



**CRISP**

Interuniversity  
Research Centre  
for Public Services



## NOTE

### **Seminars of the Big Data Knowledge Hub** **Skills mismatch across EU countries** **using job ads and survey data at the micro-level**

*Notes of the online seminar promoted by the Big Data Knowledge Hub of the European Network on Regional Labour Market Monitoring (ENRLMM). November 24, 2022*

The third of the Seminars of the Big Data Knowledge Hub took place on November 24, 2022. The aim of these series of seminars is to offer an opportunity to deepen the Network's knowledge on how to use Big Data for labour market research and consulting by presenting practical cases and demonstrations.

The project "Skills mismatch across EU countries using job ads and survey data at the micro-level" was presented by **Francesco Trentini** (University of Milano-Bicocca, Italy) and accompanied by **Lorenzo Malandri** (University of Milano-Bicocca, Italy). The open discussion counted with the participation of: **Claudia Plaimauer** (3s Unternehmensberatung GmbH), **Joanna Napierala** (Cedefop), **Michel van Smoorenburg** (UWV), **Aleksandra Webb** (University of the West of Scotland, UK) and **Christa Larsen** (IWAK Goethe University Frankfurt am Main, Germany).

#### **Introduction**

**Eugenia Atin** (Speaker of the Big Data Working Group of the ENRLMM) after the initial greetings and thanks to the participants, contextualises this session in the work being done by the Big Data Working Group of the ENRLMM (European Network on Regional Labour Market Monitoring).

**Christa Larsen** (Founder and Coordinator of the ENRLMM), then presents the European Network on Regional Labour Market Monitoring (ENRLMM) <http://regionallabourmarketmonitoring.net/>, that exists since 2007, and focuses on regional and local monitoring, especially on the demand side (the needs of companies). In 2016, Mario Mezzanzanica from the University of Milano Bicocca introduced a new way to explore data extracting information from the internet, with the focus on online job advertisements. Over the years the approach is growing and maturing,



The Big Data Working Group was also born then, in 2016, to help the network members in the application of big data techniques in their monitoring projects. And then, 2 years ago, the Group designed the Big Data Knowledge Hub <https://bigdatahub.uvt.ro/> which is a collaborative platform for mutual exchange and learning. The Knowledge Hub is the place where all the members of the EN RLMM can look for guidance when aiming to use big data in their labour market monitoring projects. It is an easy accessible source of information on the techniques used by other reference labour market observatories for a particular topic or challenge. The "Seminars of the Big Data Knowledge Hub" are an initiative within this Hub, to make it more interactive.

This third session was dedicated to a project that our partners from Italy are working on. Francesco Trentini from the CRISP research centre of the University of Milano Bicocca presented and described the process used to identify skill mismatches using job ads and survey data at the micro-level.

### **Presentation by Francesco Trentini**

Francesco Trentini first gives an overview of what he is going to speak about:

- Skill mismatch. Definition and used measures
- Novel approach using
  - PIAAC – Program for the International Assessment of Adult Competences (OECD)
  - Online job ads from WIH-OJA (Cedefop and Eurostat)
  - Bridging them with ESCO Skills pillar using AI
- Application of the mapping to measure skill mismatch in 17 European countries

The research has been developed in the framework of the EU funded H2020 [Pillars – Pathways towards inclusive labour markets](#).

We start with the definition of Skill mismatch which broadly is the deviation of workers' competencies from those required by employers to perform a job. So, we have both sides, the supply (the competences of the workers) and what is demanded by employers.

How is it possible to identify skill mismatch? It is complex in terms of the availability of good data.

Data on the supply (the skills of workers) usually comes from the highest level of education attained. On the demand side, there are surveys that are directed to employers, however, usually they lack the timely, frequent and detailed requirements.



In general, we can identify different approaches to measurement:

- Normative approaches: it is possible to ask experts to state the needed level of education or qualification to perform a job
- Objective approaches (or statistical): use the distribution of realised matches and measure the distance from, e.g., the median
- Subjective approaches: workers are asked to self-assess the coherence between their education/qualification and the task that perform on the job

In general, all the variety of approaches have important limitations: low granularity and high heterogeneity within education classes used to identify competencies on the supply side.

