Co-funded by the European Union

Methodological note - Unwrapping green, digital and entrepreneurial (GDE) skills

By UNIPI

Methodological note
Unwrapping green, digital and entrepreneurial (GDE) skills

Project result number D2.1 by UNIPI

28 October 2022

Co-funded by the European Union

Table of contents

1. Executive Summary

This report presents the work done for listing the green, digital, and entrepreneurial (GDE) skills and knowledge. We decided to start the project with a task able to guide the scope of the rest of the project: defining the most important GDE skills. The goal is to measure the relevance of GDE skills, to link this information to the Open Educational Resources we will collect. Educators and students will be able to have a quantitative and multi-perspective (science, companies and job market) view on the importance of the skills they want to teach/learn. Given the variety and the different characteristics of each group of skills, we follow a two-step approach, related to Task 2.1 "*Categorising ESCO skills on GDE skills*" and Task 2.2 "*Enhancement and revision of the identified skills by the companies*", as we explain below. Also, given the limited time frame to perform this preparatory task (5 months) we adopted data driven approaches, previously developed by the research group of University of Pisa. This document is a Methodological note, explaining to other organizations how to implement task 2.1, in such a way that anyone that is already skilled in programming and data analysis, can replicate the task (e.g., for other skill domains). As a plus, we decided to include in the note also a similar description for task 2.2, given the relevant novel method we developed in ENCORE (that mixes questionaries to companies with information coming from literature and from Skill Panorama) and the interesting results achieved also on this task. For these reasons, the document is structured as follows.

Section 2 discusses how we search for GDE skills using a methodology already developed in previous works of University of Pisa (Fareri et al., 2021; Chiarello et al., 2021). The methodology uses scientific literature to automatically collect skills on a specified domain, and then links these skills to European Skills/Competences, Qualifications and Occupations (ESCO). The link to ESCO is done using Natural Language Processing tools able to measure the semantic similarities between skills. Since GDE skills are transversal and cross-domain, to retrieve a relevant set of scientific articles the query design process is the basis for retrieving technical documents. For this reason, a query for each domain (green, digital and entrepreneurship) is launched on Scopus, one of the most large and reliable scientific paper repositories. After a manual revision of a sample of the collected papers, the articles are used to extract skills, and to link these skills to the ESCO framework. However, for enlarging the sources from where the skills are collected, we use different database. For Digital skills, we rely on the classification of digital skills already identified by ESCO. For Green skills, we also use the classification already given by Occupational Information Network (O*NET), the USA framework of skills and jobs. Finally, for Entrepreneurial skills, the EntreComp framework is used.

Section 3 presents the validation and assessment process we perform for understanding the relevance of the skills in each domain (i.e., GDE). The process is composed of two main steps. The first is explained in Section 3.1 and it relies on a questionnaire proposed to organisations (companies, research and governmental institutions), where we ask companies to list the GDE skills relevant for their business, without showing them the results of Section 2. In this way, organisations are not biased by the results of Task 2.1. This step will further enhance the recall (likelihood of considering all the relevant GDE skills) of the skill listing activity of T2.1. During the second step explained in Section 3.2, we assess the relevance of the skills collected in Section 2 and updated in Section 3.1 using the European platform Skills Panorama. This step allows us to have a data-driven assessment (coming from the analysis of job vacancies done by CEDEFOP for the platform), avoiding possible biases of the contacted organisation (e.g., too narrow focus on specific domain, low awareness on GDE skills).

The results of task 2.1 (Section 2) and task 2.2 (Section 3) are available online[1]. This table contains more that GDE 1.500 skills, all enriched with the information collected during the first two tasks of the ENCORE project.

[1] https://docs.google.com/spreadsheets/d/1b67q4O8XEcH2SNlhq9dZ0ZLWLcg094uev-vTkW7_dI0/edit#gid=0

2. Categorising ESCO skills on GDE skills

This section presents the method adopted for task T2.1 ”Categorising ESCO skills on GDE skills”. The aim is to describe the method and the results underpinning the task of extracting Green, Digital and Entrepreneurial (GDE) skills, described as follows:

- Green Skills (G): knowledge, competences and attitudes linked to the transition to a circular and greener economy;
- Digital Skills (D), knowledge, competences and attitudes linked to the confident, critical and responsible use of, and engagement with, digital technologies for learning, at work, and for participation in society;
- Entrepreneurial Skills (E): knowledge, competences and attitudes linked to the capacity to act upon opportunities and ideas, and transform them into value (financial, cultural, or social) for others (source EntreComp).

In this methodological note we used scientific literature to map skills onto a specified domain, and then we linked these skills to the European Skills/Competences, Qualifications and Occupations (ESCO) framework using Natural Language Processing (NLP). Natural language processing is a branch of text mining, focused on the analysis of human language in written format (natural language in jargon), to extract information from the text, a type of unstructured data (Gupta et al., 2009).

The aim of the description is to make the approach reusable and the results reproducible by the scientific community.

The whole task is represented in Figure 1.

2.1 Method: categorising ESCO skills on GDE skills

Our approach is divided in three phases: in the first phase, called “Paper Retrieval” (Section 2.1.1), we describe the strategies used to collect scientific papers concerning GDE skills; in the second phase, called “Skills Extraction with Named Entity Recognition“ (Section 2.1.2), we describe the design and the implementation of the NLP systems used to extract skills; in the third phase, called “Results Cleaning” (Section 2.1.3) we describe the cleaning, screening, and categorisation of the skills extracted from scientific papers.

The goal of the first phase was to retrieve a set of papers on the three domains under analysis. The papers are retrieved from Scopus[2], a database of scientific publications launched in 2004 by Elsevier. The Scopus database was chosen for its high volume of papers, frequency of updates, and comprehensiveness with respect to engineering and innovation papers. Also, Scopus offers free access to the API for Universities, which makes the data download more convenient and reproducible. Scopus contains abstracts and citation data from all scholarly journals indexed.

We downloaded from Scopus the abstracts of the scientific papers together with other metadata (such as publication type, publication year, source etc…). We analysed only the abstracts of papers to identify relevant GDE skills. We preferred to focus on the abstract instead of analysing the full paper, since we were searching for GDE skills that are properly mentioned by the authors and represent key concepts of the paper, rather than general mentions that could be found throughout the document. Additionally, mining the entire text could also retrieve false mentions of GDE skills (e.g., competencies mentioned in the state of the art of the paper), thus papers that are out-of-scope or weakly linked to the three domains under the analysis. This methodological choice (limiting the analysis to the

[2] https://www.scopus.com/. Accessed on Oct. 25, 2022.

abstracts) is also supported by other works that are focusing on detecting emerging competences using text mining techniques on scientific literature (Fareri et al. 2021, Chiarello et al., 2021).

We adopted a keywords-based search strategy for the three domains under analysis. The keywords used and the exclusion criteria for retrieving all and only papers related to GDE are described in the following sections. The number of retrieved scientific documents is specified for each strategy.

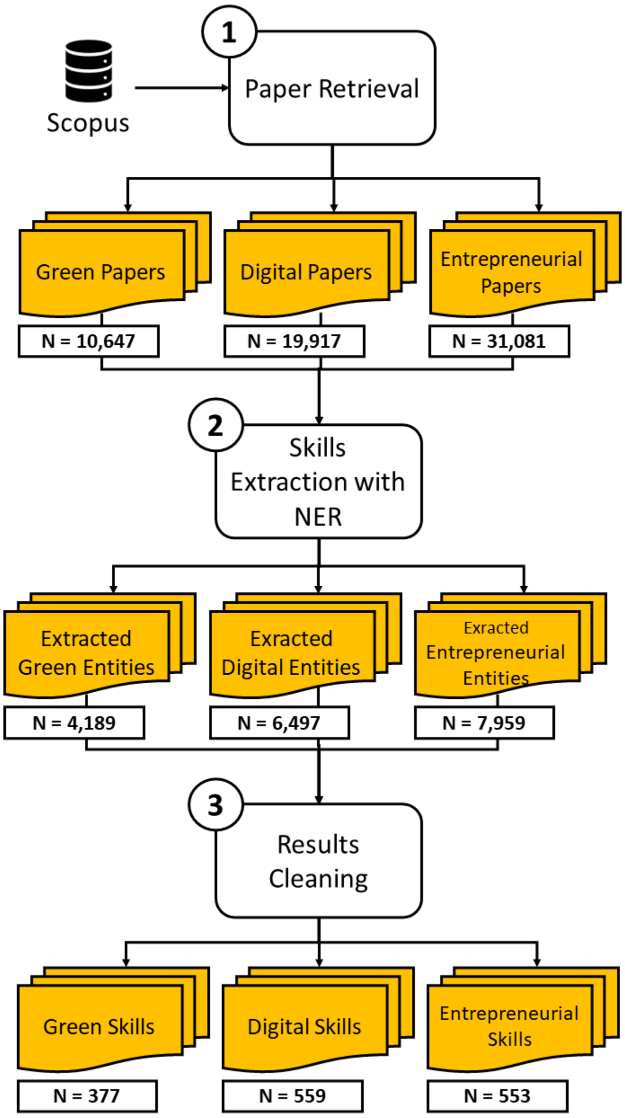

Figure 1. Workflow diagram to describe the steps of Task 2.1.

2.1.1 Paper Retrieval

Search Strategy: Green

In order to search Scopus for papers on Green related topics, we used European policies and directives to identify the main topics. We report in Appendix A the complete list of European policies and directives taken into consideration. Then, we identified keywords for each topic and revised them manually by removing generic, repetitive, or too specific terms. Finally, we added keywords related to the sphere of skills and abilities. We used the following query searching along with title, abstract and keywords of the documents that led to retrieve 10,647 papers[3].

```
TITLE-ABS-KEY ("circular economy" OR "ecodesign" OR "eco design" OR "industrial symbiosis" OR "waste prevention" OR "waste reduction" OR "secondary raw material" OR "resource recovery" OR "sustainability criteria" OR "bioeconomy" OR "bio economy" OR "biorefinery" OR "bio refinery" OR "renewable resource" OR "waste recycling" ) AND ( "skill*"  OR "competence*" OR "knowledge" OR "capabilit*" OR "attitud*" OR "abilit*" )
```

Query 1 - Papers on Green Skills

Search Strategy: Digital

We defined the query to retrieve Digital papers by narrowing the description of this type of skills (reported in section Introduction) into the following elements:

- Who: students, workers, citizens;
- What: knowledge, skills and attitudes;
- Where: digital domain.

According to these elements, we identified relevant keywords and relevant research fields related to Digital. The fields have been selected considering the indication of both the relevant ones (LIMIT-TO) and the not-relevant ones (EXCLUDE). We used the following query to retrieve papers related to Digital that led to retrieve 19,917 papers[4].

```
TITLE-ABS-KEY (digital* OR "industry 4.0") AND TITLE-ABS-KEY ("skill*" OR  "competence*" OR "knowledge" OR "capabilit*" OR "attitud*" OR "abilit*")) AND TITLE-ABS-KEY ( "human resource*" OR "HR" OR worker OR workforce OR "student*" OR "citizen*" )) AND ( LIMIT-TO ( SUBJAREA,"COMP" ) OR LIMIT-TO ( SUBJAREA,"ENGI" ) OR LIMIT-TO ( SUBJAREA,"BUSI" ) OR LIMIT-TO ( SUBJAREA,"DECI" ) OR LIMIT-TO ( SUBJAREA,"SOCI" ) OR EXCLUDE ( SUBJAREA,"MATH" ) OR EXCLUDE ( SUBJAREA,"PSYC" ) OR EXCLUDE ( SUBJAREA,"MEDI" ) OR EXCLUDE ( SUBJAREA,"PHYS" ) OR EXCLUDE ( SUBJAREA,"EART" ) OR EXCLUDE ( SUBJAREA,"CENG" ) OR EXCLUDE ( SUBJAREA,"NURS" ) OR EXCLUDE ( SUBJAREA,"AGRI" ) OR EXCLUDE ( SUBJAREA,"CHEM" ) OR EXCLUDE ( SUBJAREA,"BIOC" ) OR EXCLUDE ( SUBJAREA,"DENT" ) OR EXCLUDE ( SUBJAREA,"PHAR" ) OR EXCLUDE ( SUBJAREA,"NEUR" ) OR EXCLUDE ( SUBJAREA,"VETE" ) OR EXCLUDE ( SUBJAREA,"IMMU" ) OR EXCLUDE ( SUBJAREA,"HEAL" ) )
```

Query 2 - Papers on Digital Skills (see Appendix B for a detailed description of the constituent parts of this query)

Search Strategy: Entrepreneurial

To retrieve papers about Entrepreneurial (henceforth Entrepreneurship) we started from a simple search of the keyword “entrepreneur” to observe the results. This query produced too many results: 107,234 papers. Assuming that this amount might include a relevant number of out-of-scope papers, we added the following keywords to obtain a more specific subset of results: “skill”, “competenc”,

[3] Last update: 16/06/2022

[4] Last update: 16/06/2022

"abilit", "knowledge", "attitude" and "capabilit". We used the following query to retrieve papers related to Entrepreneurship that led to retrieve 31,081 papers[5].

```
TITLE-ABS-KEY(entrepreneur* AND (
skill* OR competence* OR knowledge OR attitude OR abilit* O
R capabilit* ))
```

Query 3 - Papers on Entrepreneurship Skills

Table 1 shows the different types of documents and the number of results for each GDE filed.

