SkiLMeeT

Conceptual Framework

Ronald Bachmann, Christina Gathmann, Federico Fabio Frattini, Vassil Kirov, Piotr Lewandowski, Francesco Vona and Ulrich Zierahn-Weilage

Reviewer: Mairi Bowdler

Grant agreement no. 101132581

Deliverable: D1.2 Technical Report Concept

Due date: 31.03.2025

Funded by the European Union

Document control sheet

Project Number:	**101132581**
Project Acronym:	SkiLMeeT
Work-Package:	1
Last Version:	28.03.2025
Issue Date:	31.03.2025

Classification

Draft	
Final	x
Confidential	
Restricted	
Public	x

Keywords

Concept, data sources, indicators, measurement issues

Legal notice

This project has received funding from the European Union's Horizon Europe 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. Introduction 4

2. Concepts 4

3. Data 6

4. Indicators and Measurement 9

5. Conclusion 10

6. References 11

1. Introduction

The aim of this document is to provide an overview of SkiLMeeT's conceptual framework. The document thus provides a reference point for all analytical work packages (WPs 2-5) and is also used in WP6 in the context of the engagement of stakeholders (Task 6.2), the research community (Task 6.3), and communication activities regarding the broader public (Task 6.6).

The SkiLMeeT project has three main scientific objectives:

1) Create indicators of skill gaps and shortages using various data sources (Work package 3).
2) Analyse the drivers of skill gaps and shortages (WP 4).
3) Analyse pathways to reduce skill gaps and shortages (WP 5).

In the following, the document provides an overview of the strategy used in SkiLMeeT to deal with the key concepts, the data sources, as well as indicators and related measurement issues.

2. Concepts

SkiLMeeT conducts cutting-edge research on a number of topics related to the measurement of skill gaps and shortages, the drivers of these gaps and shortages, and pathways to reduce them:

- The measurement of labour and skill shortages and mismatches at the sectoral and educational levels (Task 3.3).
- Similarities and distances between occupations (Task 3.4).
- Matching efficiency (Task 3.5).
- Digitalisation, demographic change and globalisation as drivers of skill needs and shortages (Tasks 4.1, 4.2, 4.7, 4.8).
- The green transformation as a driver of skill needs and shortages, resulting in reskilling needs (Tasks 4.3 – 4.5).
- Working conditions and job-to-job transitions (Task 5.1).
- Skill transferability across jobs and occupations (Tasks 5.2 and 5.3).
- Skill gaps and mismatch, and worker mobility and migration (Tasks 5.4 and 5.7).
- Educational choice of youth (Task 5.5).
- Training, resilience to technology shocks and skill transitions (Tasks 5.6 and 5.8).

These topics are related to a multitude of studies stemming from different literature strands. As the SkiLMeeT deliverable "D1.1: Overview of literature and Initiatives" shows, these literature strands are extensive and use specific definitions, indicators, and measurement procedures. The SkiLMeeT project therefore must strike a balance between relating to existing literature and initiatives on one hand, and providing a coherent approach to analysing skill gaps and shortages within the SkiLMeeT project on the other.

In addition to the growing scientific literature on the topic, SkiLMeeT also relates to a number of policy initiatives regarding skill gaps and shortages. This includes, but is not limited to, initiatives from the following organisations:

- The SkiLMeeT sister projects SKILLAB, SkillsPulse and TRAILS,[1]
- The European Labour Authority (ELA) and the European Employment Services (EURES) network (see e.g. ELA, 2024),
- Cedefop (see e.g. Cedefop, 2024),
- European Commission, Directorate-General for Employment, Social Affairs and Inclusion (see e.g. European Commission, 2024).

Relating the SkilMeeT project to existing literature and initiatives is important for several reasons. Firstly, it allows for the application of well-established concepts and methods. Second, it clarifies the contribution of the SkiLMeeT research to the existing literature, in turn simplifying the process of publishing the results in peer-reviewed scientific journals. Additionally, it facilitates the communication of the SkiLMeeT results by utilizing established concepts and methods.

At the same time, SkiLMeeT aims at making its components as coherent as possible. This is achieved through various means, including discussions at the virtual project meetings, at SkiLMeeT seminars and conferences, during multi-institution research projects meetings, and other instances.

