ESCO Skill-Occupation Matrix Tables:
linking occupation and skill groups

Technical Report – April 2021

Employment,
Social Affairs
and Inclusion

Table of Contents

1. The context

As the number of ESCO implementers increases, the range of uses of the classification is widening. The ESCO dataset is currently employed by public and private implementers as the reference language for employment and education, being widely recognised as the dictionary for the European labour market. However, implementers might not always look for something as granular as ESCO skills or occupations, as they may prefer smaller-sized and aggregated datasets which are easier to handle for their own purposes. Thus, the European Commission aims to unravel the complexity of the ESCO dataset through more dynamic illustrations of how ESCO concepts can be connected and used at more aggregated levels.

In this document, the Commission is showcasing potential interrelations (defined as matrix tables) based on the hierarchical structure offered by the ESCO and ISCO-based hierarchies (for both skills and occupations).
The ESCO Skill-Occupation Matrix Tables presented below, allow users to **connect ISCO-08 occupation groups (rather than one single occupation) to ESCO skills hierarchical groups (rather than one single skill).** Starting from the most granular level of the ESCO classification, these tables **show the share of ESCO skills within ISCO-08 occupation groups**.

For example, one may now investigate the share of 75 skill groups (level 2 of the ESCO skill hierarchy) among 125 occupation groups (level 3 of the ISCO occupation hierarchy). The infographic below (Figure 1) shows the new possibilities available today for implementers and other stakeholders, compared to the previous state of use of ESCO.

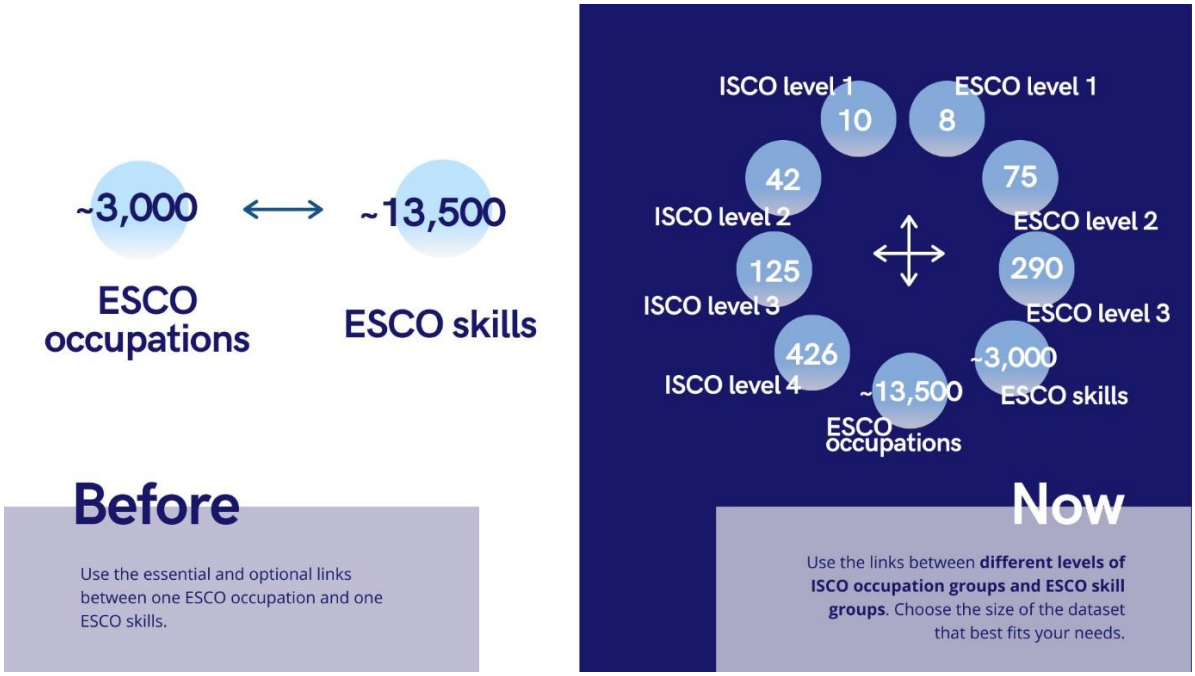

Figure 1. Comparison of available datasets, before and after the exercise.

