NOTES

Seminars of the Big Data Knowledge Hub

The digital skills employers want: A real-time skills intelligence taxonomy based on online job vacancies

Notes of the online seminar promoted by the Big Data Knowledge Hub of the European Network on Regional Labour Market Monitoring (ENRLMM). June 27, 2024

The seventh of the Seminars of the Big Data Knowledge Hub took place on June 27, 2024. 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 Seminar included a presentation by **Ludivine Martin** (Luxembourg Institute for Socio-Economic Research (LISER)) who presented the project "The digital skills employers want: A real-time skills intelligence taxonomy based on online job vacancies".

The open discussion counted with the participation of: **Trang Dau Thi** (GIZ), **Christa Larsen** (IWAK Goethe University Frankfurt am Main, Germany), **Moreno Baruffini** (USI, Switzerland) and **Nathaniel Moser** (SECO, Switzerland).

Introduction

Eugenia Atin (Speaker of the Big Data Working Group of the ENRLMM) after the initial greetings and thanks to the participants, contextualised the 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) welcomed the participants in the name of the European Network on Regional Labour Market Monitoring (ENRLMM). The European Network on Regional Labour Market Monitoring has existed for 19 years, with its 20th anniversary approaching next year. This network unites scientists and practitioners dedicated to labour market monitoring, particularly at the regional level, including larger cities.

Methodologically, labour market monitoring presents significant challenges. This network was established to exchange ideas, explore various methods, and identify useful data types. Although there is ample data on the supply side of the labour market, capturing the demand side, especially at regional or local levels, is difficult through national surveys.

To address this challenge, the network has been open to new data sources beyond surveys for several years. In 2015 or 2016, the Big Data Working Group was founded to explore the use of internet data, particularly from online job vacancies. The group has exchanged numerous ideas and tracked developments both methodologically and in terms of applications, given many members' involvement in policy consulting, making the application perspective crucial.

The European Network on Regional Labour Market Monitoring welcomes anyone interested. This year's topic is the lack of skilled labour, and discussions will include mismatch analysis using multi-level approaches. An anthology on the topic has been prepared, with contributions to be presented and discussed during the meeting in Lugano in September. The conference is open to all at no cost. More information is available on the network's website (http://regionallabourmarketmonitoring.net/).

The network eagerly anticipates hearing from Liser, one of its founding members. Annual meetings have been held at Liser in Luxembourg, which is of particular interest due to its international and trans-regional labour market. Luxembourg's focus on cross-border transparency aligns well with the network's goals. The presence of a Liser representative at today's seminar is greatly appreciated.

Presentation by Ludivine Martin

Ludivine Martin has been a Research Scientist in the Labour Market Department of the Luxembourg Institute for Socio-Economic Research (LISER) since July 2008. She holds a PhD in Economics from the University of Rennes 1 (France) and is authorised to supervise PhDs in both Economics and Psychology. Her research focuses on how digitisation and human resource management practices affect the workplace and employees' job quality, job satisfaction, motivation, performance and skills. The development of her current research program is based on various research grants supported by the FNR (CORE TWAIN, BRIDGES LOWSKIM, INTER calls ENGAGE and Heart@Work) and the European Commission (H2020 UNTANGLED, Horizon Europe; WeLaR, SkiLMeet, WinWin4WorkLife, Erasmus+ LMIEUniv). She has recent publications in Applied Economics, Economic and Industrial Democracy, European Economic Review, European Review Of Applied Sociology, Industrial Relations, PlosOne, Social Science Computer Review, The Information Society.

Ludivine Martin (LISER) dives into the paper she has been working on with her colleagues, Thiago Brant from Liser and Kamil Filipek from Maria Curie Skłodowska University in Poland. This paper is part of a research project funded by the Luxembourg National Research Fund, focusing on the skills needed in the market, particularly green and digital skills. For this presentation, she will concentrate on digital skills.

The labour market is becoming increasingly digitalized, especially since the 2020 lockdowns, which led to a surge in technology adoption. A graph of the slide shows data they collected in early 2020 in Luxembourg, illustrating how people adapted to new technologies. Tools like web conferencing, instant messaging, groupware, workflow, and enterprise resource planning saw significant increases. These are not necessarily expert-level digital skills but are increasingly in demand. The lockdown provided an opportunity for people to discover and use new digital skills.

Unfortunately, many workers still lack sufficient digital skills. Data from December 2018 shows a significant proportion of people in the EU27 and our case study countries (Belgium, France, Germany, and Luxembourg) recognize the need for more digital skills. However, national-level data is currently all they have, and they aim to refine this to a regional or city level once the data is cleaned.

The Eurobarometer survey from December 2019 highlighted barriers to digital skill development, such as lack of time, uncertainty about which skills are needed, lack of accessible training, and cost concerns. Many workers know they need to improve their digital skills but feel hesitant to take action.

In their research, they noticed that most studies focus on broad categories of digital skills, typically dividing them into basic digital skills, computer literacy, and advanced skills like programming or app development. However, this categorization is too broad and lacks specificity.

Some research expands on these categories to include skills like information, data communication, digital content creation, safety, and problem-solving. Other studies use occupational dictionaries, such as the O*NET in the U.S. or the European ESCO, to define digital competencies, but these often provide only a single digital competency indicator or divide skills into user, practitioner, and developer levels.

Recently, there has been a surge in literature focusing on AI skills. Notable contributions include works by Alekseeva and Babina for the European Commission, and a new OECD paper by Lane & Williams. However, these often include outdated analysis methods like cluster analysis, which they do not fully agree with.

Their paper addresses three main research questions:
1. Are existing typologies of digital skills sufficient to track labour market changes?
2. Which digital skills are in high demand by companies in our four case studies?
3. What about AI skills—are they really what employers want?

They use data from the Lightcast database, which includes online job vacancies from September 2018 to the present, focusing on data from 2023. They have access to the raw job vacancy texts, allowing them to extract sentences and keywords.

Their analysis includes data from Germany, Belgium, France, and Luxembourg. They use existing dictionaries and taxonomies, such as ESCO, which lists skills like using word processing software. However, they found that job ads typically use more practical terms, like "Word" instead of "word processing software."

They also incorporate keywords from Stack Overflow and AI-related taxonomies, covering areas like computational linguistics and computer vision. Their approach involves first applying these taxonomies, then refining them by examining job ads and capturing new keywords while excluding irrelevant ones.

Based on these existing keywords, they developed a dictionary of digital skills, competencies, and abilities, which they categorized into broader domains to effectively illustrate market trends. Their dictionary includes over 1850 digital skills. Most of these keywords are in English, especially for technologies, such as GPT, Teams, and Excel, which are not translated into other languages. However, for terms like basic computer skills, they needed translations in all the languages used in our case studies—French, German, Dutch, and English.

They then created a Python script using natural language processing to match these keywords within job texts. This method allowed us to search for and identify keywords in job descriptions accurately.

They captured 12 domains, ranging from basic generic knowledge (e.g., Excel, Word) to more specific areas such as online business, marketing, and user experience. These domains include skills needed for managing an online presence and selling products online, such as e-commerce, Amazon, and SEO. Additionally, business and project management skills are highly requested, including software for planning and supply chain management, such as SAP and SharePoint.

They also identified digital skills in industry 4.0, including automation and computerized management systems. Technical digital skills include data management, data analysis, software and app development, computer programming, digital security, and privacy. With the advent of ChatGPT, they added a 12th category for generative AI. For instance, they distinguish programming languages like Python as general skills rather than AI-specific.

An example job ad for a production engineer highlights relevant digital skills such as Linux, Azure, and Redis. They exclude terms like "Facebook" and "LinkedIn" when they do not pertain to job requirements but appear in context like "join our team on Facebook and LinkedIn."

Their analysis of job vacancies in 2023 across Belgium, France, Germany, and Luxembourg reveals a strong demand for basic and generic digital knowledge. For instance, in Belgium, the demand for such skills exceeds 30%, followed by business and project management skills above 10%, with variations among countries. Contrary to the literature's emphasis on AI skills, our findings indicate these are less frequently requested, except in Luxembourg where they make up about 5% of digital skills demanded.

Their breakdown of the 12 domains includes:

- Basic generic knowledge: Excel, Word.
- Online business, marketing, and user experience: Amazon, SEO, Google Analytics, Photoshop.
- Business and project management: SAP, enterprise resource planning, customer relationship management.
- Applied digital skills in industry 4.0: Automation, CNC, PLCs.
- Data management: Databases like SQL, data quality, Elasticsearch.
- Data analysis: R, SAS, general terms like data analysis.
- Software and app development: Testing, HTTP, DevOps.
- Computer programming languages: Python, JavaScript, Java.
- Digital security and privacy: Cybersecurity, GDPR, data protection.
- AI: artificial intelligence; machine learning; deep learning; virtual reality; image processing; computer vision, tokenization, word2vec
- IT infrastructure management: Linux, hardware, VMware.
- Generative AI: AI proficiency, computer vision, algorithms.

Notably, Belgium stands out with high demands for specific digital skills across these domains.

To conclude, existing taxonomies are insufficient for capturing the dynamic nature of digital skills in the labour market. Their research, spanning three years, provides a comprehensive framework, categorizing digital skills into 12 distinct domains. This approach aligns with literature highlighting the high demand for basic and generic skills, followed by business and project management tools.

Regarding AI skills, their data from four markets shows that these are not in high demand, with less than 2.5% in three countries and around 5% in Luxembourg. Their next steps involve determining the demand for digital skills by occupation, but this requires further data cleaning due to inaccuracies in occupational classifications.

They are currently working to ensure that they accurately capture the real skills requested in job ads. Once this is achieved, they will publish a paper analysing the

evolution of skill demand from 2018 to 2024. Additionally, they have drafted a paper examining the complementarity and substitutability of skills in job ads to help prioritize training programs. For instance, some job ads list numerous skills under a single broad category. They are developing a framework to identify which digital skills are related. This way, when designing training programs, they can first focus on foundational skills and then progressively introduce more advanced ones as proficiency increases.

OPEN DISCUSSION

Eugenia Atin (Big Data Working Group) thanks Ludivine Martin for her presentation and opens the floor for questions and comments. She starts with a question about the 12 domains. It's typically hard to classify digital skills into different domains. The classification made here is impressive and very suitable, especially for someone who has struggled with determining where to place a given digital skill. Were these 12 domains created by LISER? Ludivine Martin (LISER) answers yes, initially, the team used classifications found in academic literature and other papers, but these proved insufficient. Together with colleagues, they compiled an extensive list of digital skills and assigned them to specific domains. They consulted with experts to define these broad domains and categorized the digital skills accordingly. With the rise of GPT and language models, they benchmarked their human classification against models like ChatGPT, Mistral, and Falcon. When discrepancies arose, they researched the purpose of the skills online to ensure accurate classification within the intended domain.

Trang Dau Thi (GIZ) wants to know if Ludivine Martin has utilized Lightcast Open Skills for developing /enriching the dictionary. Ludivine Martin answers that they have used a part of the dictionary of Lightcast because it provides a list of keywords based on Stack Overflow but decided to add more keywords.

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) expressed appreciation for the intriguing focus on the digital skills domain. She found it interesting because it closely relates to applying skills within companies, offering a perspective beyond individual abilities. Christa also commended the reasonable structure of the approach. She then inquired about translating different techniques to specific domains, which Ludivine acknowledged as a reasonable consideration.

Regarding the dataset, Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) raised a question about whether the job ads included only commercial companies or also encompassed the public sector or social organizations. Ludivine clarified that the data came from both big companies and job portals. While social organizations were not explicitly covered, discussions with the Employment Agency Luxembourg revealed that some data originated from both the agency and online sources.

Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) further explored the differences observed across countries, particularly the high prevalence of SAP application in Germany. Ludivine explained that the dataset captured the same branches across all countries, emphasizing consistency. The variations in patterns were not due to branch selection but rather factors like past dependencies and different development stages in each country.

Ludivine Martin (LISER) reflects on the evolving landscape of digital skills, emphasizing the critical role of past dependencies and sustained technology adoption. She appreciates the feedback on their recent paper and acknowledges the discrepancy between discourse and real-world application in sectors like automation and automotive. Christa Larsen (IWAK Goethe University Frankfurt am Main, Germany) adds insights on regional disparities in digitalization, noting the prevalence of small to medium-sized enterprises in Germany with varying levels of technological integration. The discussion touches on examples like a startup in Luxembourg and Germany focused on automated battery recycling, highlighting the demand for engineers over manual labourers. They conclude that while digital skills like web conferencing are widely adopted post-2020, advanced AI skills remain less prevalent in job markets.

Moreno Baruffini (USI, Switzerland) posed a technical question about how Ludivine's team handled the inclusion and exclusion of certain words, particularly in lengthy texts. Ludivine clarified that they compiled digital skills into a dictionary and observed high demand for terms like Facebook, LinkedIn, and JavaScript. However, they also noticed instances where these words were used solely for company promotion (e.g., "follow us on Facebook"). To address this, they maintained an exclusion list while analyzing job ads.

Regarding teleworking and homeworking statistics, Moreno Baruffini (USI, Switzerland) inquired whether Ludivine's research focused specifically on these types of work or had broader implications. Ludivine explained that her research centered on digitalization and organizational practices within firms, examining their impact on workers' job quality, well-being, and skills. The data collected, both from Luxembourg and European sources, allowed them to graph the surge in digital skills usage during the pandemic.

Eugenia Atin asks if there are other ongoing projects on this topic. Nathaniel Moser (SECO, Switzerland) responded affirmatively, indicating that they have a small ongoing project focused on matching vacancies and job seekers, using competencies as a basis. He mentioned a related project from the University of Zurich led by Marlis Buchman (How digitalisation is changing task and skill profiles of occupations (nfp77.ch)), which employs a natural language model to assign ESCO competencies

from job descriptions. This project is still in the prototype stage, and they are evaluating its effectiveness in a matching context.

Trang Dau Thi (GIZ) brought up a question about tools for automatically detecting emerging trends, to which Ludivine Martin responded that manual examination is still preferred due to challenges with current AI models like GPT, which often produce hallucinations. The group discussed the absence of automatic tools for detecting emerging trends, expressing hope for future developments.

Eugenia Atin (Big Data Working Group) concluded by suggesting that if anyone had ideas or wished to share a topic or project, they should get in touch for future seminars. She thanked everyone and wished them a nice afternoon and summer.

Bilbao, June 2024

References

Speaker: Ludivine Martin
Luxembourg Institute of Socio-Economic Research (LISER)
Email: ludivine.martin@liser.lu
EUROPEAN NETWORK ENRLMM
Email: jenny@jennykipper.de
Email: c.larsen@em.uni-frankfurt.de
Website: www.regionallabourmarketmonitoring.net
BIG DATA WORKING GROUP
Eugenia Atin
Email: e.atin@prospektiker.es
Website: https://bigdatahub.uvt.ro/