[1] Detailed information on the sister projects can be found at the following websites: https://skillab-project.eu/, https://www.skillspulse.eu/, https://www.trails-project.eu/.

3. Data

SkiLMeeT uses both traditional (e.g. survey-based or administrative) and innovative (e.g., big data or text) data sources to investigate skills demand, supply, and mismatch, and quantifies their drivers such as the green and digital revolution. Table 1 provides an overview of the main data sources that are used in the project. Furthermore, data will be generated from text documents using Natural Language Processing (NLP) and Machine Learning (ML) techniques.

Table 1: Main data sources used in SkiLMeeT

Name	Countries	Period	Provider	WP
European Labour Force Survey (EU-LFS)	EU + some non-EU countries	1995-present	Eurostat	2-5
European Working Conditions Survey (EWCS)	EU + some non-EU countries	2005, 2010, 2015, 2021	Eurofound	5
European Union Statistics on Income and Living Conditions (EU-SILC)	EU countries	2005,2020	Eurostat	5
Programme for the International Assessment of Adult Competences (PIAAC)	31 countries (incl. 20 EU countries)	2012/13, 2022/23	OECD	2-4
Online Job Vacancies (OJV)	Germany, France, Luxemburg, Belgium	2018-present	Lightcast	2-5
LinkedIn - Professional Network Data on Companies and Individuals	EU + about 200 countries worldwide	2015-present	LinkedIn	2-5
Job Vacancy Statistics (JVS)	EU	2001-present	Eurostat	3, 4
European Skills, Competences, Qualifications and Occupations (ESCO)	EU	2017-present	EC (DG EMPL)	2, 3
European Patent Office (EPO) Data	EU	1990-2018	EPO	2, 4
Community Production (PRODCOM)	EU	1995-2017 / 2018 - present	Eurostat	4

The traditional data sets such as the EU-LFS, EU-SILC, EWCS and PIAAC are the foundation for the construction of indicators and for analyses of the distribution of workers across occupations and sectors by country (used e.g. in Task 3.1: Heterogeneity of skill levels and workers tasks in Europe), their socio-demographic characteristics, and working conditions (e.g. Task 5.1: Working conditions and well-being for workers facing a job-to-job transition).

In addition to the multi-country data sets listed in Table 1, for specific project tasks, country-specific data sets were used, such as German administrative data on workers' labour-market trajectories (Sample of Integrated Labour-market Biographies, SIAB, provided by the Institute for Employment Research, IAB); Italian linked employer employee data (Comunicazioni obbligatorie, COB, provided by the National Institute for the Analysis of Public Policies, INAPP); the Italian Labour Force Survey and the Italian occupational database including tasks and skills (ICP), both also provided by INAPP; the French establishment survey on energy consumption and CO2 emissions (provided by The National Institute of Statistics and Economic Studies, INSEE), and French occupation data, through the Annual Declaration of Social Data, DADS (also provided by INSEE).

Text-based data sets are used for different purposes. Online job vacancies (OJV) are generally used to measure the demand side of the labour market. In SkiLMeeT, two specific applications are focused on: measuring occupational distances as outline in Task 3.4 (Mapping the similarities and distances between occupations), and the identification of green skills. For the latter, we apply and adapt the NLP methodology of Saussay et al. (2022). Data from LinkedIn and from the training curricula for Germany (legal texts from the German Ministry of Justice) are used to analyse the supply of specific skills (Task 2.3: Data generation for labour and skills shortages in the EU).

We also use several data sources to identify technology shocks. This includes robot data from the International Federation of Robotics (Task 5.5: The effect of automation technologies on the educational choices of youth) and patent data from the European Patent Office (EPO) and from Google Patents, following Kelly et al. (2021). In addition to identifying technology shocks (e.g. Task 4.2: Drivers of skills needs and shortages in the labour market), patent data is also used to identify skill demand in the context of the green transition (Task 4.3: Skills needs and shortages in the context of the green transformation).