In order to ensure a correct understanding and support the reproducibility of this exercise, Section 2 describes the methodology adopted and how to obtain the same results working independently. The ESCO Skill-Occupation Matrix Tables can be downloaded from the ESCO Portal. If needed, the ESCO team is available to provide further support and share XLS tables of the requested size (e.g. a table showing matches between ISCO level 2 and ESCO level 3). Section 3 presents the visual product developed for this exercise to ease the understanding of the tables. Finally, section 4 gives more insights on the possible use of these tables by ESCO implementers who share different goals.

2. Methodology

2.1 Data gathering

All the employed datasets are available in the ESCO portal (Download section):

Version	Type	Language	ESCO pillar	Name	Use
V1.0.8	CSV	All	All pillars	Occupations/Skills (19MB)	Relationship (essential/optional) between occupations and skills
			Skills/competences	Skills/competences (4MB)	Distribution of skills: narrow/broader relationships and hierarchical position

			Occupations	Occupations (1MB)	Distribution of occupations: narrow/broader relationships and hierarchical position

Table 1: ESCO datasets needed for this exercise.

2.2 Data cleaning

The datasets have been further structured and have undergone some data cleaning activities:

- **All the knowledge concepts have been removed**. Since the knowledge concepts are mapped to a different hierarchy than the skills hierarchy, they have not been considered for this exercise.
- The dataset *Occupations/Skills* has been divided into two datasets, one for the essential relationships between skills and occupations, and one for the optional relationships. The two datasets have then been analysed separately, as further explained in the following section.
- **All the skills that are not mapped to the skills hierarchy have been removed from the dataset.** These are: Language skills, Attitudes skills, Transversal skills. This activity did not need any manual work, since they are automatically not considered when working with the skills hierarchy using the skills/competences dataset.

2.3 Data merging and matching

The *Skills/competences* and *Occupations* datasets have been then re-worked in such a way to obtain structured data for every level of the two hierarchies. The resulting datasets are structured as such:

- Skill/competences: a first column with skills' URIs, followed by three columns for each of the skill hierarchy's level (see Table A1 in the Annex Section).
- Occupations: a first column with occupations' URIs, followed by four columns for each of the ISCO hierarchy's level. An additional column with the 5th ESCO level can be created, upon request (see Table A1.2 in the Annex Section).

For all the following activities, the pandas Python library has been employed.
All the steps that follow have been repeated twice: one time for the essential relationships, one for the optional relationships.

First, a matrix has been created using the *Skills/competences* dataset. The matrix has skills' URIs in the rows and skill groups in the columns – the level of the skill groups (i.e. whether to use level 1, 2, or 3) can be chosen based on needs. In the matrix, the values represent the distribution of each skill within skill groups, with binary values 1-0, where:

- Value 1 means that skill X is mapped to skill group Y,
- Value 0 means that skill X is not mapped to skill group Y.

It is important to consider here that one skill could be mapped to more than one skill group – this happens when one skill is narrower to another skill which is mapped to a different skill group. In this case, two are the suggested options:

1. Each row has value x ≥ 0. Here, the cell has value 1 whenever skill X is mapped to skill group Y.
2. Each row has value 1 ≥ x ≥ 0. In this case, the following formula should be applied:
   For every cell = 1 in row i:

$$value\ cell_i = \frac{1}{\sum value\ cells_i}$$

Either options suggest similar results, and for simplicity reasons option 1 was adopted.

Second, data from the first matrix has been merged with the Relationship matrix. As a result, a new matrix has been created, which matches ESCO occupations with ESCO skill groups. In the matrix, the rows list occupations' URIs and the columns list skill groups. The values represent the distribution of each skill group within one occupation.

Third, data from this final matrix has been merged with data from the Occupations dataset. The resulting matrix matches ISCO occupation groups (in the rows) with ESCO skill groups (in the columns). Once one repeats this exercise for both essential and optional relationships, two matrices are obtained. At this point, all the values in the matrix of optional relationships have been multiplied by 0.5, hence reducing their impact in the overall result.

The final step is to sum up the two matrices, then values need to be normalised by rows, using the following formula:

For every cell in row i:

$$value\ cell_i = \frac{value\ cell_i}{\sum value\ cells_i}$$

This allows to see the frequency of one skill group in one occupation relative to the sum of the frequency of every skill group for that same occupation.