Table 1. Number of results per type of documents.

Document type	Number of results		
	Green	**Digital**	**Entrepreneurship**
Article	6,911	9,567	21,490
Conference Paper	1,123	7,489	4,506
Book Chapter	1,802	1,384	2,685
Review	536	495	1,281
Book	99	492	601
Conference Review	66	388	194
Editorial	37	41	119
Note	29	36	88
Short Survey	33	10	60
Retracted	1	5	30
Erratum	5	4	14
Letter	4	1	6
Report	1	1	4
Undefined	0	1	1
Article in Press	0	3	0

2.1.1 Skills extraction with Named Entity Recognition

To retrieve skills, we used Named Entity Recognition (NER). NER is an Information Extraction (IE) technique whose objective is to recognize information units (e.g. foods, person names, companies, geographical entities) in unstructured text (Nadeau and Sekine, 2007). We applied two NER methods:

1. *gazetteer-based NER*: a semi-supervised method which exploits lists of known GDE skills (Pawar et al., 2012) to retrieve them in texts and map their mentions to knowledge base resources. The creation of lists is an operation where manual effort and domain-specific knowledge are commonly required.

[5] Last update: 16/06/2022

2. *rule-based NER*: an unsupervised and top-down method where knowledge-based systems are built exploiting regular expressions and morphosyntactic information. These systems apply rules able to extract certain types of entities from texts (Jiang et al., 2011).

In the next sections we describe in detail the two approaches used.

Gazetteer-based NER

The gazetteer-based method consists in exploiting gazetteers of terms (i.e. GDE skills). To select a proper list of terms related to skills we use the following sources: ESCO (European Skills, Competences, Qualifications and Occupations)[6], a multilingual classification which includes 13,890 skills, 3,008 occupations, and many qualifications.

- O*NET (Occupational Information Network)[7]: an occupational framework developed by the U.S. Department of Labor. It is made of 974 occupations classified based on the Standard Occupational Classification (SOC) system and their corresponding skills, knowledge, abilities and technologies. From their database, we selected the following data to be considered as competences: Knowledge, Skills, Abilities, Work Activities, Work Context, Work Values, Work Styles, Emerging Tasks, Technology Skills & Tools. The reason for such a choice lies in the lexical pattern of those data, which is aligned with the above-described structure of ESCO concepts.
- EntreComp (Entrepreneurship Competences)[8]: a European framework of competences. It depicts the entrepreneurial capabilities in a transversal perspective by defining 3 competence areas, a list of 15 competences, learning outcomes and proficiency levels.

For the ESCO classification, we chose the "skills" category by selecting the preferred labels (the typical name of a skill), and their alternative labels. These have been included in the final list of ESCO skills, which contains 97,652 elements.

For the O*NET framework, we chose the following categories: "Knowledge", "Skills", "Abilities", "Work Activities", "Work Context", "Work Values", "Work Styles", "Emerging Tasks" and "Technology Skills & Tools". The reason for such a choice lies in the lexical pattern of those data, which is aligned with the above-described structure of ESCO concepts. We have increased the list of skills by considering further information provided for each competence as alternative labels (replicating the pair structure of preferred label - alternative labels present in ESCO). The final O*NET list of skills is made of 10,379 elements.

For the EntreComp framework, for each skill we selected the related hint and threads. The first clarifies the meaning of the competence, the second ones identify the context of application. The larger perspective of this taxonomy makes it necessary to check the ambiguity of the labels, therefore we revised the list removing labels with too general meaning (e.g., imagine and vision). We removed 45 labels. The final EntreComp list of skills is made of 61 elements.

Finally, merging the different sources, we used a complete list of 108,092 skills for the gazetteer-based method. We developed a Python script using the packages *pandas* and *TrieRegEx*[9]. The script takes as input the complete list of skills and a dataset of scientific papers; it creates a regular expression[10]

[6] https://esco.ec.europa.eu/en. Accessed on Oct. 25, 2022.
[7] https://www.onetonline.org/. Accessed on Oct. 25, 2022.
[8] https://joint-research-centre.ec.europa.eu/entrecomp-entrepreneurship-competence-framework_en || https://publications.jrc.ec.europa.eu/repository/handle/JRC101581 || https://publications.jrc.ec.europa.eu/repository/bitstream/JRC101581/lfna27939enn.pdf. Accessed on Oct. 25, 2022.
[9] https://pandas.pydata.org/ || https://github.com/ermanh/trieregex. Accessed on Oct. 25, 2022.
[10] Standard language to specify search patterns in text (Stephen Kleene)

pattern containing all the skills and iterates over all the papers' abstracts, searching the pattern in each of them. The script returns a file (*.csv* format) containing, for each paper, the list of skills found in its abstract.

The execution time for processing all the scientific papers (considering the three domains together) was equal to 180 seconds on a machine running an Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz processor and 32 GB RAM. The execution of the Gazetteer-based extraction brought to the results presented in Table 2.

Table 2. Statistics of the extracted entities from each field (gazetteer-based method).

Field	Number of scientific papers	Number of extracted entities	Number of unique extracted entities
Entrepreneurial	31,081	171,569	2,557
Green	10,647	64,596	1,889
Digital	19,917	119,549	2,641
Overall	**61,645**	**355,714**	**3,966**

Rule-based NER

The second family of NER systems we employed is based on rules. In a general sense, this means we defined morphosyntactic patterns that identified the presence of a skill in a text. This type of extraction can be made in several ways and there is literature about it (Hearst, 1992; Roller et al., 2018; Fareri et al., 2021). This methodology, if properly adapted to each case, turns out to be quite effective especially in terms of coverage.

Recent developments in NER demonstrate that the textual context in which an entity appears reveals recurrent patterns (Devlin et al., 2018). In the case of skills, the surrounding text that precedes the entity is a key clue to detect the skill itself. Indeed, as demonstrated by Fareri et al. (2021), there exist pieces of text that can be considered as clues for the presence of a skill. Let us consider the example presented in Fig. 2.

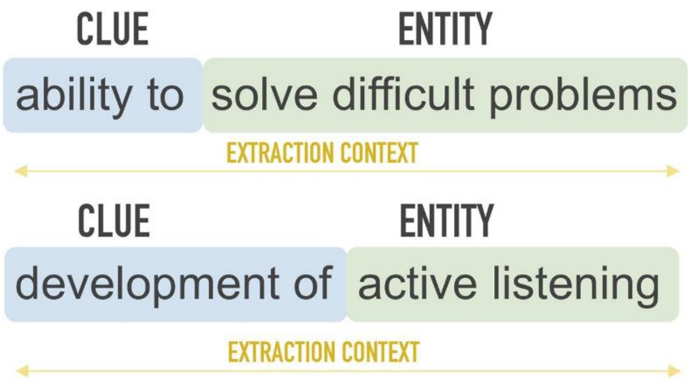

Figure 2. Example of extraction context of a skill; source: Fareri et al. (2021).

The two skills "solve difficult problems" and "active listening" are surrounded by words that introduce them ("ability to" and "development of"), creating the so-called extraction context. Hence, the extraction context is made of two elements that constitute the pillars of a rule-based NER system that detect skills:

- Entity: a linguistic sign (i.e., one word or a set of words) that references the soft skill;

- Clue: a set of terms, lexical expressions, or recurrent patterns correlated with the appearance of the soft skill.

In the present work, we designed a system based on the following clues: “ability to”, “able to”, “knowledge in”, “knowledge of”, “level of” and “proficiency in”[11]. Moreover, we extracted the subsequent text that appears after these clues using a stop criterion based on the dependency parsing[12]. We stopped the last relation of the dependency parser (namely the last word that is connected to the clue). As an example, in the sentence shown in Figure 3, our strategy highlights the text “solve difficult problems” as it is the last relation connected to the clue “ability to”.

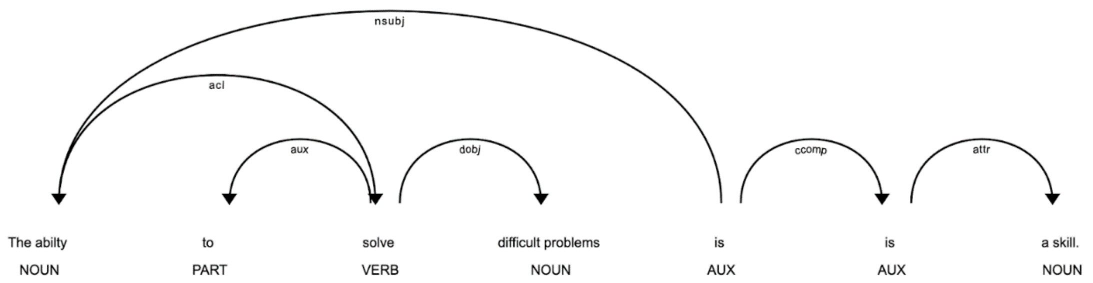

Figure 3. Example of dependency parsing of a sample sentence.

We developed a Python script to perform this rule-based extraction, using the Spacy package (Honnibal and Montani, 2017). The script extracts from the abstracts of the papers the skills according to the rules above explained.

The execution time for processing all the scientific papers (considering the three domains together) was equal to 360 seconds on a machine running an Intel(R) Core (TM) i7-6700 CPU @ 3.40GHz processor and 32 GB RAM. The execution of the Gazetteer-based extraction brought to the results presented in Table 3.

Table 3. Statistics of the extracted entities from each field (rule-based method).

Field	Number of scientific papers	Number of extracted entities	Number of unique extracted entities
Entrepreneurial	31,081	7,202	5,485
Green	10,647	2,624	2,338
Digital	19,917	5,051	3,935
Overall	**61,645**	**14,877**	**11,125**

Finally, table 4 presents the statistics of the entities extracted from each scientific paper, merging the results of both the gazetteer-based method and the rule-based one.

[11] According to Fareri et al. (2021), these clues are proposed.
[12] Dependency parsing is the process of determining how the words in a sentence are related to each other (Penn Treebank Dependency Parsing. (n.d.). Retrieved from http://www.ling.upenn.edu/advice/annotation/penn-treebank-dependency-parsing.html). Accessed on Oct. 25, 2022.

Table 4. Statistics of the extracted entities from each field (both gazetteer-based and rule-based method).

Field	Number of extracted entities	Number of unique extracted entities
Entrepreneurial	178,771	7,959
Green	67,220	4,189
Digital	124,600	6,497
Overall	**370,591**	**14,940**

2.1.3 Results Cleaning

The entities extracted from the scientific papers were processed as follows. After the skill extraction, a list of resulting skills was checked. The results contained both skills and elements which were not actual skills. Therefore, we applied an (i) automatic data cleaning and (ii) a data screening in order to produce a cleaner and more accurate result. A unique set of elements was produced by merging all the results together and removing duplicates. The total number of different entities extracted was 14,941. We executed the cleaning operations on this list of unique elements. This process consisted of a pipeline of three steps, which are explained in the next subsections: (1) Automatic Data Cleaning; (2) Data Screening; (3) Identifying Synonyms.

Automatic Data Cleaning

In this first step, we cleaned the results by removing automatically:

- entities which appeared in very few papers (less than 0.05% of the papers);
- very short entities, composed of less than three characters (e.g., *c*, *r*, *it*, *ai*);
- entities composed of numbers only;
- entities composed of stop words only;
- entities composed of domain stop words only.

Table 5. Criteria used to screen data for entities that are skills, and criteria used to screen data for entities that are non-skills.

Entities evaluated as skills	Entities evaluated as non-skills
Subjects and disciplines (e.g., Geology)	Badly written entities (e.g., “A production process)
Technologies, software and applications (e.g., “Google”)	Languages (e.g., “French”)
Programming languages (e.g., “Python”)	Single-word adjectives and verbs (e.g., “agile”, “casts”)
Acronyms of technologies (e.g., “SAAS”)	Entity which is probably not used as skill in papers (e.g., “algorithm”, “argument”, “capacity”)
	Religions (e.g., “Islam”)
	Materials (e.g., “Gold”)
	Sports (e.g., “Soccer”)

Furthermore, we manually revised the excluded elements, in order to avoid the loss of important information. After these operations, the number of entities decreased to 1,510.

Data screening

This second step consisted of a manual task performed by three students (a Ph.D student at the School of Engineering of the University of Pisa, a Ph.D. student of the Computer Science Department of the University of Pisa and a Msc. student of the course Data Science and Business Informatics of the University of Pisa), where the entities list has been screened to distinguish between actual skills and entities which are not skills. Each student has been provided with the table containing the 1,510 entities obtained in the Automatic Data Cleaning step. For each of them, the assignment was the following:

"*Read each extracted entity and decide whether the entity is a competence or not*".