Combining these diverse data sets is important to conducting the quantitative analyses required in SkiLMeeT. The data provides information at different levels of aggregation, namely individual, occupational, and sectoral levels (e.g., PIAAC data on skills and OJV data on skill demand to analyse

skills mismatch at the occupational level). We use multiple methods of linking data selecting the most appropriate approach for the specific data set and research question at hand. The basic approach for combining data sources uses common classifications (e.g. occupational or industry codes) or correspondence tables (e.g. a crosswalk from the US to the European occupational classification), that allow linking data from surveys to text-based data, or survey data to administrative data, or text-based data to administrative data. If there is no common classification or correspondence table, we rely on alternative methods. For example, to match patents, which are coded according to the International Patent Classification (IPC) with workers in industries that use the knowledge embodied in a given patent, we rely on Algorithmic Links with Probability (ALP) proposed by Lybbert and Zolas (2014). This method uses the definition of an industry, which describes what the industry produces, and the keywords of the patents in the IPC code. To link data on knowledge embodied in patents to skills learned in an occupation (vocational curricula) or tasks performed in a job, we extend this approach to text-to-text linkages. This develops measures of similarities between documents. We use established NLP methods such as bag-of-words, N-grams, term frequency-inverse document frequence (tf-idf), topic models, as well as word embeddings, to identify which words and combinations of words characterize documents. Based on the word/s embeddings that characterize the documents, the distance between documents is computed by assessing how well the characteristic words in both documents match with each other.

Finally, to link French online job postings to firm characteristics, we use probabilistic matching and geocoding to link each job posting to an individual firm id (SIREN) using the name and location of the plant. Based on this unique firm identifier, we can then link the data to worker biographies (including information on hirings, turnover or wages) and firm balance sheet data (like revenues).

In addition to quantitative data, we rely on qualitative information. This is particularly important in the analysis of pathways to reduce skills gaps and shortages. We therefore conduct occupational case studies to examine how transitions contribute to a better skills match, and identify the factors that determines successful transitions, including upskilling and reskilling of workers. The qualitative research strategy is described in detail in the SkiLMeeT deliverable D1.3: Research Protocols.

4. Indicators and Measurement

After collecting and combining the necessary data, we develop indicators for skills supply, skills demand, labour shortages, and skills mismatch. We also examine the drivers of these shortages and mismatches such as the green and digital revolutions.

For indicators on the supply of skills, we quantify the skills supply of the current workforce using surveys on skills (e.g., PIAAC data) and data from online platforms (LinkedIn) to map and compare the supply of cognitive and non-cognitive skills across European countries and how they evolve over time (Task 3.1: Heterogeneity of skill levels and worker tasks in Europe). To quantify the skills acquired in vocational training programs in Germany, we construct indicators using various NLP methods (e.g., N-grams, tf-idf, topic models, or word embeddings) to identify and condense the skills listed in training programs into suitable indicators (Task 2.5: Collecting data on vocational training curricula and linked technological change embedded in patents).

We rely on OJV and LinkedIn data, as well as survey data on skills to build new indicators of the demand for skills. Using OJV and LinkedIn data involves the development of taxonomies, where we use the original texts of the vacancies. In addition to broad skill groups like social skills or routine task intensity, we focus especially on the identification of skills related to the digital and green transition (Task 3.2: Indicators for green and digital skills and specialized versus diversified skill sets). To this end, we create extensive lists of relevant keywords to identify online job vacancies that require digital and green skills (see Deliverable D2.1: Dataset with keywords for skills). We supplement this algorithm-based approach with manual annotation to verify the accurate identification of vacancies requiring specific skills. We then calculate indicators on the occupations and industries requiring digital or green skills, the variation of green or digital skills within an occupation or the spatial distribution of digital or green jobs and how these evolve over time. We also describe changes in skills used at work at the worker-level in different occupations and sectors using survey data (Task 3.1).

We use several indicators to analyse skills mismatch in European labour markets, which are based on the estimation of matching functions and Beveridge curves for specific EU27 countries (Task 3.5: Indicators of matching efficiency in European countries) and workers' self-assessment of mismatch (Task 3.1). Beveridge curves are the basis for creating indicators of labour market tightness. The matching functions, in turn, allow us to compare matching efficiency across countries and thus shed light on countries' ability to match individuals to jobs. We first estimate national matching functions

and Beveridge curves using data on vacancies and unemployment merged to the EU-LFS. We then provide more detailed indicators on labour market tightness and matching efficiency by estimating occupation-specific Beveridge curves and matching functions, while accounting for skills differences between occupations.