So, here we present a new approach where we have identified two sources of data:

1. On the supply side, the PIAAC questionnaire which consists of 2 parts:
  - a. Survey of Adult Skills: direct assessment of skills where participants are asked to complete certain tasks (e.g. use the browser, search for job online) and their performance is measured.
  - b. The [Background questionnaire](#) of individual characteristics: skill use in different domains: at home or at work, this is the part where we have focused on
2. On the demand side, the information on job ads provided by the web intelligence hub (WIH-OJA), the online job ad system, by Eurostat and Cedefop.

There is clearly one structural point to be solved: the two data sources are very valuable, but they express skills using different taxonomies. So, how can we bridge the two, avoiding aggregation and subjective alignment PIAAC question with ESCO skills?

First, we need to understand how the data looks like in both data sets.

### **Supply:**

PIAAC is the Program for the International Assessment of Adult Competences (OECD)

In the background questionnaire, there are different modules that cover different domains in the use of skills. It is a self-declared intensity (frequency or extent) of skill use in different domains. So, a person is asked about how he or she depicts himself. It applies at work, and it covers Literacy, Numeracy and ICT at work and everyday life. The skills are implicitly referenced to by items in the questionnaire.



There are 84 questions in these sections that are suitable for analysis. Moreover, PIAAC is used in the literature to measure skill content in a global perspective (Lewandowski et al. 2022) and used by the OECD to calculate the risk of automation (Nedelkoska and Quintini, 2018).

Figure 1 Example of a PIAAC question. Question F\_Q02b

Layout	Item group table
	<b>How often ^DoesDid your ^JobLastjob usually involve ...</b>
<b>F_Q02b</b> (JRA) (B)	<b>instructing, training or teaching people, individually or in groups?</b>
	1 Never
	2 Less than once a month
	3 Less than once a week but at least once a month
	4 At least once a week but not every day
	5 Every day
	DK
	RF

Figure 1 presents the typical structure of a PIAAC question. The question asks how often do you (in your current or previous work) are involved in instructing, training or teaching people individually or in groups. We can identify that the question broadly refers to the semantic field of “teaching”, but still there is no standardised skill expressed.

## **Demand:**

Considering the demand side, we use Online Job Ads (OJAs) from the WIH-OJA dataset of Eurostat and Cedefop. It is a collection of online job ads from the whole European statistical system: 27 European countries + UK, evolving to soon cover EFTA countries, as well. The time series is since 2018Q4.

We have data on occupations and related skills as they emerge from online job postings. Skills and occupations are codified using the ESCO Skill Pillar and ESCO Occupation Pillar respectively.

## **So, how can we link PIAAC and OJA?**

Our first contribution is using ESCO and a mix of methods using Artificial Intelligence.

- The linkage is done using AI in a framework that combines various methods: embeddings, selection of the best embedding, taxonomy alignment and experts' validation
- PIAAC questions are processed to tag the most similar ESCO Skills.
- The embedding is trained on OJA UK data and the matching is done on the English language.



What we obtain is a validated dataset that covers 21 PIAAC questions and the mapped ESCO skills, enriched with alternative Labels. For instance, coming back to our example of question F\_Q02b, many ESCO terms are associated. The results of the matching are shown in Figure 2.

Figure 2 ESCO Skills associated with the PIAAC question F\_Q02b

F\_Q02b: instructing training or teaching people individually or in groups?

- coach young people
- coach youngsters
- educate others
- educate young people
- facilitate young peoples education
- facilitate young peoples mentoring
- instruct colleagues
- instruct others
- instruct young people
- instructing others
- teach others
- teach young people
- train others
- train young people
- train youngsters
- tutoring

Many more questions / skills were matched but the expert validation reduced the list to these 21. Only the high-quality matches were selected and are presented in *Table 1*



Table 1 Mapped and validated PIAAC questions

PIAAC Question Id	PIAAC Question Description	Label	Group	Type of variable
F_Q02b	instructing training or teaching people individually or in groups?	Teaching people	General	Frequency (time units)
F_Q02d	selling a product or selling a service?	Selling	General	Frequency (time units)
F_Q04a	persuading or influencing people?	Influencing people	General	Frequency (time units)
F_Q05a	[...] «Problem solving» [...]. How often are you usually faced by relatively simple problems that take no more than 5 minutes to find a good solution?	Simple problems	Problem solving	Frequency (time units)
G_Q01b	read letters memos or e-mails?	Read letters memos or mails	Literacy	Frequency (time units)
G_Q01g	read bills invoices bank statements or other financial statements?	Read financial statements	Literacy	Frequency (time units)
G_Q01h	read diagrams maps or schematics?	Read diagrams maps or schematics	Literacy	Frequency (time units)
G_Q02a	write letters memos or e-mails?	Write letters memos or mails	Literacy	Frequency (time units)
G_Q03b	calculate prices costs or budgets?	Calculating costs or budgets	Numeracy	Frequency (time units)
G_Q03c	use or calculate fractions decimals or percentages?	Use or calculate fractions or percentages	Numeracy	Frequency (time units)
G_Q03d	use a calculator - either hand-held or computer based?	Use a calculator	Numeracy	Frequency (time units)
G_Q03g	use simple algebra or formulas?	Use simple algebra or formulas	Numeracy	Frequency (time units)
G_Q03h	use more advanced math or statistics such as calculus complex algebra trigonometry or use of regression techniques?	Use advanced math or statistics	Numeracy	Frequency (time units)
G_Q04	you use a computer in your job?	Experience with computer in job	ICT	Yes (1) / No (2)
G_Q05a	use email?	For mail	ICT - Internet	Frequency (time units)
G_Q05d	conduct transactions on the internet for example buying or selling products or services or banking?	Conduct transactions	ICT - Internet	Frequency (time units)
G_Q05e	use spreadsheet software for example Excel?	Spreadsheets	ICT - Computer	Frequency (time units)
G_Q05f	use a word processor for example Word?	Word	ICT - Computer	Frequency (time units)
G_Q05g	use a programming language to program or write computer code?	Programming language	ICT - Computer	Frequency (time units)
I_Q04d	I like learning new things	Like learning new things	Learning strategies	Extents
I_Q04l	I like to figure out how different ideas fit together	Figure out how different ideas fit together	Learning strategies	Extents



As we can see in the Group column, we cover a variety of different domains: general, problem solving, literacy, numeracy, ICT (Internet, Computer) and learning strategies. We don't cover manual skills like dexterity or precision or such. This is one of the limitations of the current mapping.

We then decided to publish these data sets, so it can be downloaded and used freely by all the community: <https://crisp-unimib.github.io/PIAAC2ESCO/>

We also have a publication<sup>1</sup> and this was developed within the PILLARS consortium.

In the framework of this research with these colleagues, we used the PIAAC2 ESCO crosswalk to link data from OJA to PIAAC questions in order to derive some descriptive insights on skills mismatch on 17 European countries in 2019 and to go into the detail of the relation with automation and training.

On the one side we have PIAAC which comprises 250,000 observations (4,000 - 8,000 per country). It was run in 2012 and 2014. On the other side, the demand side we have WIH-OJA including 17,966,812 observations from 2019. Therefore, in order to have comparable data, we extrapolated the PIAAC observations of 2014 and 2017 to 2019 using changes in the US (observed in 2014 and 2017) as inflation parameters.

How can we measure the skill mismatch?

We follow two recent publications<sup>2</sup> that calculate for each skill and occupation, how relevant is that skill compared to other occupations. So, what we derived from PIAAC and OJAs is the relative frequency of observation of one skill per occupation. This is, in OJAs, for each job ad posted, while in PIAAC for each respondent, to compute whether this is more frequent compared to all other occupations.

We identify the degree of specialisation of occupation by observing the cases in which skills are relatively more requested in an occupation compared to the average demand in all occupation. This measurement can be done on the demand side (OJA, skills

---

<sup>1</sup> Guo, Y., Langer, C., Mercorio, F., & Trentini, F. (2022, September). Skills Mismatch, Automation, and Training: Evidence from 17 European Countries Using Survey Data and Online Job Ads. In *CESifo Forum* (Vol. 23, No. 5, pp. 11-15). Institut für Wirtschaftsforschung (Ifo). [[link](#)]

<sup>2</sup> Alabdulkareem, A., M. R. Frank, L. Sun, B. Al Shebli, C. Hidalgo and I. Rahwan (2018), "Unpacking the Polarization of Workplace Skills", *Science Advances* 4(7), DOI: 10.1126/sciadv.aao6030.

Giabelli, A., L. Malandri, F. Mercorio and M. Mezzanzanica (2022), "GraphLMI: A Data Driven System for Exploring Labor Market Information through Graph Databases", *Multimedia Tools and Applications* 81, 3061–3090. <https://dx.doi.org/10.1007/s11042-020-09115-x>



requested by employers in online job ads) and on the supply side (PIAAC skills used at work). Then we can calculate for each occupation-skill pair the distance (a simple difference) between the importance of skill requested on the market and the importance of that skill on the job.

We use the Relative Comparative Advantages (RCA) adapted to the case of survey and OJA<sup>3</sup>. RCA is a measure of specialisation. Formally, for Occupations  $\bar{O} = \{O_k, k=1, \dots, m\}$  and skills  $\bar{S} = \{S_j, j=1, \dots, l, \dots, p\}$  we calculate the relative skill-frequency  $sf$  for the occupation  $k$  – skill  $j$  pairs:

$$sf(o_k, s_j) = \frac{\sum_{i=1}^n I(o_i = o_k) \cdot I(s_i = s_j)}{\sum_{i=1}^n I(o_i = o_k)}$$

The skill frequency for OJA counts the number of ads of an occupation  $k$  that mention the skill  $j$  (numerator) on the number of total job ads observed in that specific occupation. For PIAAC we map the Likert scale<sup>4</sup> to an indicator variable such that if skill is used at least once a month it is set to value 1, 0 if it is never used.

The RCA of an occupation  $k$  - skill  $l$  couple is:

$$rca(o_i, s_l) = \frac{sf(o_i, s_l) / \sum_{j=1}^p sf(o_i, s_j)}{\sum_{k=1}^m sf(o_k, s_l) / \sum_{k=1}^m \sum_{j=1}^p sf(o_k, s_j)}$$

Therefore, the RCA compares the relative frequency of skill  $l$  in an occupation  $k$ , with the relative frequency of that same skill  $l$  in all other occupations. Its domain is  $[0, +\infty)$ , being 1 its neutral point: for values below this threshold the occupation is not specialised in that skill, while for values above the threshold it is specialised, i.e. the skill is more requested or used compared to the average in all occupations.

The RCA is not linear, so we map it to its percentile rank of belonging. For instance, with the example of the skill “teaching others” for occupation “Teaching professionals” we rank the skill among all occupations and find, that it ranks in the top

---

<sup>3</sup> Giabelli, A., Malandri, L., Mercorio, F., & Mezzanzanica, M. (2020). GraphLMI: A data driven system for exploring labor market information through graph databases. *Multimedia Tools and Applications*, 1-30.

<sup>4</sup> The response scale in PIAAC differs by the type of variable that is considered. In the cases relevant to us, there are three scales. The first, which is used to measure frequency in time units, is a 5-point Likert scale (1 - Never, 2 - Less than once a month, 3 - Less than once a week but at least once a month, 4 - At least once a week but not every day, 5 - Every day). The second scale is used to measure the extent to which the respondents think the statement represents them, again measured on a 5-point Likert scale (1 - Not at all, 2 - Very little, 3 - To some extent, 4 - To a high extent, 5 - To a very high extent), and the third scale is simply binary (Yes/No).



5% of the demand (OJA) for that specific skill. Analogously, on the supply side (PIAAC) we see that it also ranks in top 3% .

Our measure of skill gap is the difference between the RCA percentile rank in supply (PIAAC) and RCA percentile rank in demand (OJA) . In case of positive skill gaps we talk about skill shortage; vice versa we talk about skill surplus. [Indeed, it is appropriate to define them as under-skilling and over-skilling respectively]

PIAAC reports the occupation at the ISCO08 II digit, which constraints us to run the analysis at this level — OJA are currently disclosed at occupation ISCO08 digit IV. Concerning the geographical location, we have information at NUTS2 level on the PIAAC side — OJA have potentially detail at the city level, despite they have complete information at the NUTS2 level.

Skill gaps are therefore calculated at the ISCO08 II digit, skill, and country (or NUTS2 level) level.

[The subsequent part is described in detail in the publication Guo, Y., Langer, C., Mercorio, F., & Trentini, F. (2022, September). Skills Mismatch, Automation, and Training: Evidence from 17 European Countries Using Survey Data and Online Job Ads. In *CESifo Forum* (Vol. 23, No. 5, pp. 11-15). Institut für Wirtschaftsforschung (Ifo). [\[link\]](#)].

Through this method we have calculated the average skill gap by occupation (ISCO 2 digits), for the pool of 17 countries and mapped each occupation to a broad category of belonging in terms of task context (cognitive non-routine, cognitive routine, manual non-routine, manual routine).

What we observe is that there is a relation in terms of the belonging to a category and the intensity of the skill gap. In this case, the skill shortage is more prevalent among manual workers than cognitive workers.

We have also started to inspect the relation between skill shortages and on the job training, so trying to understand if there are relations between the distance in terms of importance in some skills between demand and supply. If there are mechanisms that are related to this divergence. What we observe is that on aggregate (pool of 17 countries), there is a negative relation between the share of participants in on-the-job training in 2012 (PIAAC survey) and the skill shortages that we observe 7 years later. The relationship with the type of tasks of workers is still very present, for example, teaching professionals have high surplus of skills compared to what is demanded and a high share of these workers participate in training, so there is a negative relation.

In the past, PIAAC has also been used by the OECD to calculate the risk of automation, so we have used the PIAAC items of the questionnaire to assign a likelihood of



possibility of automation. On aggregate, we put in relation the two measures and we obtain a positive relation between the skill shortage and the risk of automation. So, some occupations are more prone to changes (tasks changing inside jobs) and this changing content of jobs, makes the skill gap more pronounced.

They have also made a first inspection of the regional variation/ heterogeneity in these skill gaps. Not all regions are covered, but we observe a high heterogeneity between regions but also within country. There are a few countries that in general display skill surplus (BE, DE, FR, IE, SE), while for Greece for example there are skill shortages (except in the region of Athens). There are countries with a mix (ES, IT, LT, PL), some regions have skill shortages and in others just the opposite.

## Conclusions

- Web data can enrich and complement existing datasets (enrich the data and use AI to add new taxonomies and make the data linkable)
- AI-methods can simplify complex tasks and support human experts (putting in relation almost 14,000 ESCO skills and 80 skills not explicitly stated in a PIAAC questionnaire)
  - There is a cost reduction and bounded risk of non-systematic errors
  - We also need a transparent design and account of the process
- What we can learn is very relevant:
  - From the application we learn that jobs are changing in terms of tasks and the skill composition of jobs changes. Skill shortages are negatively related to training and positively related to automation risk.
  - There is a high heterogeneity at the country and regional level calls for more depth in understanding the role of institutional features of the countries labour markets.

## OPEN DISCUSSION

Claudia Plaimauer (3s Unternehmensberatung GmbH) understands that they are using two different measuring units, in PIAAC, there are respondents, real people but in OJAs we don't know how many vacancies there are, how many people might fill the advertisement. So how do they reconcile this?

Francesco Trentini (University of Milano-Bicocca, Italy) responds that on the job-ads side, we observe the number of posted ads with a certain characteristic, so we can actually count the number occurrences among job ads. The skill frequency is how many times we observe a specific skill-occupation pair in all job ads. So, we don't see the



number of people that apply but the number of ads employers post in that specific occupation that stated a specific skill.

Lorenzo Malandri (University of Milano-Bicocca, Italy) adds that in job ads we are interested in how many companies need a skill, the number of open positions, not the number of people that apply. Also, we observe job ads, not vacancies. Because inside the demand there a lot of phenomena occurring and we observe only a small part of the vacancies (not filled internally, etc.). So, we need to focus more on the intensity in the demand of specific skills rather than others which is more relevant than a matter of representativeness in terms of which vacancies will be filled.

Joanna Napierala (Cedefop) asks for clarification, if in the calculation of the RCA index did you include only the 21 skills that are matched, or all skills extracted. Francesco Trentini (University of Milano-Bicocca, Italy) replies that they used them all [Amend: on the OJA case we have calculated the RCA using all the skills and then filtered the relevant matches, while on the PIAAC side we first filtered the questions and then calculated the RCA. Therefore we will correct the procedure to make it consistent.]. Moreover, we have a great deal of relevant information that we are currently not using: the unmatched questions/skills in both PIAAC and OJA. We have not actively considered how to take them into account in the analysis but is one of the things that that they will further explore.

How many more skills are in the PIAAC database? In those 4 modules there are 84 questions, 70 could be used, all of them were matched but then only 21 were validated by the experts. So, they only kept those with higher quality.

Michel van Smoorenburg (UWV) is rather surprised by the results, especially the negative relation between on-the-job training participation and the skill shortages. He would expect a positive relation because when there are large skill shortages you may be motivated to improve and extend training participation within a company. Francesco Trentini (University of Milano-Bicocca, Italy) replies that on-the-job training is observed in 2012 and skill shortages are observed 7 years later, so there is a delay in the observation. So, we could interpret that the low share of participation in on-the-job training in 2012 led to a more pronounced skill shortage. So, it allowed to identify which sectors were less prone to skill shortages. The mechanisms are not clearly inferable from the results, we would need to go more into detail in the institutional context of each country, how firms behaved in that period... but one thing that emerges the more the workforce participates in the training, the more it is up to the request of the market. Which occupations are those that employers take more care of, there are also some dynamics, how training is deployed, the stability of contracts, many mechanisms that could be further explored. I would not say that training is effective in reducing skill shortages, that would be a bit too much, but there are some clear relationships that we could further explore.



Aleksandra Webb (University of the West of Scotland, UK) would want to know more about the expert validation process. Francesco Trentini (University of Milano-Bicocca, Italy) replies that there are embeddings involved. For each PIAAC question, the expert received the top 10 most similar ESCO skills suggested by the system and the expert had to vote on a Likert scale to which extent the ESCO skill correctly represents the PIAAC question. The text of the question was manipulated beforehand. So, the expert voted and then all the matches that were below the top level of the Likert scale were discarded. So, we kept those where all experts had the same judgement that the match was reliable. And who were the experts? They were a group of colleagues of the PILLARS consortium. The system could be further extended to include other stakeholders in the validation.

Michel van Smoorenburg (UWV) has a question about the graph of Europe, he is surprised by the results because when you look at the tightness of the labour market and the unemployment figures, you see major skill shortages in the northern and central countries, and a high level of employment in Spain, South of Italy, Greece... And in this graph, it is just the opposite. Francesco Trentini (University of Milano-Bicocca, Italy) replies that here what we observe is how is the workforce ready for the request of the market, we are not observing unemployed individuals. PIAAC helps us identify the skill use on the job, we are not observing the unemployed. This graph shows the relationship between the skills of workers and the demand of the market, so the importance that is given to the skills to perform a job and the demand of employers. Michel asks if we could say that in the southern countries the employee is underskilled and in the northern and central they are overskilled. Yes, Francesco says that this could be derived from the observation. Michel suggests validating these results by other research, through employer surveys, as this would be very interesting. Francesco thinks that there already exists research about this but doesn't know if there are already surveys. If anyone knows about this kind of surveys/ research, please let Francesco know.