2.4 Result: Skill-Occupation Matrix Tables

The result consists in a matrix table for every chosen level of occupation groups and skill groups. The size of the matrix depends on the level selected from the hierarchy. The hierarchies are structured as follows:

Content (developed by)	Level	Size (i.e. number of groups)
Occupation groups (ISCO)	01	10
Occupation groups (ISCO)	02	42
Occupation groups (ISCO)	03	125
Occupation groups (ISCO)	04	426
Occupations (ESCO)	Most granular level	2,942
Skill groups (ESCO)	01	8
Skill groups (ESCO)	02	75
Skill groups (ESCO)	03	290
Skills (ESCO)	Most granular level	13,485

Table 2: Composition of ESCO and ISCO hierarchies.

As an example, the following is the matrix table for the first level of the occupation hierarchy and the first level of the skills hierarchy. The table shows in percent values the distribution of skill groups

among occupation groups. Looking at *Clerical support workers*, one observes that workers in this group have a higher share of *Communication, collaboration and creativity* skills, while they have a low share of *Handling and moving* skills.

	communication, collaboration and creativity (S1)	information skills (S2)	assisting and caring (S3)	management skills (S4)	working with computers (S5)	handling and moving (S6)	Constructing (S7)	working with machinery and specialised equipment (S8)
Armed forces occupations	24.8%	16.3%	24.3%	23.3%	2.6%	0.5%	0.2%	7.9%
Managers	28.8%	20.9%	12.9%	32.7%	2.6%	1.4%	0.0%	0.6%
Professionals	36.3%	22.3%	12.8%	17.7%	6.3%	2.0%	0.2%	2.4%
Technicians and associate professionals	28.4%	23.1%	14.4%	17.4%	4.0%	4.4%	1.0%	7.2%
Clerical support workers	28.8%	21.9%	15.9%	21.9%	5.3%	2.8%	0.1%	3.2%
Service and sales workers	31.1%	13.8%	24.5%	15.2%	1.0%	11.7%	0.0%	2.6%
Skilled agricultural, forestry and fishery workers	12.2%	16.1%	10.7%	9.9%	1.3%	37.1%	0.8%	11.9%
Craft and related trades workers	14.7%	20.1%	10.2%	6.2%	3.0%	20.6%	6.7%	18.5%
Plant and machine operators and assemblers	10.5%	19.8%	13.3%	5.2%	2.7%	20.4%	2.1%	26.0%
Elementary occupations	11.2%	12.5%	18.7%	6.9%	1.4%	34.3%	3.8%	11.1%

Table 3: Matrix table ISCO-01 occupations and ESCO-01 skills

2.5 Limitations

Due to the different size of occupational groups, in this case one cannot compare the distribution of skills within the whole dataset – this means that **columns of the tables should not be compared**. If one would rather be interested in the distribution of occupations per skill group, the whole methodology could be repeated except for the final step, where the normalisation should be conducted by columns instead of rows.

2.6 Methodological note

Skill groups vary by size, and the methodology applies the normalisation of the size only in the last step. Skill groups with a significantly bigger size may be expected to show higher shares of frequency among different occupation groups. However, this difference in size does not affect the quality of the result and interpretation of the matrix. In fact, different groups collect skills that are sector-independent and differ in terms of i) tools and equipment used, ii) type of object on which the work is performed, and iii) the function or outcome of the task or activity. This applies as long as one occupation is matched with the correct number of skills, i.e. avoiding that one occupation is mapped with two skills that may be considered as duplicates. In this respect, the Commission is involved in a continuous effort to ensure the respect of high-quality standards.

3. The visual representation

The Skill-Occupation Matrix Tables are represented in a more user-friendly way using heatmaps. Using the Plotly library, each value of the matrix tables is represented as a heatmap pixel. Values are then assigned to colours in a blue scale depending on their share level. When the user moves the cursor above the heatmap, additional data is showed, namely: full name of skill group, full name of occupation group, share value. Heatmaps can be created for every size of the matrix table. One heatmap is currently showed in the Latest News section of the ESCO portal.

