

Dataset online platforms

“Occupational Skill Similarity and Occupational Transitions in six European countries”

Authors: Xavier Pinho, Sadegh Shahmohammadi

Grant agreement no. 101132581

Deliverable: D2.2

Due date: 31.03.2025



**Funded by
the European Union**

Document control sheet

Project Number:	101132581
Project Acronym:	SkiLMeeT
Work-Package:	2
Last Version:	31.03.2026
Issue Date:	24.07.2026

Classification

Draft	
Final	x
Confidential	
Restricted	
Public	x

Legal notice

This project has received funding from the European Union's Horizon 2022 Research and Innovation programme under grant agreement No. 101132581. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

Table of contents

1	Overview	4
2	Occupational Skill Similarity Dataset	4
2.1	Unit of observation	4
2.2	Time coverage	5
2.3	Data sources	5
2.4	Variables.....	5
3	Occupational Transitions Dataset	6
3.1	Unit of observation.....	7
3.2	Time coverage	7
3.3	Data sources	7
3.4	Variables.....	7

1 Overview

The European labour market is evolving rapidly, reshaping career trajectories and changing the skill requirements associated with different occupations. Addressing these developments requires data that captures detailed and comparable labour market characteristics across countries.

Online professional profile data represents a valuable source of information for analysing these trends. To support research on the European labour market, we have transformed large-scale online profile data provided by Revelio Labs into two complementary formats. These formats enable researchers to investigate key labour market dynamics through: (1) cross-country comparisons and (2) analyses of occupational transitions within individual countries. The countries are: Netherlands, Germany, Poland, Italy, Luxembourg and Bulgaria.

The dataset is developed within the SkiLMeeT project.

2 Occupational Skill Similarity Dataset

The Occupational Skill Similarity Dataset measures the similarity of occupation-specific skill profiles between countries and between countries and ESCO. For each occupation, the dataset compares the most frequently reported LinkedIn skills associated with workers in a country to the skills associated with the same occupation in another country or in ESCO.

Skill similarity is computed using semantic embeddings generated from skill descriptions and summarized through multiple similarity and distance metrics. The dataset enables researchers to assess how closely labour market skill profiles align across countries and with the ESCO occupational framework.

2.1 Unit of observation

The unit of observation is an occupation comparison pair.

Each record represents one occupation (identified through an ESCO occupation code) and compares the skill profile of that occupation between two entities:

- Country vs. ESCO (e.g., Italy vs. ESCO)
- Country vs. Country (e.g., Italy vs. Bulgaria)

2.2 Time coverage

The dataset does not represent a specific reference year or time interval. The dataset does not measure occupations at a specific point in time. Instead, it is constructed from the complete set of available online professional profiles used in the analysis. Skill profiles are aggregated across all individuals observed in each occupation, and similarity measures are calculated on these aggregated skill distributions.

2.3 Data sources

The dataset combines information from:

- Revelio Labs online professional profile data, providing occupation and skill information from professional profiles.
- ESCO (European Skills, Competences, Qualifications and Occupations), providing occupation codes and essential occupation-specific skills.
- SentenceTransformers (all-MiniLM-L6-v2) embeddings, used to represent skills and calculate semantic similarity measures.

2.4 Variables

Variable	Description
JOB_TITLE	ESCO occupation title being compared.
JOB_CODE	ESCO occupation code.
SIMILARITY	Cosine similarity between the aggregated skill profiles of the two entities. Higher values indicate greater similarity.
DIFFERENCE	Euclidean distance between the aggregated skill profile vectors. Higher values indicate greater difference.
MEAN_PAIRWISE_DISTANCE	Average cosine distance between top skills from the two entities.
MEDIAN_PAIRWISE_DISTANCE	Median cosine distance between top skills from the two entities.

Variable	Description
AVG_DISTANCE_FROM_CENTROID_[ENTITY A]	Average distance of Entity A skills from their skill-profile centroid. Measures skill dispersion.
AVG_DISTANCE_FROM_CENTROID_[ENTITY B]	Average distance of Entity B skills from their skill-profile centroid. Measures skill dispersion.
[ENTITY A]_SKILL_COUNT	Number of skills included in Entity A's occupational profile.
[ENTITY B]_SKILL_COUNT	Number of skills included in Entity B's occupational profile.
[ENTITY A]_SKILLS	Most frequent skills associated with the occupation in Entity A.
[ENTITY B]_SKILLS	Most frequent skills associated with the occupation in Entity B.
[ENTITY A]_UNIQUE_USERS	Number of unique workers contributing to Entity A's occupational skill profile.
[ENTITY B]_UNIQUE_USERS	Number of unique workers contributing to Entity B's occupational skill profile.

3 Occupational Transitions Dataset

The Occupational Transitions Dataset captures job-to-job mobility patterns observed in online professional profiles. Individual career moves are aggregated into transition types, defined by movements between a source occupation and a destination occupation.

A transition type represents movements between different occupations (e.g., Office Manager → Human Resources Manager) or within the same occupation (e.g., Office Manager → Office Manager), reflecting career progression, employer changes, or occupational stability.

The dataset allows researchers to analyse the frequency, characteristics, skills, and demographic composition of occupational mobility across countries and labour market segments.

3.1 Unit of observation

The unit of analysis is a transition type, defined as a unique combination of:

- Source occupation
- Destination occupation
- Occupational classification systems (ESCO and O*NET)

Each record aggregates all individuals observed making that transition.

3.2 Time coverage

The dataset does not represent a specific reference year or time interval. The dataset does not represent transitions observed within a specific reference period. Instead, it aggregates all occupational transitions identified in the available employment histories of individuals in the underlying dataset.

3.3 Data sources

Online profile data provided by Revelio Labs (LinkedIn profiles).

3.4 Variables

Variable	Description
Transition	Occupational movement from the origin occupation to the destination occupation.
Frequency	Number of individuals observed making the transition.
Percentage	Share of the transition among all observed transitions.
From (ESCO label)	ESCO occupation title of the origin occupation.
From (ESCO code)	ESCO occupation code of the origin occupation.
From (ONET label)	O*NET occupation title of the origin occupation.
From (ONET code)	O*NET occupation code of the origin occupation.
To (ESCO label)	ESCO occupation title of the destination occupation.

Variable	Description
To (ESCO code)	ESCO occupation code of the destination occupation.
To (ONET label)	O*NET occupation title of the destination occupation.
To (ONET code)	O*NET occupation code of the destination occupation.
Sex	Distribution of individuals by sex within the transition.
Country	Distribution of individuals by country within the transition.
Skills	Frequency distribution of skills associated with individuals making the transition.
Compensation Change	Average change in compensation associated with the transition (in USD).
Dates	List of transition dates observed for individuals making the transition.
To_Green	Indicates whether the destination occupation is classified as green.
From_Green	Indicates whether the origin occupation is classified as green.
To_Digital	Indicates whether the destination occupation is classified as digital.
From_Digital	Indicates whether the origin occupation is classified as digital.



Skills for labour markets in the digital and green transition

SkiLMeeT is a research project carried out by eight research institutions, exploring how the digital and green transformation of European economies are changing the demand for skills. We analyse the scope of skills shortages and mismatches, identify their drivers and propose innovative ways to reduce skills gaps.

Follow us:



[linkedin.com/company/SkiLMeeT](https://www.linkedin.com/company/SkiLMeeT)



bsky.app/profile/skilmeet.bsky.social



twitter.com/SkiLMeeT

www.skilmeet.eu



Funded by
the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

SkiLMeeT Partners:

