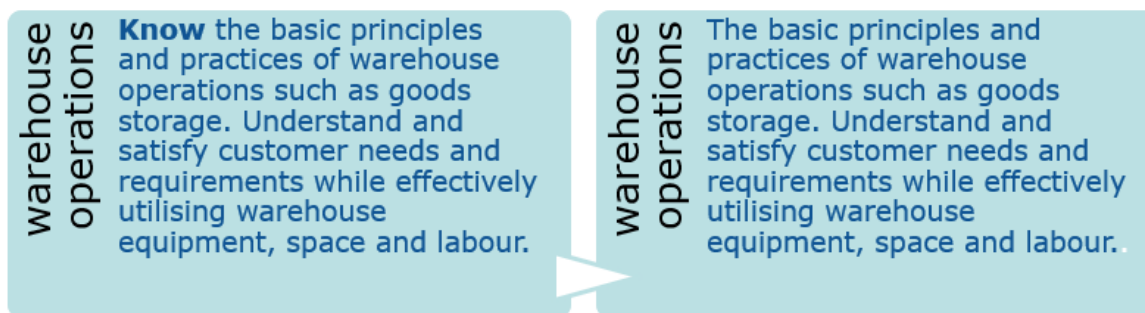


Figure 3. Example of improvement actions for the ESCO knowledge pillar

Addressing duplicates skills

As the ESCO Skills Pillar collects information on almost 14,000 concepts, it is not unexpected that multiple skills may share significant similarities.

While preserving the variety and richness of the ESCO skills thesaurus is a precondition to offer a labour market terminology that can work for both labour market and education and training actors, in some cases ESCO implementers prefer to use a limited set of ESCO skills and adapt the level of granularity of the ESCO skills pillar to their specific context or use case.

Concepts in ESCO are considered duplicates when they describe the same activity within the same or a very similar working context. Addressing duplicates concepts is one of the first tasks to be handled to improve the quality of ESCO. In such cases, the skills should be merged into one unique concept, or one or more of the duplicates should be removed.

The methodology used to carry out this work is based on machine learning (ML) model suggestions and on results validation. The ML model is employed to detect the duplicates and obtain external statistics to support the decision making, while the final decision and subsequent improvement actions are taken by the ESCO Secretariat.

The work planned to detect and remove duplicate skills is carried out in three steps, as shown in the figure below:

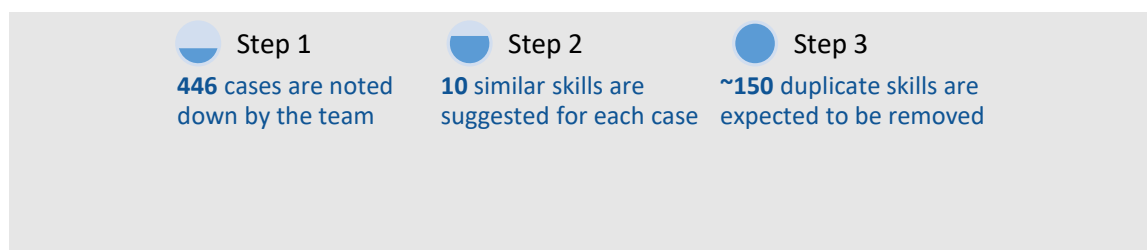


Figure 4. Detect and remove duplicate skills

- First, 446 skills were manually identified as candidate duplicates, since the same scope was repeated multiple times in different concepts. This work was carried out within the quality review work strand that took place in view of the release of ESCO v1.1.0.
- Given these critical concepts as input, the second step involves the use of ML model which assigns scores of semantic similarities to each ESCO concept. For each case, 10 similar skills/knowledge concepts are suggested. Moreover, to support the process of manual validation that takes place at the last step, statistics from Europass and EURES on the frequency of skills are extracted and added for each suggestion from the model. This helps the validators to detect which skill, among duplicate cases, is adopted with higher frequency by users of the ESCO classification.
- Finally, manual validation is performed to remove duplicate skills. The validators decide whether two or more skills are duplicates and proceed with a concrete action. This final step is developed in two different ways: if two concepts are exact duplicates, one of them is removed; if two concepts complete each other, they are merged. Likewise, in both cases, existing metadata (such as PTs, NPTs, and relations with occupations) are reviewed and adjusted to the resulting concept.

As a result of this process, **159 duplicate skills** are expected to be made obsolete in ESCO v1.2.

Improve the skills-occupations relations

For this part, the aim is to improve the quality of the relations between skills and occupations, the two pillars of ESCO. For some occupations, there is an “under-representation” in terms of the number of skills assigned to them, while others have the opposite issue, an “over-representation” of the occupation with an incoherent number of skills associated to it.

To address this challenge, the ESCO Secretariat will compute statistics to select outliers out of the distribution of skills per occupations. Subsequently, Machine Learning (ML) models provide suggestions of skills that may be removed from over-represented occupations and skills that may be added to under-represented ones.