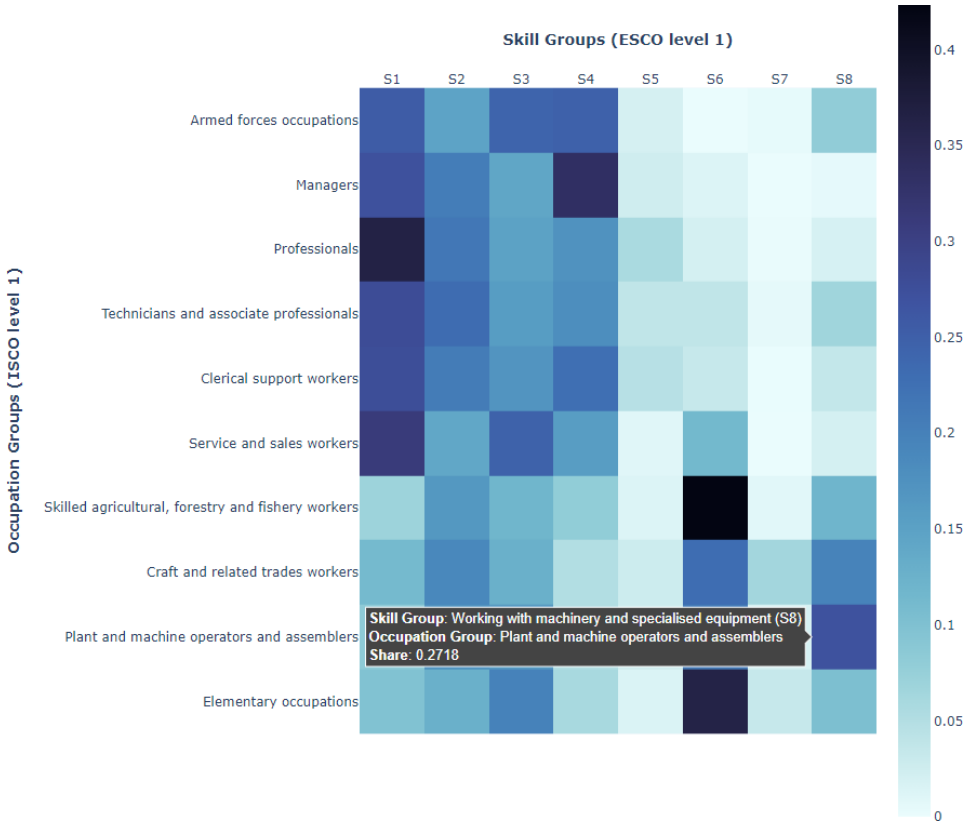

Figure 2: Heatmap ISCO-01 occupations and ESCO-01 skills

4. Use cases

In this section, we consider possible uses of the Skill-Occupation Matrix Tables for three groups of ESCO implementers and their use cases.

Figure 3: Examples of use cases

4.1 Public and private employment services

Public Employment Services (PES), Human Resource departments, and IT software providers developing products to manage the workforce are among the main implementers of the ESCO classification. They usually are organisations of heterogeneous sizes, and for this reason in some cases they would prefer to adopt ESCO at higher hierarchical levels. They can consult the matrix tables for the following purposes:

1. **Ease the mapping to ESCO to classify and manage employees**. Connecting a classification or mapping employees using ESCO may sound challenging when handling a dataset of about 3,000 occupations and 13,500 skills. As a, solution implementers can use higher hierarchical levels of the ESCO or ISCO classification. It is sufficient to choose the preferred hierarchical level, and eventually conduct a more detailed selection on the occupation groups desired. The tables will then allow to receive the information provided by ESCO concerning the skill distribution among occupations, as well as descriptions for each group of skills and occupations.
2. **Improve matching algorithms.** The values offered by the matrix tables can be further analysed and used as inputs to build models that suggest potential career paths, help balancing human resources, and, in general, connect people to jobs. For example, the values can be employed as weighting factors and using higher hierarchical groups may improve the quality of matches.