The students were allowed to use external sources for achieving their goal. To decide whether an entity is a competence or not, several criteria have been designed. Table 5 shows the criteria adopted to evaluate entities as skills and those adopted to evaluate them as non-skills.

After each person screened the list, we discussed cases where there was not full agreement and made a final decision for each. This decreased the number of entities evaluated as actual skills to 1,107 elements.

Identifying Synonyms

The final step of this process was to try to convert all the resulting skills into ESCO competences. This was done for two reasons: first, we wanted to incorporate possible synonyms of a certain skill into a standard form, using ESCO as our main dictionary of standard skills; second, we wanted to analyse which entities could not be incorporated into ESCO skills, as these might be competences not yet included in the ESCO classification, or even entirely new competences.

We performed this task through the following operations. First, all the 1,107 elements were automatically searched in the ESCO skills list to find possible exact matches. 995 entities were matched with ESCO skills: no further operations were needed on them. Second, for the remaining entities that did not produce an exact match, a semi-automatic step was performed by exploiting the semantic similarity: the Bidirectional Encoder Representations from Transformers (BERT) model (Devlin et al., 2018) was used to map each extracted entity and ESCO competence into a vector and multidimensional semantic space. Converting skills into vectors is effective for identifying similarities between them, as recently demonstrated by Chiarello et al. (2021). The cosine similarity between all the extracted entities' vectors and all the ESCO skills' vectors was calculated as the cosine of the angle between the two semantic vectors (Han et al., 2011). The final step of the semantic similarity process involved taking the four most similar ESCO skills for each extracted entity. A manual screening step was used to determine if one of the four selected ESCO skills would have been a correct alternative label for each extracted competence. We found a coherent ESCO match for 63 skills. For the remaining 49 skills, none of the four alternatives were valid. A sample of the results of the task is provided in Table 6. The Extracted Competences column contains groups of similar skills: they have been assigned to the same ESCO competence through the initial automatic search or through the semantic similarity calculation process. The assigned ESCO skill is written in the other column.

After the Identifying Synonyms process, all the extracted competences have been converted into the corresponding ESCO skills assigned. The result of the whole step is a list of 804 ESCO skills, which incorporates the initial list of 1,107 extracted entities.

Table 6. Sample of the Identifying Synonyms step results; an ESCO skill has been assigned to each of the Extracted Competences.

Extracted Competences	Corresponding ESCO Skills
Working with others Work collaboratively	Cooperate with colleagues
Work together Cooperate Work in teams Work in groups Working in teams Team-working	Work in teams
Programming Computer programming Programming languages	Computer programming

2.2 Results of the categorisation of ESCO skills on GDE skills

A total number of 370,591 skills were retrieved from 61,645 papers during the Skills Extraction phase. 14,940 unique entities were identified after duplicates were deleted. To delete noise which could be present in the list (e.g., entities which are not competencies), further cleaning operations were applied. The number of entities decreased to 1,510 after the first automatic cleaning. These remaining elements were manually screened. 1,107 of them were evaluated as relevant competences for GDE domains.

We measured the accuracy of the method by calculating the percentage of competences manually tagged as correct. Considering this metric of accuracy, the score obtained is 73.31%. After the manual revision, we used semantic similarity algorithms to group skills that have the same meaning but are expressed with different spelling. This task led us to assign a unique label for each group of skills. The chosen label is the preferred label of ESCO whether the skill related concept is contained in ESCO. Finally, we obtain 804 unique skills.

In the following sections we describe the results of the extraction for each field. Moreover, Table 7 presents the complete statistics of the whole process, from the Paper Retrieval to the Results Cleaning.

The complete result of this task is available online[13].

Table 7. Number of results retrieved in each phase. The phases are divided into tasks, and the results are shown for each phase.

Phase	Task Description		Number of results			
			G	D	E	Total
Paper Retrieval	a	Paper Retrieval (No. of papers)	10,647	19,917	31,081	**61,645**
	b	Paper that contains at least one skill (No. of papers)	10,565	19,724	30,370	**60,659**
	c	Paper that contains at least one skill (%. of papers) (**b/a**)	99.23%	99.03%	97.52%	**98.40%**
Skills Named Entity Recognition	d	Gazetteer-based method (No. of entities)	1,889	2,641	2,557	**3,966**
	e	Rule-based method (No. of entities)	2,338	3,935	5,485	**11,125**

[13] https://docs.google.com/spreadsheets/d/1b67q4O8XEcH2SNlhq9dZ0ZLWLcg094uev-vTkW7_dI0/edit#gid=0

(Unique)	f	Both methods (No. of entities)	38	79	83	**151**
	g	*Total Number of skills (No. of entities)* ***(d + e - f)***	*4,189*	*6,497*	*7,959*	***14,940***
Skills Named Entity Recognition (Total)	h	Gazetteer-based method (No. of entities)	64,596	119,549	171,569	**355,714**
	i	Rule-based method (No. of entities)	2,624	5,051	7,202	**14,877**
	l	*Skills Named Entity Recognition (No. of entities)* ***(h + i)***	*67,220*	*124,600*	*178,771*	***370,591***
Results Cleaning	m	Automatic Data Cleaning (No. of entities after Automatic Revision)	683	1,004	1,025	**1,510**
	n	*Data Screening (No. of skills after Manual Revision)* ***(p + q + r)***	*461*	*728*	*718*	***1,107***
	o	*Precision of Skills Named Entity Recognition* ***(m/l)***	*67.50%*	*72.51%*	*70.05%*	***73.31%***
	p	Identifying Synonyms exact matching (No. of skills in ESCO)	430	641	650	**995**
	q	Identifying Synonyms similarity matching (No. of skills in ESCO)	22	48	42	**63**
	r	Definition Comparison (No. of skills not in ESCO)	9	39	26	**49**
	s	**Total Number of Skills (Using preferred labels)**	**377**	**559**	**553**	**804**

2.2.1 Green Skills

The top 20 skills extracted from the Green field dataset of papers are summarised in Table 8. The metric of accuracy reported a score of 67.50%. The most important skills retrieved in Green papers are those which can be classified as Soft Skills (such as “think creatively” and “communication”), which are shared with the two other domains as well; moreover, many skills associated with environmental issues (“environmental engineering”, “electricity”, “energy”, “waste management”, “fossil fuels”), but also with economics and innovation (“circular economy” and “innovation processes”) are retrieved.

Table 8. Top 20 relevant skills extracted from Green papers. For each skill, the label (column *Skill*), number of occurrences in the papers (column *Relevance (# Papers)*) and the percentage of Papers containing it (column *% of Papers,* calculated by dividing the number of occurrences of the skill by the total number of Green papers) are shown.

Skill	Relevance (# Papers)	% of Papers	Skill	Relevance (# Papers)	%of Papers
think creatively	2,300	21.60%	waste management	508	4.77%
environmental engineering	1,813	17.03%	bioeconomy	486	4.56%
circular economy	1,806	16.96%	fossil fuels	476	4.47%
innovation processes	716	6.72%	lead others	463	4.35%
packaging engineering	699	6.57%	analyse	360	3.38%
electricity	637	5.98%	mechanical systems	335	3.15%
electricity principles	637	5.98%	characteristics of waste	331	3.11%
energy	587	5.51%	business model	305	2.86%
chemistry	559	5.25%	monitor assessment	300	2.82%

communication	513	4.82%	communication principles	286	2.69%

2.2.2 Digital Skills

The skills extracted from the Digital field dataset of papers are summarised in Table 9. The metric of accuracy reported a score of 72.51%. Similarly to the other domains, among the most important competences it is possible to find Soft Skills. Being in the digital field, most of the other relevant skills are related to the technologies ("computer technology", "printed circuit boards testing methods", "hardware testing methods", "in-circuit test", "computer programming"). There is also a significant presence of pedagogical skills ("e-learning", "pedagogy").

Table 9. Top 20 relevant skills extracted from Digital papers. For each skill, the label (column *Skill*), number of occurrences in the papers (column *Relevance (# Papers)*) and the percentage of Papers containing it (column *% of Papers,* calculated by dividing the number of occurrences of the skill by the total number of Digital papers) are shown.

Skill	Relevance (# Papers)	% of Papers	Skill	Relevance (# Papers)	% of Papers
think creatively	5,445	27.34%	computer programming	844	4.24%
communication	4,243	21.30%	e-learning	748	3.76%
communication principles	3,047	15.30%	pedagogy	740	3.72%
packaging engineering	2,192	11.01%	mathematics	714	3.58%
computer technology	2,037	10.23%	originality	712	3.57%
printed circuit boards testing methods	1,369	6.87%	statistics	711	3.57%
hardware testing methods	1,362	6.84%	computer science	680	3.41%
in-circuit test	1,362	6.84%	lead others	581	2.92%
innovation processes	1,238	6.22%	metrology	549	2.76%
analyze	1,001	5.03%	think critically	534	2.68%

2.2.3 Entrepreneurial Skills

The skills extracted from the Entrepreneurial field dataset of papers are summarised in Table 10. The metric of accuracy reported a score of 70.05%. In the Entrepreneurial field, the knowledge of *entrepreneurship* is clearly the most present skill. Beside some economics and innovations related knowledge ("marketing principles", "statistics", "innovation processes") which can be related to the economics area, all the other most relevant competences appear to be Soft Skills ("think creatively", "originality", "communication principle"), especially related to the leadership aspects ("leadership principles", "lead others").

Table 10. Top 20 relevant skills extracted from Entrepreneurship papers. For each skill, the label (column *Skill*), number of occurrences in the papers (column *Relevance (# Papers)*) and the percentage of Papers containing it (column *% of Papers,* calculated by dividing the number of occurrences of the skill by the total number of Entrepreneurship papers) are shown.

Skill	Relevance (# Papers)	% of Papers	Skill	Relevance (# Papers)	% of Papers
entrepreneurship	13,466	43.33%	lead others	1,263	4.06%

innovation processes	6,633	21.34%	environmental engineering	1,167	3.75%
think creatively	5,850	18.82%	statistics	1,039	3.34%
originality	3,037	9.77%	economics	1,026	3.30%
communication	2,414	7.77%	mechanical systems	1,020	3.28%
packaging engineering	1,771	5.70%	computer technology	887	2.85%
marketing principles	1,713	5.51%	business model	821	2.64%
analyze	1,548	4.98%	establish collaborative relations	745	2.40%
leadership principles	1,520	4.89%	history	718	2.31%
communication principles	1,313	4.22%	collect samples	677	2.18%

2.3 Main conclusion of the Task 2.1

In Task 2.1 we collect and revise a list of skills from scientific papers concerning the Green, Digital, and Entrepreneurship domains, using NLP techniques. We use two methods to identify GDE skills in the abstract of scientific publications: gazetteer-based methods and rule-based methods, that allows us to list 804 unique skills (377 for Green, 559 for Digital, and 553 for Entrepreneurship domain).

The main limitations of our method are three. First, the representativeness of the findings is largely dependent upon the comprehensiveness of the sources analysed and the quality of the query to collect the documents to analyse. Second, increasingly there is no single source of data that can address the requirements of a wide variety of users of labour market skills intelligence, especially in the growing fields such as those which are analysed in our project (Green, Digital and Entrepreneurship). Third, the method provides limited information about how a skill is relevant for a given domain. In fact, we can only count the number of papers in which a skill occurred. This leads us to have some bias in measuring the relevance of a skill for a given domain. For this reason, in task 2.2 the skills will be further validated by companies and other sources (Skills Panorama).

The method can have practical applications for a wide range of users. In the industrial context, it is important to have access to updated information. Our approach can help in this direction, since it provides a quantitative indication of emerging skill needs. In fact, scientific papers are likely to anticipate new technological trends. Moreover, policy makers can use the output and the methodology of the current work package for planning education and training provision. Finally, the system is clearly useful for educators and learners who want to identify the skills and qualifications required to enter an occupation or job, and training providers looking to update their courses. This last group of stakeholders is the one intended to impact in the ENCORE project.

3. Enhancement and revision of the identified skills by the companies

This section presents the methodological note for the task T2.2 ”Enhancement and revision of the identified skills by the companies”. After the extraction of Green, Digital and Entrepreneurial (GDE) skills from scientific literature (T2.1), we aimed at understanding the relevance of the identified competencies and validating the results. To assess this task, we decided to ask organisations to answer a questionnaire, whose questions are designed to make employees list GDE skills that are important for the organisation (without showing them the results of T2.1, to avoid their answers to be biased by our outcomes). The questionnaire is available at the following link: https://forms.gle/67cv3p5sStySLgJv7 and is outlined in Section 3.1.

Furthermore, we assessed the relevance of the extracted skills using the European platform Skills Panorama. Skill Panorama, explained in Section 3.2 allows us to understand the prospects of the GDE skills: an indicator that measures how much a skill will be required in the future.

Finally, Section 3.3 presents and comment the final table obtained after the enhancement and revision of the skills identified in Section 2.

The whole task is synthesised in Figure 4.

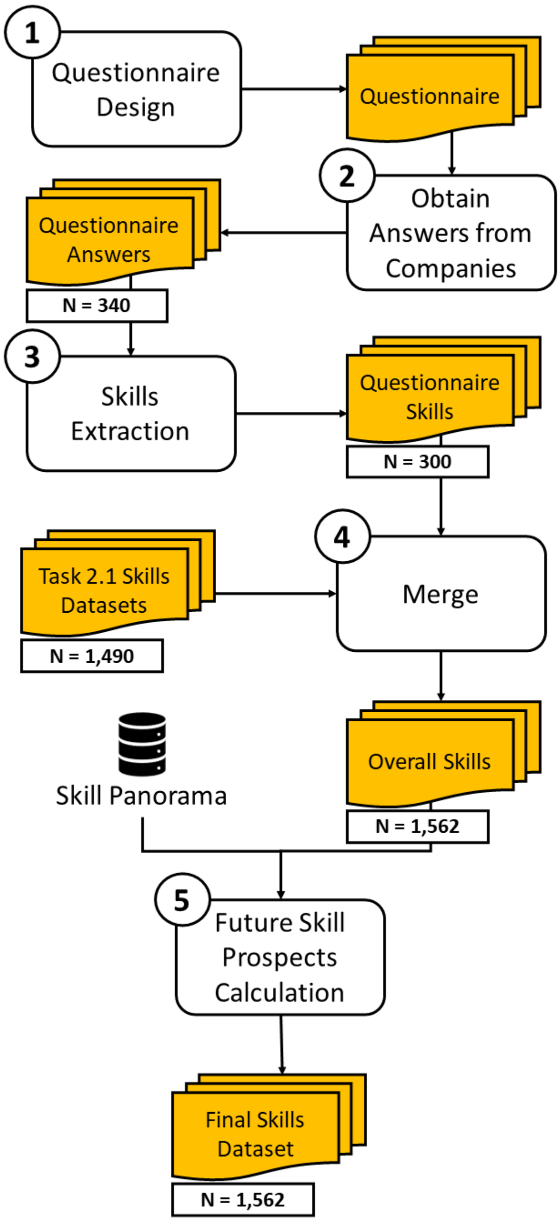

Figure 4. Workflow diagram to describe the steps of Task 2.1.

3.1 Skills Need Questionnaire of ENCORE project

In this section, we firstly explain the structure of the questionnaire together with the design and dissemination process followed (3.1.1). Then we move to outline the methodological workflow we perform for analysing the answers of the questionnaire with Natural Language Processing techniques (3.1.2). Finally, we present which are the most relevant skills for the GDE domains based on the questionnaires (3.1.3).

3.1.1 Questionnaire Design and Structure

The structure of the questionnaire can be logically divided into two main parts. The first focuses on the objective of the survey and includes the most important questions about the skills needed. We name this part as "*Skills Need Part*" in order to facilitate the reading process of the rest of the report. The second and third parts contain information about the workplace where the employees works and personal data about the respondents, respectively.

The questionnaire also includes an explanatory section on the questionnaire itself and further information on confidentiality of the study, as well as a section about the General Data Protection Regulation (GDPR) for protecting the data of the respondent.

The questionnaire is developed in 7 languages in order to allow as many people as possible to answer the questions. This is important in order to get a representative sample of the population, avoiding limiting our population for the language barriers. These 7 languages are the following: Deutsch, English, Español, Français, Italiano, Nederlandse taal, Svenska.

The main section of the questionnaire (i.e., "*Skills Need Part*") is composed of 3 questions per each domain that aim to understand which are the most relevant Green, Digital and Entrepreneurial skills. These questions are:

1. *What skills are you missing the most in your organization?*
2. *What are the skills you are investing the most for upskilling/reskilling?*
3. *What are the skills you are searching for the most when hiring?*

These 3 questions are repeated for each GDE domain.

The rationale behind the three questions is to evaluate the importance of the skills without asking directly for this information. The hypothesis is that the importance of a skill for a company can be driven by the facts that the skills is (1) missing, (2) the company is investing in that skill in terms of upskilling and reskilling and, (3) the company is searching for that skill when hiring.

The questionnaire was launched on 5 September 2022 and closed on 23 October. The dissemination process of the questionnaire was performed with different approaches to reach a bigger audience for the ENCORE questionnaire. The whole consortium worked on the dissemination of the questionnaire using different channels.

First, the questionnaire was shared through Social Networks by the partners. The social networks used are Facebook, Twitter, and LinkedIn. On Facebook the partners posted the questionnaire only on Facebook groups interested in the topic of Human Resource Management. Instead, the partners and its members shared the questionnaire via LinkedIn and Twitter using institutional and personal accounts. In this case, the questionnaire has been posted at least one time for each week in the period of dissemination.

Second, the questionnaire was shared via email and text message by the partners of the ENCORE project. In particular, UNIPI shared the questionnaire with mailing lists of previous ERASMUS+ project where it participated: ULISSE (https://ulisseproject.eu/), SPRINT 4.0 (https://www.sprint40.eu/), ASSETs+ (https://assets-plus.eu/) and E-STEM (http://www.superfastlearning.eu/). Moreover, UNIPI sent an email with the companies of Galileo Aggregator for Technology and Enterprises (https://www.gatecentre.eu/) and Tuscany manufacturing district 4.0 (https://distrettogate40.it/).

FBK shared an ad-hoc email with 27 representatives of medium-large public and private companies working in Human Resources or as digital transformation and innovation experts mainly operating in North Italy (Trentino alto Adige, Lombardy, Veneto, Piemonte, Emilia Romagna) and abroad (Europe, UK, USA). BMU contacted by email all participants in its master's programs: 34 people participating in the Industry 4.0 Design Master's program (https://www.masterindustry40.it/); 45 people participating in the Scalability Master's program (https://masterscalability.it/). Moreover, BMU shared the questionnaire within the Pact For Skills group on Linkedin (https://www.linkedin.com/groups/12664287/). KFRU sent an email to the Alumni club of ESB Business School who has more than 4.000 people (https://www.esb-business-school.de/en), most of them in management positions. Furthermore, KFRU sent the survey to more than 60 companies working in Human Resources. USAL and UNIPD shared the questionnaire with the professors of their university and with companies and research institutions in their network. Myla shared the questionnaire via email with its network composed of about 67,000 companies.

Finally, the questionnaire was also filled by phone call interviews. The calls were made by UNIPI that called more than 300 Italian and European companies.

Each questionnaire, asking about information on green, digital and entrepreneurial skills, can be answered by more than one person of the company (considering their knowledge about the domains). Since we have done a fraction of the questionnaires via phone, we were able to estimate that each questionnaire has been done by a mean of 2 persons (minimum 1 and maximum 4 during the calls).

3.1.2 Questionnaire skills analysis: Data and Methodology

This section presents the method adopted for analysing the result of the questionnaire. As mentioned before, the purpose of this work was to analyse the answers of the ENCORE questionnaires completed by organisations and understand which the most relevant skills were reported for the three areas Green, Digital and Entrepreneurial.

The analysis considers only the "*Skills Need Part*" of the questionnaire, that is directly related to understanding which are the most relevant GDE skills. We provide in Appendix C the distribution of the answers for the questions about workplace and respondent data.

To analyse the "*Skills Need Part*", we needed to convert the skills explained through natural language by companies into a standard form. We matched them with the competencies extracted from scientific papers in task T2.1 and to the ESCO skills list. In Section 3.1.2.1 we show the data used for the questionnaire. In Section 3.1.2.2, we outline the methodology for analysing the collected answers.

3.1.2.1 Data

The input dataset is a spreadsheet containing the answers to the Encore questionnaires done by the companies. The questionnaire was filled by about 680 peoples from companies, reaching a total number of answers of 340.

Each row contains the answers of a single questionnaire completed. There are 134 columns: each of them corresponds to a question of the survey (language choice, country selection, sector of the company, number of employees, job profile of the interviewed, age, gender, education, company email, date and the 9 main questions regarding green, digital, and entrepreneurial skills; this set of question is repeated 7 times, for the 7 languages available in the survey). There are 340 rows, corresponding to the 340 questionnaires completed by the companies. Furthermore, we utilised the dataset produced in task T2.1, concerning the list of skills identified into scientific papers.

3.1.2.2 Methodology

This analysis aimed at comparing the list of distinct skills obtained from the answers given to the questionnaire with the skills extracted in the task T2.1, obtained from scientific papers. This required the application of a semantic similarity measurement through text embeddings and a manual revision.

3.1.2.2.1 Data pre-processing

The first operations executed were aimed at preparing the dataset for the analysis. First, we removed the separation between the seven languages: for each question of the questionnaire, we merged the content of the columns containing the answers to the question in the different languages into a single column. The number of columns decreased to 21.

Then, we created a further version of this dataset, focused on skills: we merged the content of the nine columns containing the answers to the skill questions into a single column, and we separated each skill contained in the answers using items separators, such as ",", "and", "e", "und", and so on. In this manner, we obtained a dataset where each row records one skill inserted as answer to the questions of the survey into the column *skill*, and still maintain all the other parameters related to the specific questionnaire each skill belongs to.

Table 11. Sample of the dataset: each skill of each questionnaire is recorded in a row.

Questionnaire index	Skill	Language	Country	...
1	Artificial intelligence	Italiano	Italia	...
1	Programmazione	Italiano	Italia	...
1	Autonomia	Italiano	Italia	...

Table 11 presents a sample of the dataset obtained, which contains 5,450 rows. Then, we applied a text cleaning step to improve the list of skills contained in the *skill* column of the dataset. Since many answers of the survey were not actual skills, we produced a list of noisy elements to be deleted (some examples are "I don't know", "No one", "No") to reduce the amount of non-relevant entities registered in the dataset. Furthermore, we deleted missing values, extra/multiple spaces, and some other specific symbol from the skills list. The number of skills decreased from 5,450 to 5,132 after the cleaning. However, some skill occurred in more answers, for this reason, we removed the duplicates, since they would have been useless in the automatic skill similarity step. The final dataset contained 1,492 unique entities.

3.1.2.2.2 Automatic skill similarity

Once we had the list of 1,492 unique skills to the questionnaire, we performed a semantic similarity analysis to try to match them to our standard skills, i.e., the GDE skills resulted from Section 2, collected from scientific publications (804 unique skills as we can see from Table 7).

We executed this task by using *SentenceTransformer,* a Python framework for text embeddings. Since the answers to the questionnaire were in seven languages, we needed a model able to map similar inputs written in different languages close in vector space. We used the model *distiluse-base-multilingual-cased-v2*[14], which supports 50+ languages. Both the 1,492 questionnaire skills and the 804 scientific papers' skills have been mapped to a 512-dimensional dense vector space and the similarity has been calculated between each pair using cosine similarity algorithm. We obtained a table demonstrating, for each questionnaire skill, the semantic similarity with every single scientific paper's skill. In this way, we could identify to which standard skill they could have been matched.

3.1.2.2.3 Manually revision of skill similarity

We decided to evaluate manually if each questionnaire skill could have been matched to the scientific papers' skills. We selected, for each questionnaire skill, the four most similar scientific papers' ones. By reading them, we decided if one of the four identified skills was a good match, or if no one of them

[14] https://huggingface.co/sentence-transformers/distiluse-base-multilingual-cased-v2. Accessed on Oct. 25, 2022.

would have been correct. For some of the skills with no good matches, we identified a similar skill on the ESCO database: we added them to the list of matched skill, recording the corresponding ESCO label as well. Furthermore, after producing the list of entities with no correct matches, we evaluated if any of them could have been considered a new skill, because currently absent on standard classifications such as ESCO.

3.1.2.2.4 Measuring the relevance of GDE skills

Finally, as last task of the methodology, we measured the relevance of the GDE skills resulting from the questionnaire. The relevance of a skill for the GDE domains has been calculated based on the occurrences of the skill for the questions related to the "*Skills Need Part*". We considered the number of occurrences as a metric for the relevance of the skills, since our hypothesis is that the more a certain skill has been cited in the questionnaires, the more it is relevant for the organisations. We have three different questions in "*Skills Need Part*" (see Section 3.1.1 for more details) for each domain (i.e., 3 for the Green domain, 3 for the Digital one, and 3 for the Entrepreneurship one). A respondent can list the same skill for each one of these questions in the same answer. This can happen because the skill can be at the same time the most missing skill, the most required for upskilling/reskilling, the most searched in the hiring process. For this reason, we calculated the relevance of a skill for a given domain as the sum of occurrences in the three questions.

3.1.3 Results of enhanced and revised Skills from questionnaire

After the manual revision of the skill similarity, we obtained the results of the analysis. Among the 1,492 distinct questionnaire skills, we had:

- 1,184 skills already found in Section 2 or matched to an ESCO label;
- 305 skills removed, since that they were related to: (1) concepts that were not related with the question (e.g., *Internet instead of paper*); (2) too general concepts (e.g., *Agility*, *Moderation*); sentences expressing that the respondent did not know how to answer (e.g., *I don't know*, *Nothing*);
- 3 skills evaluated as potential new skills that should be taken into consideration in task 2.3 "*skill2ESCO*": *Industry 4.0; Software development; Docker*.

The final list of skills was composed of the 1,184 skills matched plus the 3 potential new skills. By considering the label of the skills listed in Section 2, these 1,187 entities corresponded to 300 unique competencies (many questionnaire answers have been matched to the same skill, since they were the same concept described with different words or in another language). Considering the three different areas, Table 12 shows the number of unique skills for each domain. The sum of the three numbers exceeds the total of 300 because several skills have been found in more than one domain.

Table 12. Number of unique skills (matched with the Scientific Paper's ones) found in the answers of the questionnaires for each Domain and overall.

Domain	Number of unique skills matched with the Scientific Papers' ones
Green	128
Digital	129
Entrepreneurial	142
Overall	300

Table 13, Table 14, and Table 15 present the 20 most relevant skills of the three areas Green, Digital and Entrepreneurship, respectively.

Table 13. The 20 Green skills with the highest occurrences in the questionnaires answered by companies. For each skill, the label (column *Skill*), the number of occurrences (column *Relevance for questionnaire,* being the sum of the occurrences of the skill in the three Green-related questions) and the percentage of questionnaires where the skill has been found at least once (column *% of Questionnaires*) are reported.

Green Skill	Relevance for questionnaire	% of Questionnaires
waste management	190	44.41%
circular economy	159	35.88%
energy efficiency	152	30%
bioeconomy	103	24.71%
ecological principles	51	12.35%
ecology	25	5.59%
manage waste	25	6.47%
promote sustainability	24	5%
electricity consumption	23	5%
renewable energy technologies	18	4.71%
energy	13	3.24%
climate change impact	9	1.76%
solar energy	9	1.76%
assume responsibility	8	4.12%
environmental threats	8	1.76%
work efficiently	8	2.35%
manage routine waste	7	1.76%
security regulations	6	3.24%
sustainable development goals	6	1.47%
water reuse	6	1.47%

Table 14. The 20 Digital skills with the highest occurrences in the questionnaires answered by companies. For each skill, the label (column *Skill*), the number of occurrences (column *Relevance for questionnaire,* being the sum of the occurrences of the skill in the three Digital-related questions) and the percentage of questionnaires where the skill has been found at least once (column *% of Questionnaires*) are reported.

Digital Skill	Relevance for questionnaire	% of Questionnaires
analyse big data	198	39.71%
principles of artificial intelligence	191	36.76%
cyber security	162	34.12%
computer programming	121	26.76%
data analytics	65	15.88%
machine learning	47	10.59%

data mining	42	9.12%
Python (computer programming)	42	8.53%
business intelligence	32	10%
operate digital hardware	30	7.35%
project management	22	10%
manage data	15	3.82%
cloud technologies	14	2.35%
analyze	13	3.53%
analyse scientific data	12	2.65%
perform data analysis	12	2.35%
automation technology	7	1.47%
Internet of Things	7	1.18%
security regulations	7	3.24%
building automation	6	0.88%

Table 15. The 20 Entrepreneurial skills with the highest occurrences in the questionnaires answered by companies. For each skill, the label (column *Skill*), the number of occurrences (column *Relevance for questionnaire,* being the sum of the occurrences of the skill in the three Entrepreneurial-related questions) and the percentage of questionnaires where the skill has been found at least once (column *% of Questionnaires*) are reported.

Entrepreneurial Skill	Relevance for questionnaire	% of Questionnaires
leadership principles	228	45,88%
work in teams	166	32,35%
think creatively	150	30,88%
negotiate compromises	71	16,18%
marketing principles	69	17,35%
communication	42	9,41%
meet commitments	35	7,35%
moderate in negotiations	23	5%
project management	19	10%
involvement	16	3,53%
think critically	14	2,65%
motivate employees	13	3,24%
assertiveness	11	2,94%
personnel management	11	2,65%
entrepreneurship	10	1,47%
international trade	10	3,24%
strategic planning	10	2,35%
assume responsibility	9	4,12%
independence	9	2,35%

solve problems	9	2,65%

3.2 Skills Panorama

The aim of this section is to assess the relevance of the skills collected in T2.1 (updated after the steps of Section 3.1) using the European platform Skills Panorama[15]. This will allow us to have a data-driven assessment of skills, to present together with the results coming from the questionnaire. This section includes a description of the approaches used and the obtained results.

3.2.1 Methodology for measuring the Skill Future Prospect

European platform Skills Panorama is an interactive web portal offering skills intelligence on countries, occupations, and sectors in the European Union. It is an initiative of CEDEFOP, the European Centre for the Development of Vocational Training. The design of the platform started in 2015. Skills Panorama provides quantitative and qualitative data and insights from the analysis of job vacancies, which can support research studies on skills and relevant stakeholders in the labour market and in the education area, such as policymakers, learning providers or career practitioners, with data-driven information on employment and jobs. Skills Panorama include 56 indicators, 23 sectors, 48 occupations and 28 countries. Currently, there are also 26 skills dashboards and 89 curated data insights. The approach focuses on the perspective of the labour market, so it is centred on the concept of occupation, which is "a set of jobs with similar tasks and duties as well as qualification and skills requirements". The databases and the dashboards are structured on the ISCO-08 classification[16].

For the purpose of our analysis, we selected the Future Job Prospects indicator, that is a comparison between future job openings and the current employment in each occupation. It is based on the forecast of future job opportunities, i.e., Cedefop Skills Forecast[17], which provides information on the future labour market trends in Europe. The score is considered high if it is above 60, meaning that there will be high job prospects, and so more job opportunities can be opened for the given occupation compared to the ongoing situation. The score is considered low if it is below 40, meaning that there will be low job prospects, and so less job opportunities for the given occupation compared to the ongoing situation. A score between 40-60 means that the number of job openings is closer but below the current employment[18].

We are interested in measuring the future need of a skill, instead of an indicator on the occupation. For this reason, we developed a methodology for transferring the information from the occupation perspective (as provided by Skills Panorama) to the skill point of view. The final output of the methodology is the index we call "Future Skill Prospect".

We first prepared the data for calculating the "Future Skill Prospect". We defined a table for joining occupations of ESCO with those of Skill Panorama. Even though both datasets are based on ISCO-08 classification, some of the labels are slightly different. The crosswalk table between ESCO and Skills Panorama is available in the Appendix D of this document.

Then we joined our list of GDE skills included in ESCO with the groups of occupation. For each skill, ESCO provides a list of occupations which require a given skill to properly perform the typical activities

[15] https://skillspanorama.cedefop.europa.eu/

[16] https://www.cedefop.europa.eu/en/news/cedefop-skills-intelligence-skills-panoramas-new-home. Accessed on Oct. 25, 2022.

[17] https://www.cedefop.europa.eu/en/tools/skills-intelligence/datasets#cedefop-skills-forecast. Accessed on Oct. 25, 2022.

[18] https://www.cedefop.europa.eu/en/tools/skills-intelligence/future-job-prospects?country=EU27_2020&year=2020-2030#1. Accessed on Oct. 25, 2022.

of that job. Those connections can be essential or optional. We used these fine-grain connections between skills and occupations of ESCO, considering only the links tagged as essential.

However, the skills and knowledge considered transversal for the ESCO framework were not directly related to any ESCO occupation, since that they are related to all occupations. For this reason, these skills were considered as connected to all occupations of ESCO for the next steps of the methodology.

The occupations in ESCO are organised in a hierarchy following the ISCO-08 classification. A four-digit code is associated with each occupation. We climbed up the hierarchy to reach the level of the two-digit code and then the one-digit code. This is due to the need to make ESCO and Skills Panorama “talk” the same language, and so have the same level of detail.

Finally, we customised the Future Job Prospects indicator because our perspective focuses on competences rather than occupations, as explained above. Specifically, we counted the number of occupations of ESCO associated with a given skill (referring to the four-digit code) in each group of higher level (i.e., two-digit code and one-digit code). We refer to one-digit codes as *Level 1* and two-digit codes as *Level 2*. Then, we computed for each skill the following measure, called Future Skill Prospects:

$$Future\ Skill\ Prospects\ =\ \frac{\Sigma\left(Future\ Job\ Prospects_{Level\ 2} * Number\ of\ ESCO\ Occupation_{Level\ 2}\right)}{\left(Future\ Job\ Prospects_{Level\ 1} * Number\ of\ ESCO\ Occupation_{Level\ 1}\right)}$$

The scores follow the same pattern of Future Job Prospects. Therefore, it varies between 0 and 100. The score is considered high if it is above 60, meaning that the skills will be in high demand for the future compared to the ongoing situation in the labour market. The score is considered low if it is below 40, meaning that the skill will have lower relevance in the labour market compared to the ongoing situation. A score between 40-60 means that the demand for a given skill is closer but below the current one.

We finally computed the average value of Future Skill Prospects to provide a unique indicator for each skill.

A sample of the results is reported in Table 16 for better explaining how the Future Skill Prospect index was calculated. In Table 16, the column *Label* reports the skills, the second column is the *Future Skill Prospect*, column *Details* Future Job Prospect *(FJP)* reports the value of Future Job Prospects for each of the occupations groups associated with a given skills (i.e., the numerator of the Future Skill Prospect formula), and finally the *Number of different occupation groups* reports the total number of occupation groups detailed in the previous column (i.e., the denominator of the Future Skill Prospect formula).

Table 16. Sample of the results of the Skills Panorama Analysis.

Label	Future Skill Prospect	Details FJP	Number of different occupation groups
communication	75,25	Legal & social associate professionals: 100; Office professionals: 62; Legal & social professionals: 62; Customer clerks: 77	4
communication principles	68,38	Hospitality & retail managers: 67; Legal & social professionals: 61; Office professionals: 61; Legal & social associate professionals: 78.5; Business managers: 67; Technical managers: 67; Office associate professionals: 78.5; CEOs, officials & legislators: 67	8
packaging engineering	63	Researchers & engineers: 63	1
computer technology	51,45	Teaching professionals: 55.67; Science & engineering technicians: 43; Researchers & engineers: 55.67	3
capacity building	87	Legal & social associate professionals: 100; Business managers: 74	2
personal development	82	Legal & social associate professionals: 100; Health professionals: 73; Legal & social professionals: 73	3
manage feedback	78,24	Legal & social professionals: 67.36; Office professionals: 67.36; Legal & social associate professionals: 100	3
radiology	78	Health professionals: 78	1

3.2.2 Future Skill Prospect Results

This section includes the results of the analysis. The results are available online[19].

The skills with the highest *Future Skill Prospect* resulted **Capacity Building[20]**, which refers to human resources management and development, it includes processes and methods to develop skills and knowledge within the organisation, promoting knowledge sharing and training activities and developing managerial procedures. Therefore, **Capacity Building** will be paramount to properly face the challenges of the current fast-evolving context, where the pace of technological changes requires employees and managers with a flexible mindset.

Many of the skills included in this analysis reach a value of *Future Skill Prospect* around 50 (indeed the mean value of this indicator is 53,20). Therefore, the demand for those skills will be quite stable in the labour market. However, with a closer look at the distribution of the *Future Skill Prospect*, reported Figure 5, some interesting insights emerge. In particular, 199 skills out of the 635 considered in this analysis, register a *Future Skill Prospect* above 60, meaning that those skills will be in high demand for the future. Those are mainly knowledge (163), such as *personal development*, *data analytics* and *sustainable development goals*, then skills (only 56), like *manage feedback*, *analyse big data* and *develop food waste reduction strategies*.

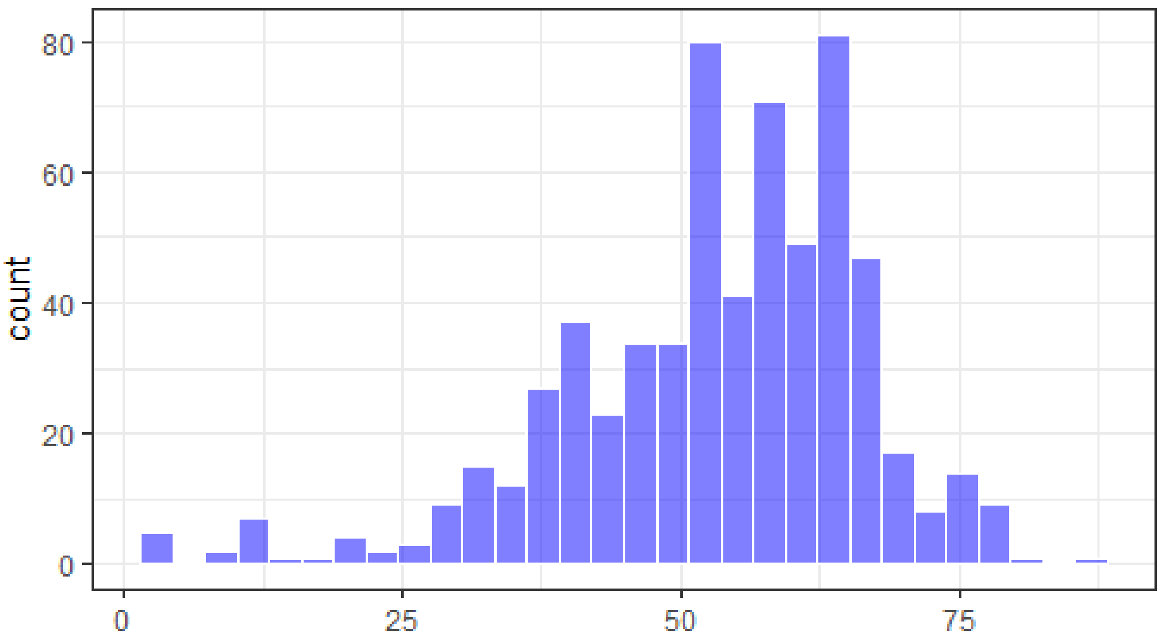

Figure 5. Histogram of the Future Skills Prospects, as resulted in the Skills Panorama Analysis.

On average, the skills are related to 3.33 occupations groups, proving the transversality of the Green, Digital and Entrepreneurial competences in the labour market. We have also analysed the relationship between the number of different skills and the number of different occupation groups which require a given competence. For example, the skill *perform project management* is linked to 13 different groups, following listed with the *Future Skill Prospect* in brackets: Researchers & engineers (63.15), Legal & social professionals: (63.15), Health professionals (63.15), Hospitality & retail managers (63.58), Technical managers (63.58), ICT professionals (63.15), Office professionals (63.15), Business managers (63.58), Teaching professionals (63.15), CEOs, officials & legislators (63.58), Office associate professionals (67.75), Legal & social associate professionals (67.75), Sales workers (33). Another

[19] https://docs.google.com/spreadsheets/d/1b67q4O8XEcH2SNlhq9dZ0ZLWLcg094uev-vTkW7_dI0/edit#gid=0
[20] http://data.europa.eu/esco/skill/c28553f4-1745-401a-82b3-7ce8460bed33. Accessed on Oct. 25, 2022.

example is the knowledge *marketing principles* which has been found in 6 occupations groups, namely Business managers (70), Office professionals (58.64), Researchers & engineers (58.64), Teaching professionals (58.64), Legal & social professionals (58.64), and Technical managers: (70). However, some competences appear very specific, such as *radiology* and *biomedicine* linked only to Health professionals, both with a *Future Skill Prospect* equal to 78; or *conflict management* linked only to Business managers with a *Future Skill Prospect* equal to 74.

Finally, the top 20 skills for the 3 areas in analysis, Green, Digital and Entrepreneurship are presented in the following tables (Table 17, Table 18, and Table 19, respectively).

Table 17. Top 20 skills in the Digital domain based on Future Skills Prospects, as resulted in the Skills Panorama Analysis.

Digital Skill	Future Skill Prospect
capacity building	87
personal development	82
manage feedback	78.24
radiology	78
psychotherapy principles	78
biomedicine	78
writing techniques	77.6
use questioning techniques	77
pathology	75.27
communication	75.25
data analytics	74
sustainable development goals	74
social innovation	74
alter management	74
conflict management	74
management department processes	74
crowdfunding	74
social entrepreneurship	74
manage several projects	74
outplacement	74

Table 18. Top 20 skills in the Entrepreneurship domain based on Future Skills Prospects, as resulted in the Skills Panorama Analysis.

Entrepreneurial Skill	Future Skill Prospect
capacity building	87
personal development	82
manage feedback	78.24
radiology	78
psychotherapy principles	78
biomedicine	78
use questioning techniques	77
ecotourism	77
use computer telephony integration	77
pathology	75.27
communication	75.25
data analytics	74
sustainable development goals	74
social innovation	74
alter management	74
conflict management	74
management department processes	74
crowdfunding	74
social entrepreneurship	74
manage several projects	74

Table 19. Top 20 skills in the Green domain based on Future Skills Prospects, as resulted in the Skills Panorama Analysis.

Green Skill	Future Skill Prospect
capacity building	87
organic chemistry	78
ecotourism	77
communication	75.25
data analytics	74
sustainable development goals	74
social innovation	74
alter management	74
management department processes	74
pharmaceutical industry	73
establish collaborative relations	72.95
sociology	72.5
social sciences	70.7
social justice	70.06
develop food waste reduction strategies	69.75
toxicology	69.67
provide information	69.16
promote sustainability	68.5
climate change impact	68.5
analyse score	68.48

3.3 Table of the GDE skills

The list of GDE skills obtained in Section 2 is enhanced and revised in Section 3. The final output is a table that reports in each domain the skills obtained in the report with three main indicators of relevance:

- **Relevance for Papers**: the indicator considers the number of papers in which the skill occurred. The indicator is calculated by rescaling the number of papers values, for avoiding distorting differences in the ranges of values. We applied min max normalisation and multiplied the obtained value by 100 to ensure that all data falls within the range of 0-100. Moreover, we used the logarithmic of the number of papers in the min max normalisation in order to compress a large range of values into a smaller scale, since that the number of papers spans multiple orders of magnitude.

- **Relevance for Questionnaire**: the indicator considers the relevance index calculated in Section 3.1 (see Table 13, 14 and 15). The indicator is calculated by rescaling the index values, for avoiding distorting differences in the ranges of values. We applied min max normalisation and multiplied the obtained value by 100 to ensure that all data falls within the range of 0-100.
- **Future Skill Prospect**: the indicator is explained in Section 3.2.

The results are available online[21]. The table contains in each row a GDE skill and the following column:

- **Label**: the preferred label used for the skill (character variable);
- **Domain**: the domain where the skill was found (character variable);
- **Relevance for Papers**: the indicator measures the relevance of the skill taking into account the number of papers in which the skill occurred (double value, ranging from 0 to 100);
- **Relevance for Questionnaire**: the indicator measures the relevance of the skill taking into account the index calculated using questionnaire (double value, ranging from 0 to 100);
- **Future Skill Prospect**: the indicator measures the relevance of the skill taking into account the future prospect of the skill (double value, ranging from 0 to 100);
- **Relevance Mean:** the average between the three relevance measures, reported in Relevance for Papers, Relevance for Questionnaire and Future Skill Prospect;
- **Found in Paper**: a binary variable for indicating whether the skill was found in paper (binary variable);
- **Found in Questionnaire**: a binary variable for indicating whether the skill was found in questionnaire (binary variable);
- **ESCO**: a binary variable for indicating whether the skill was found using Named Entity Recognition method based on ESCO gazette (binary variable);
- **O*NET**: a binary variable for indicating whether the skill was found using Named Entity Recognition method based on O*NET gazette (binary variable);
- **EntreComp**: a binary variable for indicating whether the skill was found using Named Entity Recognition method based on EntreComp gazette (binary variable);
- **Rule-based Extraction**: a binary variable for indicating whether the skill was found using Named Entity Recognition method based on rule-based method gazette (binary variable);
- **Alternative Labels**: the alternative labels used for searching skills with Named Entity Recognition methods or extracted from the questionnaire. Each alternative label is separated by a semicolon (character variable).

The results, which will be commented and analysed in the next steps of the project, are particularly interesting. It emerges a synthetic view on the three domains, with many skills in common and a strong focus on soft skills. To have a bird eye view on the results, Table 20 shows the top 10 skills in terms of mean between the three indicators we computed in the present report. As it is evident, there are 4 green, 3 entrepreneurial and 3 digital skills. This quasi-equal distribution shows the relevance of all the three sectors, that we will bring on the rest of the project. It is interesting to notice that among the top-5 skills there is a mix of vertical skills, such as *waste management* and *circular economy* for green, and *principles of AI* for digital, and of soft-skills, such as *think creatively* and *leaderhip principles*. The other 5 skills are more technical and go from *energy efficiency* to *computer programming*. The fact that the results are in line with well-known trends in the sector, is evidence of the reliability of the adopted method. Given these results, we look forward to cross these information with the outcomes coming

[21] https://docs.google.com/spreadsheets/d/1b67q4O8XEcH2SNlhq9dZ0ZLWLcg094uev-vTkW7_dI0/edit#gid=0

from educational data, i.e. OER, to give our contribution to the pedagogical community on such a relevant set of skills.

Table 20. Top 10 skills of the dataset with the respect to the Relevance Mean column.

Label	Domain	Relevance for Papers	Relevance for Questionnaire	Future Skill Prospect	Relevance Mean
waste management	Green	80,51	100	52,74	77,75
think creatively	Entrepreneurship	91,23	65,79	75,51	77,51
leadership principles	Entrepreneurship	77,06	100	51,3	76,12
circular economy	Green	96,88	83,68	45,75	75,43
principles of artificial intelligence	Digital	70,21	96,46	52,5	73,05
energy efficiency	Green	71,53	80	59,67	70,4
work in teams	Entrepreneurship	56,04	72,81	75,51	68,12
analyse big data	Digital	32,23	100	68,33	66,85
bioeconomy	Green	79,94	54,21	63	65,71
computer programming	Digital	78,34	61,11	44,11	61,18

4. Conclusion

The tasks described in this document, T2.1 and T2.2, had the aim of listing and evaluating the Green, Digital and Entrepreneurial (GDE) skills and knowledge that will be relevant for the rest of the project.

Table 21 contains the indicators we aimed to achieve, and the results obtained in the actual implementation of the activities. Thanks to the use of data driven methods, many of the indicators were largely reached.

Table 21. Indicators of objectives to be achieved within the tasks T2.1 and T2.2 and actual results obtained.

Quantitative indicators	Target Value	Achieved Results
N. of green, digital and entrepreneurial skills	75	804
N. of people from companies involved in the validation	500	680
N. of validated green, digital and entrepreneurial skills	60	300

What is relevant of this task is (besides the results that will feed the rest of project, indicating also the scope on which to focus during the rest of the activities) the developed methods. ENCORE has the aim of mixing quantitative/data-driven with qualitative/expert driven approaches, and this report is a first step in this direction. In the vision of the project, this will be the way to contribute to the teaching and recognition of the most important skills in relation to the new trends in digitalization, climate change and post-COVID economic recovery challenges. For the quantitative approaches we rely on Natural Language Processing techniques, to properly label the most relevant skills for the project (according to ESCO) and to link the skill to the OER of the ENCORE database.

In the next WP (WP3) we will analyze the text of the OER (e.g., description of lessons, content of books, transcripts of video if available) and use Natural Language Processing Tools to find OER linked to the specific skills identified in this report. The database will be designed to support the storage and indexation of the variety of types of OER that can be found in the repositories. To do so, the partnership will use a method already developed by the University of Pisa and the University of Salamanca. Next, in WP4 we will furtherly exploit the ENCORE database thanks to NLP: FBK will conduct a textual analysis of the OER to extract the concepts (i.e., building blocks of a lesson), using as a tool Wikipedia, the free encyclopedia. These building blocks will be used to automatically synthesize HyFlex/hybrid pedagogical courses. Also, WP4 is expected to leverage quantitative approaches, to create an interface between the OER database and educators/learners (ENCORE enablers). FBK will use available AI and Gamification techniques to support teachers in the process of defining and monitoring dedicated courses for their students.

The solutions described in this report are expected to facilitate learning, encourage motivation and engagement, improve students' participation and cooperation, and stimulate students to reach their GDE skills.

References

Chiarello, F., Fantoni, G., Hogarth, T., Giordano, V., Baltina, L., & Spada, I. (2021). Towards ESCO 4.0–Is the European classification of skills in line with Industry 4.0? A text mining approach. *Technological Forecasting and Social Change*, 173, 121177.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Fareri, S., Melluso, N., Chiarello, F., & Fantoni, G. (2021). SkillNER: Mining and mapping soft skills from any text. *Expert Systems with Applications*, 184, 115544.

Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. Journal of emerging technologies in web intelligence, 1 (1), 60-76.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. *Elsevier*.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. *In COLING 1992 Volume 2: The 15th International Conference on Computational Linguistics*.

Honnibal, M., and Montani, I. (2017). “spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing”. *In-press.*

Hu, K., Wu, H., Qi, K., Yu, J., Yang, S., Yu, T., ... & Liu, B. (2018). A domain keyword analysis approach extending Term Frequency-Keyword Active Index with Google Word2Vec model. *Scientometrics*, 114(3), 1031-1068.

Jiang, M., Chen, Y., Liu, M., Rosenbloom, S. T., Mani, S., Denny, J. C., and Xu, H. (2011). A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. *Journal of the American Medical Informatics Association*, 18(5):601–606.

Nadeau, D. and Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.

Pawar, S., Srivastava, R., and Palshikar, G. K. (2012). Automatic gazette creation for named entity recognition and application to resume processing. *In Proceedings of the 5th ACM COMPUTE Conference: Intelligent & scalable system technologies*, pages 1–7.