To create indicators of past and current labour and skills shortages, we combine data on new hires, the number of unemployed, and firms' challenges in filling vacancies by country, occupation and time (Task 3.3: Indicators for past, current and forecasted skill shortages and mismatches).

To shed light on the pathways out of skills shortages, we build indicators on occupational mobility (Task 3.4: Mapping the similarities and distances between occupations, and Task 3.2: Indicators for green and digital skills and specialized versus diversified skill sets). We use two alternative measures: first, we create indicators on actual worker flows across occupations using EU-SILC (as in Bachmann et al., 2020). Second, we construct a measure of skills similarity using the angular separation measure (Gathmann and Schoenberg, 2010) and factor analysis (Ingram and Neumann, 2006); and indicators of skills specialisation and diversification using the correlation of Balassa indices (Saussay et al., 2022) for all European countries.

To identify new digital and green technologies as potential drivers of skills shortages and mismatch, we rely on patent data (Task 4.2: Drivers of skill needs and shortages in the labour market, Task 4.3: Skills needs and shortages in the context of the green transformation, Task 4.4: The green transition and reskilling need). We first follow the procedure by Autor et al. (2022) and differentiate between automation and augmenting innovations to distinguish different types of technological change. Moreover, following these authors, we identify breakthrough patents for developing a causal identification strategy. To identify green and digital technological innovations, we rely on NLP searching for keywords in the title and abstracts of patent descriptions to identify patents in green, AI or robotics technologies.

5. Conclusion

This document has summarised the key concepts, data sources as well as indicators and measurement issues relevant for the SkiLMeeT project. It serves as a reference point for the research and communication activities in the project. Naturally, some procedures will have to be updated during the project as the research on skill gaps and skill shortages evolves within SkiLMeeT and beyond.

6. References

Autor, D., C. Chin, A. Salomons, B. Seegmiller (2024). 'New Frontiers: The Origins and Content of New Work, 1940–2018'. The Quarterly Journal of Economics; qjae008, https://doi-org.proxy.library.uu.nl/10.1093/qje/qjae008

Bachmann, R., Bechara, P., and Vonnahme, C. (2020). 'Occupational mobility in Europe: Extent, determinants and consequences'. *De Economist*, *168*(1), 79–108.

Cedefop (2024), 'Untangling labour shortages in Europe: unmet skill demand or bad jobs?'. Luxembourg: Publications Office. Cedefop working paper series No 20.

ELA (2024), 'EURES Report on labour shortages and surpluses 2023, Luxembourg: Publications Office of the European Union.'

European Commission: Directorate-General for Employment, Social Affairs and Inclusion (2024), 'Labour market and wage developments in Europe – Annual review 2024'. Luxembourg: Publications Office of the European Union.

Gathmann, C., Schönberg, U. (2010). 'How General is Human Capital? A Task-Based Approach'. Journal of Labour Economics, 28(1):1-50.

Ingram, B. F., & Neumann, G. R. (2006), 'The returns to skill'. Labour economics 13(1): 35-59.

Kelly, B., D. Papanikolaou, A. Seru, & M. Taddy (2021). 'Measuring Technological Innovation over the Long Run'. American Economic Review: Insights, 3 (3): 303-20.

Lybbert, T. J. & Zolas N. J. (2014). 'Getting patents and economic data to speak to each other: An Algorithmic Links with Probabilities approach for joint analyses of patenting and economic activity'. Research Policy, 43(3), pp. 530-542.

Saussay A., Sato M., Vona F. & O'Kane L. (2022), 'Who's fit for the low-carbon transition? Emerging skills and wage gaps in job and data', FEEM Working Paper 031. Milano, Italy: Fondazione Eni Enrico Mattei.

SkiLMeeT

Follow us:

linkedin.com/company/SkiLMeeT

bsky.app/profile/skilmeet.bsky.social

twitter.com/SkiLMeeT

www.skilmeet.eu

Funded by
the European Union

This project has received funding from the European Union's Horizon Europe Research and Innovation programme under grant agreement No 101132581.

SkiLMeeT Partners:

TNO innovation for life

rwi

LISER Luxembourg Institute of Socio-Economic Research

Utrecht University

ibs institute for structural research

Fondazione Eni Enrico Mattei

Institute of Philosophy and Sociology, Bulgarian Academy of Sciences

ALDGATE STRATEGY GROUP