A further round of statistics on internal (based on ESCO) and external data (e.g., online job vacancies from EURES) is used to support decision making for these cases. The final step involves manual validation. An overview of the preliminary findings of this exercise is presented below.

Cluster	Issue	Number of cases
Underrepresentation	Occupations with <5 essential skills	64
Overrepresentation	Occupations with >70 essential skills	4
	Occupations with >70 optional skills	69

Table 1 Thresholds for under and over-represented occupations

CONTENT UPDATE

The process for updating ESCO content is structured in three work packages: the preparation phase, the research and analysis phase and the implementation phase (creation of new content and changes to existing occupations). The visual below summarizes the main steps undertaken:



Figure 6. Overview of content update process

ESCO version 1.2 will build on the content update process implemented in ESCO version 1.1 and will focus in particular on:

- Impact of the twin transitions on skills and occupations
- Stakeholder input
- Data-driven update

Impact of the Twin Transitions on skills and occupations

Given the recognized effort to support the transition towards a greener and more digital economy, ESCO v1.2 will help reflecting these key drivers of change by focusing on digital scientific and green skills and promoting transversal and entrepreneurial skills and skills for life. These are the aspects informing the prioritisation of changes to implement in the next version of ESCO, as well as the subsequent research activity and stakeholders' engagement.

The content update is based on the input from domain experts, desk research, statistics extracted from job vacancies and online user profiles. Quality review of concepts already labelled as green and digital is also implemented.

While ESCO already provides for a large set of digital skills and knowledge concepts and ICT occupations, the update of the classification will focus on emerging technologies such as AI, blockchain, IoT or cloud computing, and on key sectors such as big data analysis and cybersecurity. On the other hand, ESCO green skills will be reviewed to identify gaps and improvement needs. The sectors under review, in this case, are related to renewable energies, energy-efficiency, bioeconomy, agri-food systems and circular and social economy.

The update of the green skills thesaurus will also build on the analysis of existing data sets. As an example, the ESCO Secretariat compared the current list of ESCO green skills with a dataset of green competences identified by the European Training Foundation (ETF)

through the analysis of online job ads. Once again, this work is implemented through a combination of ML models and manual validation.

Namely, ML models compare ESCO skills with the green skills identified in online job vacancies and suggest the most similar ESCO skills for each label. The models compare not only the exact wording, but also make use of the contextual information to ensure that the suggested matching concepts (among ESCO concepts) are effectively similar to those proposed by implementers. Afterwards, manual validation is carried out to review the proposed matches and decide on the way forward.

As a first step, a file with about 144 green skills was used to run the skill/knowledge suggestion model to be able to propose the most similar ESCO concepts. In a second stage, the team performed the manual validation case by case.

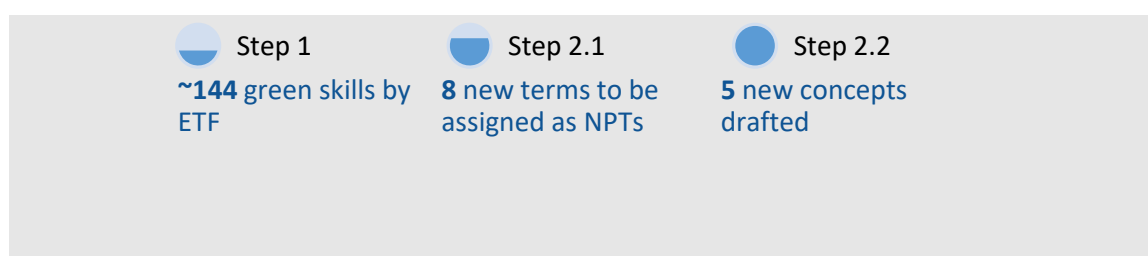


Figure 7. Overview of ESCO green skills maintenance

The review resulted in 8 new terms assigned as NPTs and the drafting of 5 new concepts. This exercise, extended on a high number of data, will allow to draw some conclusions on the overall coverage of green skills and competences in ESCO.

Stakeholder input

External feedback is provided by different domain experts. The input collected from Blueprint projects, sectoral associations, social partners, NGOs, private and public stakeholders as well as other Commission services, varies from detailed comments of occupations and skills to broader labour market analysis. During the feedback process, these external inputs are analysed and discussed in detail, to select what relevant content could be used for ESCO v1.2.

The following table offers an overview of the sectors currently under review:

SECTORS
Additive Manufacturing
Agri-food ecosystem
Batteries for electro-mobility
Bioeconomy & Agriculture innovation
Blockchain
Cloud computing

Creative & Cultural Industries
Cultural heritage
Digital (AI)
Energy-intensive industries
Energy value chains
Maritime Shipping
Microelectronic
Proximity & Social Economy
Tourism

Table 2. Labour Market sectors under review

The process of gathering inputs from stakeholders is developed in different stages. The first focuses on analysing the input collected during the feedback phase. Hereafter, the ESCO Secretariat revises the occupations and skills mapped to each sector under review, to identify existing gaps in the classification.