Lorenzo Malandri (University of Milano-Bicocca, Italy) adds that an underskilled worker could be underskilled because of the education system or because the online job market requested too many skills (more than needed). Francesco Trentini (University of Milano-Bicocca, Italy) says that there is dimension that they had to discard when tagging which occupations do not use specific skills. They discarded all the part about the intensity on jobs which can be an interesting part of the research, this could help understanding if there is a relationship between the frequency of the use of a skill and how essential it is for an occupation. To use the gradient to know more about the essential skills, maybe different evidence could emerge.

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) says that it is very interesting because the countries that have surpluses have strong VET systems, so



this could be further explored. She also wants to thank Francesco Trentini (University of Milano-Bicocca, Italy) for this approach because up until now we had some approaches either on one side or the other and this one is matching both. We always assume that there is a matching going on in the labour market, and with this methodology we try to reconstruct this matching, combining different type of data sets and developing methodologies to make them comparable. So, from the methodological side it is very interesting for us in the Network, we can learn a lot from such an approach.

Christa would also want to ask a question about the sample size of the survey data. Was there a problem going to NUTS 2 which such a small sample? Did you have to develop a method/ technology to handle this situation to be able to go to a regional level. Francesco Trentini (University of Milano-Bicocca, Italy) replies that there is of course a limitation from the sample size and that we are bounded to what we have. We left out poorly populated cells.

We did not develop any technology, it is methodological the same, but we can match at the lower level, in terms of occupation (ISCO08 II digit), region (NUTS2) and skills (ESCO Skill fourth hierarchical level).

Francesco Trentini (University of Milano-Bicocca, Italy) adds that we are working on using AI to bridge national classifications to ESCO, to provide a sound methodology that can help us align different classifications to create crossworks and this is one of the dimensions on the use of these technologies which can really help improve and enrich the existing surveys with external data sources. Using web data has the strong limitation of not knowing which population you are talking about, you have to make a lot of assumptions. Surveys as a starting point can be a very valuable and having a standard to enrich the survey.

In 2023 [amend: 2024] there will be a new release of PIAAC.

Lorenzo Malandri (University of Milano-Bicocca, Italy) highlights 3 points:

- Use of data, benefit of having quality data
- Linked data, find the right dimension to match the data
- Interaction between AI and human is necessary, AI can be the telescope and the human can do the validations

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) asks if the methodology would be available for Network members to getting inspired, it would be a good basis to explore further. Trying to reconstruct the dynamic between the supply and the demand, it would help a lot to have a description of the methodology. Francesco



Trentini (University of Milano-Bicocca, Italy) replies that concerning PIAAC to ESCO, there is a general description and a technical annex available on the website. Francesco has added more details about the methodology throughout this document so that the Network members can better understand the method.

The next seminar of the series will be delivered by Switzerland, the 23rd of February 2023 under the topics:

- Improving refugee integration through data-driven algorithmic assignment
- Implementation of Risk Assessment Tools in the Public Employment Service

More information will be circulated as always through the Network's newsletter. Also, if you are interested in sharing your experiences with the Network, please do so through the Knowledge Hub <https://bigdatahub.uvt.ro/> and contact Eugenia Atin if you wish to present your project through a Seminar.

*Bilbao, December 2022*

#### **References**

##### **Speaker: FRANCESCO TRENTINI**

University of Milano-Bicocca, CRISP and LABORatorio R.Revelli , Italy

Email: [francesco.trentini@unimib.it](mailto:francesco.trentini@unimib.it)

##### **EUROPEAN NETWORK ENRLMM**

Email: [jenny@jennykipper.de](mailto:jenny@jennykipper.de)

Email: [c.larsen@em.uni-frankfurt.de](mailto:c.larsen@em.uni-frankfurt.de)

Website: [www.regionallabourmarketmonitoring.net](http://www.regionallabourmarketmonitoring.net)

##### **BIG DATA WORKING GROUP**

Eugenia Atin

E-mail: [e.atin@prospektiker.es](mailto:e.atin@prospektiker.es)

Website: <https://bigdatahub.uvt.ro/>