Roller, S., Kiela, D., and Nickel, M. (2018). Hearst patterns revisited: Automatic hypernym detection from large text corpora. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers),* pages 358–363, Melbourne, Australia. Association for Computational Linguistics.

Acknowledgements

The ENCORE project partners would like to acknowledge all the associate stakeholders for the scientific contribute given in the design of the activities described in this project. Also, we would like to thank Irene Spada, for the fruitful scientific contribution on themes of education and skills.

Appendix A - European Policies

Table A1 presents the complete list of the European policies and directives taken into consideration for designing the query related to the green concept.

Table A1. List of the European policies and directives considered for designing the query related to the Green field.

European Policy and Directive		Reference Link
Name	**Description**	
Ecodesign Directive	Directive 2009/125/EC of the European Parliament and of the Council of 21 October 2009 establishing a framework for the setting of ecodesign requirements for energy-related products, OJ L 285, 31.10.2009, p. 10	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32009L0125&from=it#d1e527-10-1
Ecolabel	Regulation (EC) No 66/2010 of the European Parliament and of the Council of 25 November 2009 on the EU Ecolabel, OJ L 27, 30.1.2010, p. 1.	https://ec.europa.eu/environment/ecolabel/eu-ecolabel-for-consumers.html
GPP Criteria	National Green Public Procurement Guidelines	https://ec.europa.eu/environment/gpp/eu_gpp_criteria_en.htm
Smart Circular Applications	COM (2020) 67 final.	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0067&from=IT
IPPC Directive	Directive 2010/75/EU of the European Parliament and of the Council of 24 November 2010 on industrial emissions (integrated pollution prevention and control),OJ L 334, 17.12.2010, p. 17.	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32010L0075&from=it
Bioeconomy action plan	A sustainable bioeconomy for Europe Strengthening the connection between economy, society	https://op.europa.eu/en/publication-detail/-/publication/edace3e3-e189-11e8-b690-01aa75ed71a1/language-en https://op.europa.eu/en/publication-detail/-/publication/775a2dc7-2a8b-11e9-8d04-01aa75ed71a1

	and the environment : updated bioeconomy strategy	
Batteries Directive	Directive 2006/66/EC of the European Parliament and of the Council of 6 September 2006 on batteries and accumulators and waste batteries and accumulators and repealing Directive 91/157/EEC, OJ L 266, 26.9.2006, p. 1.	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32006L0066&from=it
Vehicle Directive	Directive 2000/53/EC of the European Parliament and of the Council of 18 September 2000 on end-of life vehicles, OJ L 269, 21.10.2000, p. 34.	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02000L0053-20200306, https://ec.europa.eu/environment/topics/plastics_en
Packaging and Packaging waste directive	European Parliament and Council Directive 94/62/EC of 20 December 1994 on packaging and packaging waste, OJ L 365 31.12.1994, p. 10	
Single-use Plastic products Directive	Directive (EU) 2019/904 of the European Parliament and of the Council of 5 June 2019 on the reduction of the impact of certain plastic products on the environment, OJ L 155, 12.6.2019, p. 1	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L0904&from=EN#d1e32-17-1
Plastic Strategy	COM (2018) 28 final	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0028
Construction and demolition	Regulation (EU) No 305/2011 of the European Parliament and of the Council of 9 March 2011 laying down harmonised conditions for the marketing of construction products and	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02011R0305-20140616

	repealing Council Directive 89/106/EEC, OJ L 88, 4.4.2011, p. 5.	
Social Europe for Just transition	COM(2020) 14 final	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0014
EU Taxonomy Regulation	Regulation (EU) 2020/852	https://ec.europa.eu/commission/presscorner/detail/en/fs_20_39
Circular economy action plan	COM(2020) 98 final	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0098&from=EN#footnote45
Waste	Directive 2008/98/EC	https://ec.europa.eu/environment/topics/waste-and-recycling_en
Accessibility of certain products and services	Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services (Text with EEA relevance)	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019L0882

Appendix B - Queries

Table B1 provides the description of each constituent part of the query used to retrieve digital papers.

Table B1. Deployment of the elements of the query related to Digital.

Query Part	Description
`TITLE-ABS-KEY(digital* or "industry 4.0")`	Area/domain
`TITLE-ABS-KEY("skill*" OR "competence*" OR "knowledge" OR "capabilit*")`	Information
`TITLE-ABS-KEY("human resource*" OR "HR" OR worker OR workforce OR "student*" OR "citizen*" )`	Users
`( LIMIT-TO ( SUBJAREA,"COMP" ) OR LIMIT-TO ( SUBJAREA,"ENGI" ) OR LIMIT-TO ( SUBJAREA,"BUSI" ) OR LIMIT-TO ( SUBJAREA,"DECI" ) OR LIMIT-TO ( SUBJAREA,"SOCI" ) )`	Knowledge field (limit to relevant areas)
`(EXCLUDE (SUBJAREA , "MATH") OR EXCLUDE (SUBJAREA , "PSYC") OR EXCLUDE (SUBJAREA , "MEDI") OR EXCLUDE (SUBJAREA , "PHYS") OR EXCLUDE (SUBJAREA , "EART") OR EXCLUDE (SUBJAREA , "CENG") OR EXCLUDE (SUBJAREA , "NURS") OR EXCLUDE (SUBJAREA , "AGRI") OR EXCLUDE (SUBJAREA , "CHEM") OR EXCLUDE (SUBJAREA , "BIOC") OR EXCLUDE (SUBJAREA , "DENT") OR EXCLUDE (SUBJAREA , "PHAR") OR EXCLUDE (SUBJAREA , "NEUR") OR EXCLUDE (SUBJAREA , "VETE") OR EXCLUDE (SUBJAREA , "IMMU"))`	Knowledge fied (there could be publications in more than one field; exclude not relevant field)

Appendix C - Questionnaire Metadata Analysis

For each question of metadata, more specifically of modules "Workplace General Data" and "Respondent General Data", the results have been analysed and reported in the following sections.

Workplace General Data

As mentioned above this part of the questionnaire includes information about the country and the sector of the companies, and the number of employees. The following figures (bar plots) have been produced while analysing the data.

Figure C1 shows the diverse allocation of the companies. The highest number of responses came from companies located in Spain and Italy (134 and 85 responses respectively).

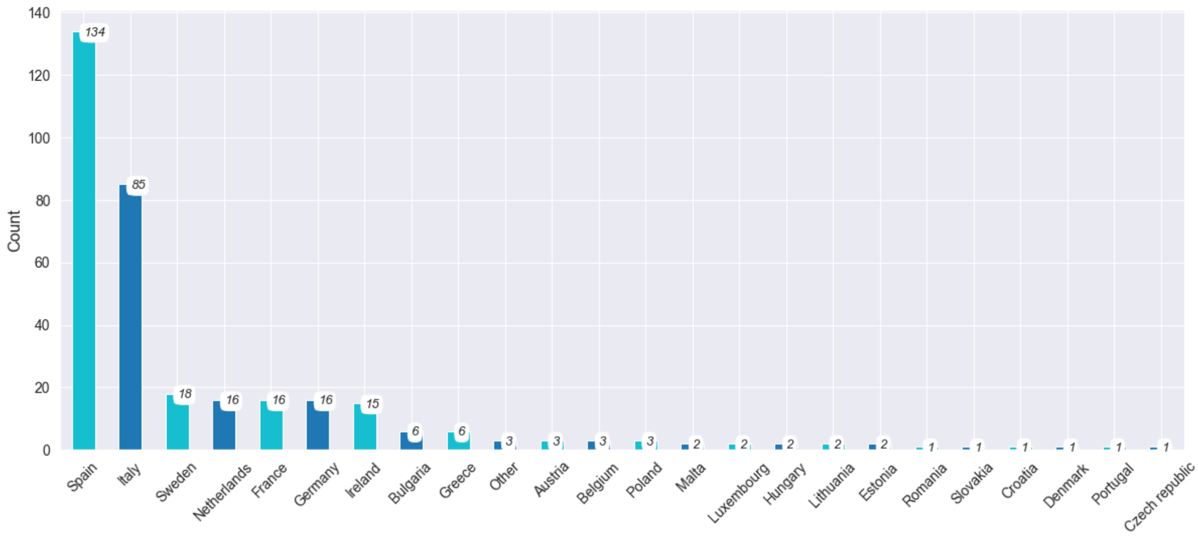

Figure C1: Countries where the companies are located.

Figure C2 shows the diversity of the business sectors. Education sector showed the highest participation rate.

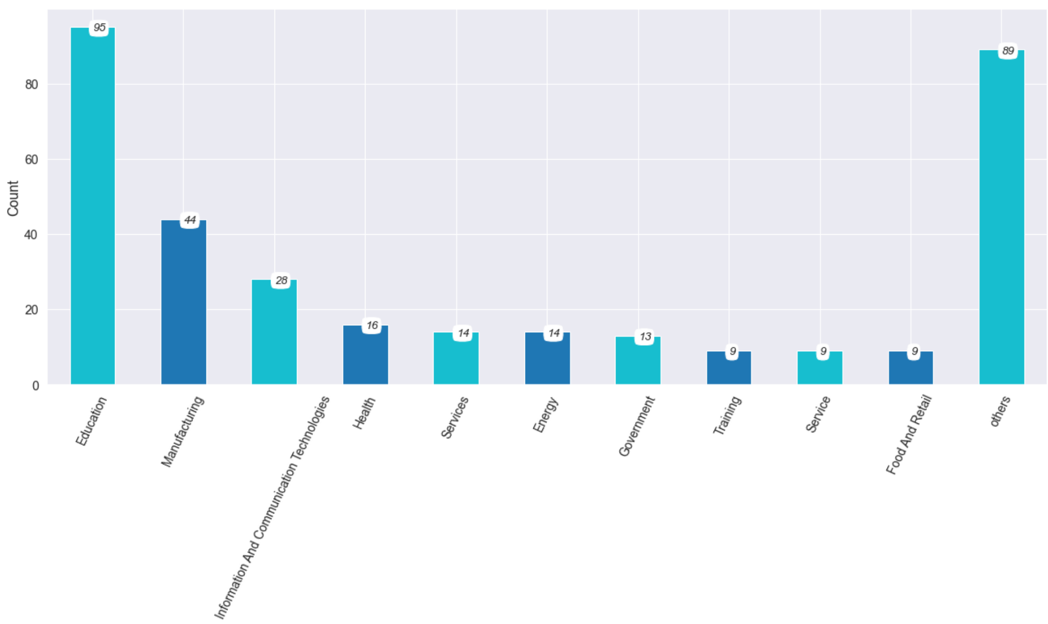

Figure C2: Sectors of the companies.

In the question regarding the number of employees, there were four possible answers to choose between. Figure C3 shows the number of all the possible responses to this question. Most of the companies which participated in this questionnaire are large organisations with more than 250 employees.

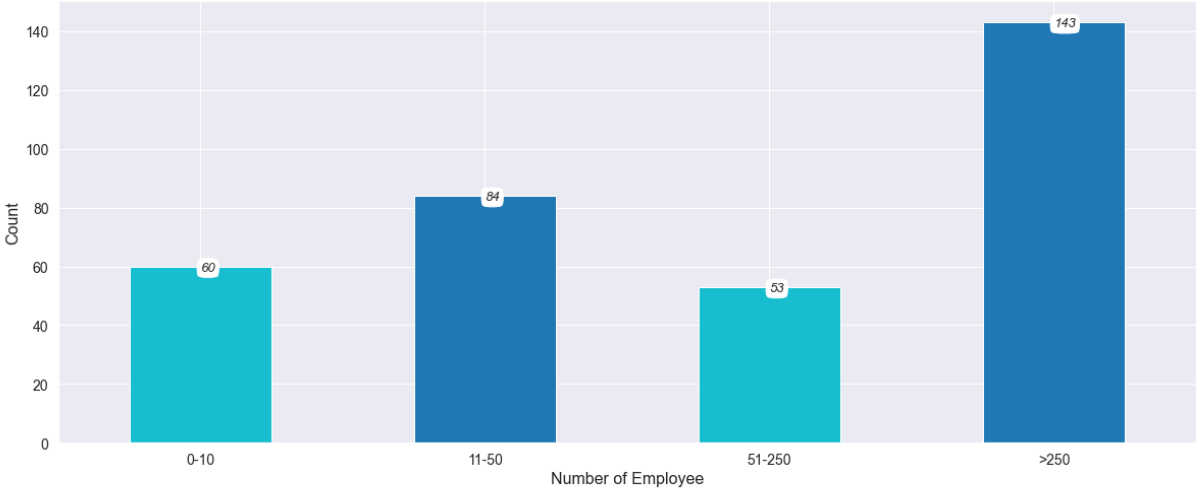

Figure C3. Number of employees in the companies.

Respondent General Data

The following figures are the demonstration of the results obtained from the sections of "respondent general data".

Figure C4 shows the number of respondents from each age range. The highest number of respondents are in the range of 50-55 years.