4.2 Research bodies

Research institutes, data companies, and others currently use the ESCO classification to gather insights from the European and International labour market. Based on feedback gathered by the ESCO team, these stakeholders may gain significant benefits from using a more flexible classification. They can consult the matrix tables for the following purposes:

1. **Connect the ESCO classification with data gathered externally.** The granularity and size of ESCO constitute a challenge when attempting to match data collected by national and international statistical institutes with ESCO's occupations or skills. This issue may be solved adopting ESCO and its matches between occupations and skills at a broader level (i.e. higher hierarchical groups).
2. **Build surveys and investigate results**. Implementers can use the matrix tables and the matches between occupations and skills to build surveys and collect data on labour market variables. The surveys can be developed in a more targeted way, focusing on different levels of granularity or designing datasets based on shared values.

4.3 Education institutions

ESCO aims at connecting the labour market to education and training systems. While already a significant number of organisations have successfully implemented the classification in their own services, the matrix tables may be relevant to explore different uses and respond to new needs. More specifically, the tables may be helpful in the following cases:

1. **Better define career prospects.** When making choices on education and study programmes, students consider the linked career prospects. From the education providers side, this information is as crucial to share as difficult to gather. One solution is to use the links between skills acquired in the study programme with skills demanded by the labour market for certain occupations. In this case, using skill groups at a higher level of the ESCO hierarchy can help education institutions matching their learning outcomes with skills, while the tables will offer detailed information of the skill groups' distribution at occupation level.
2. **Conduct graduate tracking.** As tracking graduates allows to gather information on skills use in the labour market and career paths of graduates, ESCO can serve as the common vocabulary that allows for the identification of common indicators on skills and occupations graduates have. The matrix tables would ensure accessibility to the ESCO classification.

5. Accessing the Skill-Occupation Matrix Tables

The Skill-Occupation Matrix Tables connecting occupation groups (from the ISCO hierarchy) and skill groups (from the ESCO hierarchy) can be developed independently following the methodology described above (Section 2), downloaded from the Documents section or requested to the ESCO team here or via email (EMPL-ESCO-SECRETARIAT@ec.europa.eu). Every service offered is free of charge. The Commission welcomes feedback and is available to provide technical support. Suggestions and questions can be shared in the online Forum available in the ESCO Portal.

6. Annex

Table A1: Extract from the *Skill/competences* dataset. The first column lists skills' URIs, while the three following columns list the URIs of the skills' broader hierarchical levels (ESCO hierarchy).

conceptUri	ESCO Level 3	ESCO Level 2	ESCO Level 1
http://data.europa.eu/esco/skill/0005c151-5b5a-4a66-8aac-60e734beb1ab	http://data.europa.eu/esco/skill/S4.8.1	http://data.europa.eu/esco/skill/S4.8	http://data.europa.eu/esco/skill/S4
http://data.europa.eu/esco/skill/000709ed-2be5-4193-b056-45a97698d828	http://data.europa.eu/esco/skill/S3.3.0	http://data.europa.eu/esco/skill/S3.3	http://data.europa.eu/esco/skill/S3
http://data.europa.eu/esco/skill/000c94d2-2a2e-4545-993c-6df8cb5b0316	http://data.europa.eu/esco/skill/S6.7.3	http://data.europa.eu/esco/skill/S6.7	http://data.europa.eu/esco/skill/S6

Table A1.2: Extract from the *Occupations* dataset. The first column lists occupations' URIs, while the four following columns list the URIs of the occupations' broader hierarchical levels (ISCO hierarchy).

conceptUri	ISCO Level 4	ISCO Level 3	ISCO Level 2	ISCO Level 1
http://data.europa.eu/esco/occupation/000e93a3-d956-4e45-aacb-f12c83fedf84	http://data.europa.eu/esco/isco/C8121	http://data.europa.eu/esco/isco/C812	http://data.europa.eu/esco/isco/C81	http://data.europa.eu/esco/isco/C8
http://data.europa.eu/esco/occupation/0022f466-426c-41a4-ac96-a235c945cf97	http://data.europa.eu/esco/isco/C3155	http://data.europa.eu/esco/isco/C315	http://data.europa.eu/esco/isco/C31	http://data.europa.eu/esco/isco/C3
http://data.europa.eu/esco/occupation/002da35b-7808-43f3-83bf-63596b8b351f	http://data.europa.eu/esco/isco/C2431	http://data.europa.eu/esco/isco/C243	http://data.europa.eu/esco/isco/C24	http://data.europa.eu/esco/isco/C2