Considering the feedback received and the results of the desk research, the following stage is to assess whether there is a need to change or draft new concepts in ESCO. This intervention could be carried out in three different ways:

- Creation of new skills and knowledge concepts.
- Creation of new occupations to reflect emerging trends in the labour market and fill the existing gaps.
- Update of existing ESCO occupations and skills to better reflect the demands of the labour market.

Domain experts are actively involved in this process, including the support to drafting of occupations and skills concepts metadata. This ensures that the implementation of the input received will meet labour market trends.

Data-driven update

Data science techniques are another important approach for identifying potentially missing occupations and skills.

For ESCO v1.2, the Commission will leverage this methodology in combination with the results of the mapping of national classifications to ESCO (e.g. Member State and O*NET mapping tables), online job advertisements and Europass user profile data. The candidate concepts as obtained from this data-driven approach will then be subject to manual expert validation.

- Mapping tables for national classifications:

Mapping tables establish the correspondence between national taxonomies and ESCO concepts based on four matching relations: exact matches, narrower matches, broader matches and close matches. Moreover, they provide an indication of the closer ISCO group when it is not possible to identify an ESCO concept corresponding to a national occupation.

The data taken into consideration includes all matching relations other than exact matches from all available Member State mapping tables and the O*NET crosswalk. The established relations are reviewed by applying AI models that suggest the most similar ESCO concepts based on semantic similarity.

Thresholds are applied to select those cases that need further analysis: above a fixed threshold of similarity, national occupations are considered as already covered by similar ESCO terms while, for the occupations that are below the threshold, statistics are computed to check the frequencies of the occupational titles with clusters of occupations identified in EURES online job advertisements. If frequencies are above a certain threshold, a new occupation is considered for drafting. In addition, EURES data can be explored to suggest NPTs and skills that could be associated with the newly created occupations.

- EURES online job advertisements and Europass user profiles:

Additionally, data mining of EURES online job advertisements and Europass profiles represents a complementary strategy to identify new occupations and skills.

First, clustering methods can be used to group together similar vacancies or user profiles. Next, this allows to enrich the resulting clusters of records with meta-data such as geographical coverage, frequency statistics and statistical dispersion. These clusters can then be compared to existing ESCO occupation concepts through a multilingual machine learning model and applying similarity score cutoffs (e.g. for filtering out clusters of occupations that are already present in ESCO). This allows to identify potentially missing occupations based on a list of criteria, e.g. sufficient geographical coverage, reasonable presence in the labour market, distance from existing ESCO concepts. As explained in the previous section, the EURES and Europass data can also be used to provide suggestions for alternative titles for the newly drafted occupations and to suggest corresponding skills.

TIMELINE

The figure below shows a summarized action plan for ESCO version 1.2, focusing on the challenges listed above.

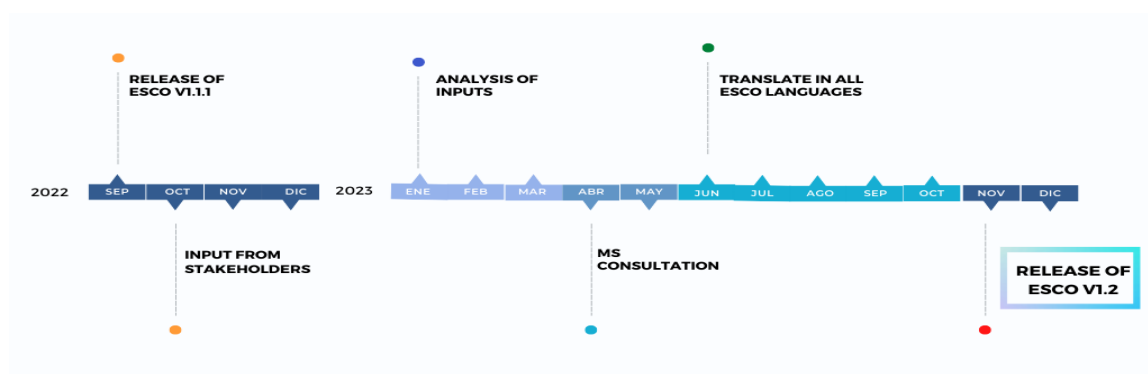


Figure 8. Summarized overview of planning for ESCO version 1.2

Overall, the analysis of input, creation and validation of new content will be completed by **March 2023**.

Member States will then be consulted on the improvements to existing ESCO concepts and on the new content. A pre-release of ESCO version 1.2 will be made available on the testing environment of the ESCO portal, together with the instructions on how to provide feedback and suggest changes to the proposed content. The pre-release will also be made available to the wider community of ESCO implementers.

The MSWG consultation on the English version will be concluded in **May 2023**. Translations will start by **June 2023** so that the new version could be published in **November 2023**.

The consultation of the ESCO Member States Working Group on the different language versions will run in parallel to the publication process. Member States will have three months to provide feedback on translations of new and modified concepts, which will then be implemented in a subsequent minor version in the **second quarter of 2024**.