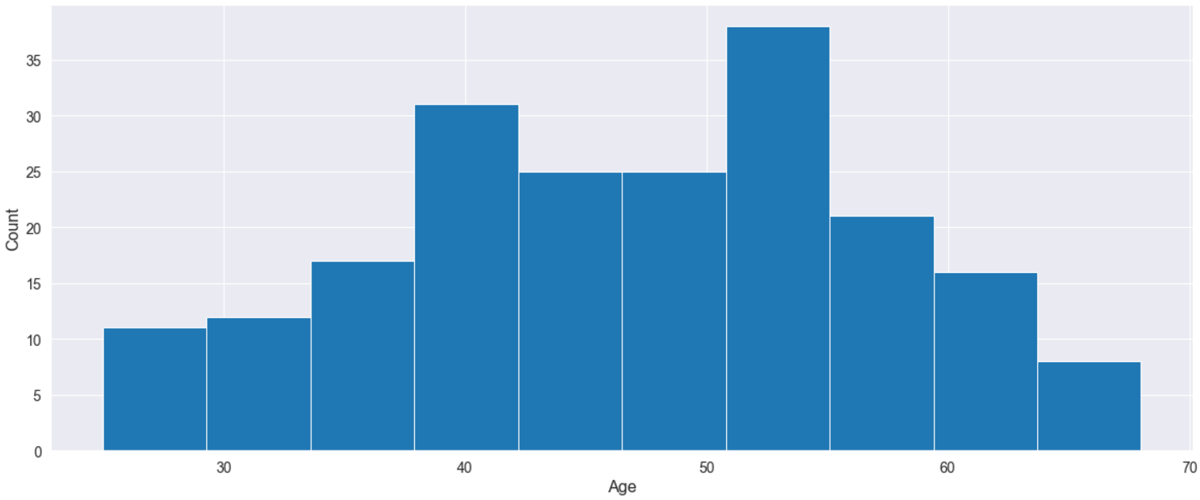

Figure C4. Age of respondent.

Figure C5 shows the level of education of the interviewees. Most of them have a PhD or master's degree. Still there are few respondents with bachelor's degrees or other, which includes cases such as employees with high school diplomas, or with certain certifications.

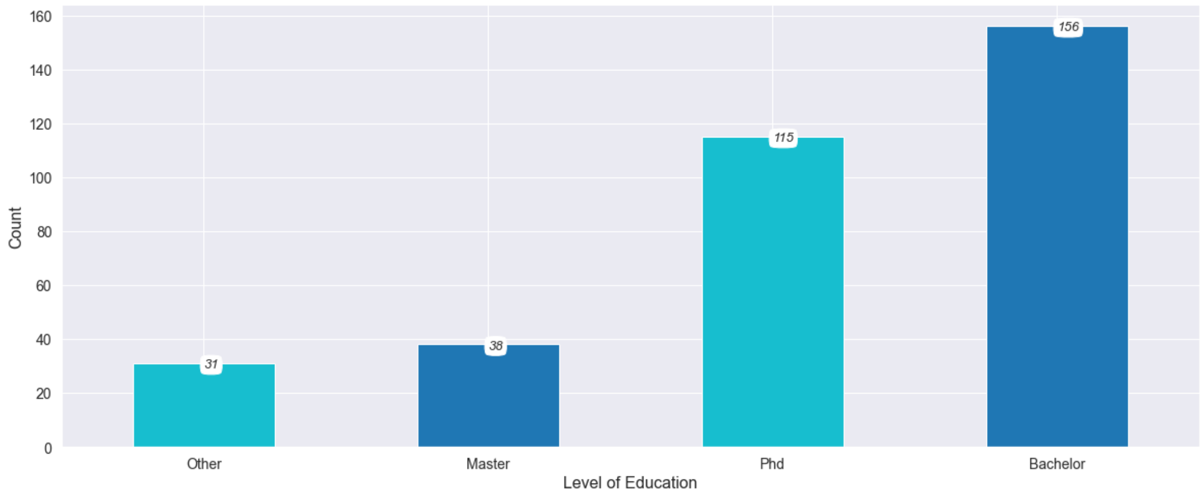

Figure C5. Level of education of the respondents.

Figure C6 shows the distribution of respondents' years of experience in the company they represented.

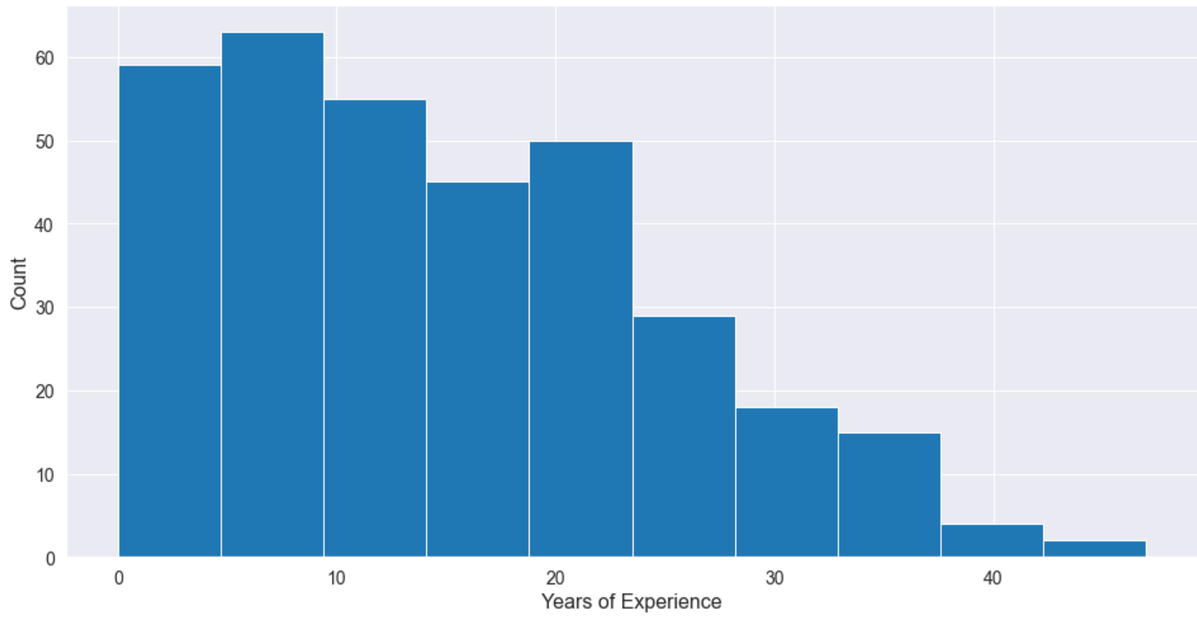

Figure C6. Years of experience of the respondent in the company.

Figure C7 shows the rate of gender distribution among respondents. 52.06% of the total interviewees were male and 38.53% female, while 9.41% of them preferred not to indicate anything.

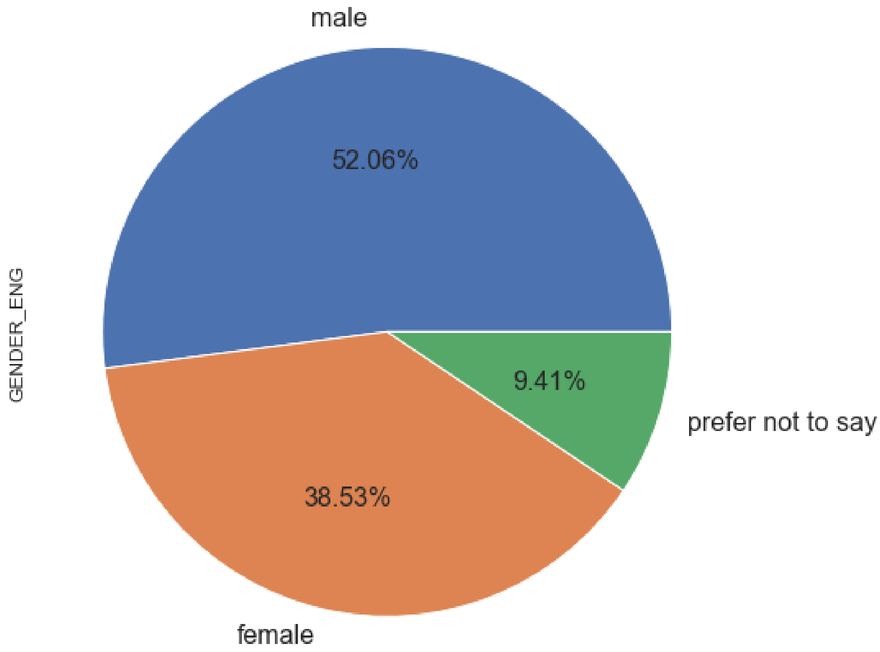

Figure C7. Distribution of the gender of the respondents.

Appendix D - Skill Panorama and ESCO crosswalk

This appendix includes the table used to join the database of ESCO and Skills Panorama based on the ISCO-08 classification.

Table D1. Table used to join the database of ESCO and Skills Panorama based on the ISCO-08 classification.

Occupations Skill Panorama Level 1	Occupations Skill Panorama Level 2	Occupations ESCO Level 1	Occupations ESCO Level 2
Associate professionals	Health associate professionals	Technicians and associate professionals	Health associate professionals
Associate professionals	ICT technicians	Technicians and associate professionals	Information and communications technicians
Associate professionals	Legal & social associate professionals	Technicians and associate professionals	Legal, social, cultural and related associate professionals
Associate professionals	Office associate professionals	Technicians and associate professionals	Business and administration associate professionals
Associate professionals	Science & engineering technicians	Technicians and associate professionals	Science and engineering associate professionals
Clerks	Accounting clerks	Clerical support workers	Numerical and material recording clerks
Clerks	Customer clerks	Clerical support workers	Customer services clerks
Clerks	Office clerks	Clerical support workers	General and keyboard clerks
Clerks	Other support clerks	Clerical support workers	Other clerical support workers
Elementary workers	Agricultural labourers	Elementary occupations	Agricultural, forestry and fishery labourers
Elementary workers	Cleaners and helpers	Elementary occupations	Cleaners and helpers
Elementary workers	Food preparation helpers	Elementary occupations	Food preparation assistants
Elementary workers	Other elementary workers	Elementary occupations	Refuse workers and other elementary workers
Elementary workers	Street services workers	Elementary occupations	Street and related sales and service workers
Elementary workers	Technical labourers	Elementary occupations	Labourers in mining, construction, manufacturing and transport
Farm and related workers	Farmworkers and gardeners	Skilled agricultural, forestry and fishery workers	Market-oriented skilled agricultural workers
Farm and related workers	Farmworkers and gardeners	Skilled agricultural, forestry and fishery workers	Subsistence farmers, fishers, hunters and gatherers
Farm and related workers	Forest & fishery workers	Skilled agricultural, forestry and fishery workers	Market-oriented skilled forestry, fishery and hunting workers
Farm and related workers	Forest & fishery workers	Skilled agricultural, forestry and fishery workers	Subsistence farmers, fishers, hunters and gatherers
Managers	Business managers	Managers	Administrative and commercial managers
Managers	CEOs, officials & legislators	Managers	Chief executives, senior officials and legislators
Managers	Hospitality & retail managers	Managers	Hospitality, retail and other services managers
Managers	Technical managers	Managers	Production and specialised services managers
Operators and assemblers	Assemblers	Plant and machine operators and assemblers	Assemblers
Operators and assemblers	Drivers & vehicle operators	Plant and machine operators and assemblers	Drivers and mobile plant operators
Operators and assemblers	Machine & plant operators	Plant and machine operators and assemblers	Stationary plant and machine operators
Professionals	Health professionals	Professionals	Health professionals
Professionals	ICT professionals	Professionals	Information and communications technology professionals

Occupations Skill Panorama Level 1	Occupations Skill Panorama Level 2	Occupations ESCO Level 1	Occupations ESCO Level 2
Professionals	Legal & social professionals	Professionals	Legal, social and cultural professionals
Professionals	Office professionals	Professionals	Business and administration professionals
Professionals	Researchers & engineers	Professionals	Science and engineering professionals
Professionals	Teaching professionals	Professionals	Teaching professionals
Service and sales workers	Care workers	Service and sales workers	Personal care workers
Service and sales workers	Personal service workers	Service and sales workers	Personal service workers
Service and sales workers	Protection workers	Service and sales workers	Protective services workers
Service and sales workers	Sales workers	Service and sales workers	Sales workers
Trades workers	Construction workers	Craft and related trades workers	Building and related trades workers, excluding electricians
Trades workers	Electroengineering workers	Craft and related trades workers	Electrical and electronic trades workers
Trades workers	Handicraft & printing workers	Craft and related trades workers	Handicraft and printing workers
Trades workers	Metal & machinery workers	Craft and related trades workers	Metal, machinery and related trades workers
Trades workers	Other manufacturing workers	Craft and related trades workers	Food processing, wood working, garment and other craft and related trades workers

Co-funded by the European Union

Disclaimer: 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 the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.