SKILLAB: Creating a Skills Supply and Demand Data Space

KOSTANTINOS GEORGIOU, School of Informatics, Aristotle University of Thessaloniki, Greece

ROSARIA ROSSINI* and MARCO JAHN*, Eclipse Foundation Europe, Germany

DIMITRIOS TSOUKALAS, Centre for Research and Technology Hellas, Greece

VASSILIS VOULGARAKIS, VILABS LTD, Cyprus

SOFIA TSEKERIDOU, Netcompany-Intrasoft SA, Luxembourg

MIHAELA ALUAS, Asociatia Cluj IT, Romania

APOSTOLOS VONTAS*, VILABS LTD, Cyprus

DIONYSIOS KEHAGIAS*, Centre for Research and Technology Hellas, Greece

NIKOLAOS MITTAS, School of Sciences, Democritus University of Thrace, Greece

APOSTOLOS AMPATZOGLOU, School of Applied Informatics, University of Macedonia, Greece

VALIA KORDONI, Humboldt-Universitaet Zu Berlin, Germany

ALEXANDER CHATZIGEORGIOU*, School of Applied Informatics, University of Macedonia, Greece

LEFTERIS ANGELIS*, School of Informatics, Aristotle University of Thessaloniki, Greece

The advent of Industry 5.0 as a defining concept for the future, which advocates a human-centric coalescence of humans and technology or software, renders the skilled workforce the most important asset in any organization or business. The society is 'forced' to adapt itself to technological change and progress for setting the necessary skillsets for the workforce. In order to follow the digital transformation, it is necessary to evoke the reshaping, evolution, or replacement of traditional and possibly obsolete processes at intra- or inter-organizational levels in multiple aspects, introducing innovative ways of re-defining the workforce. To do so, the key piece are data. In this context various platform collect and organize data, also exploiting the new era of Data Spaces (DS). SKILLAB will act as a smart tool for handling, honing, and widening the competencies of the personnel of companies, forecasting future skill gaps and providing European citizens with a tool for upskilling and reskilling, exploiting DS.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI)**; **Collaborative and social computing**; • **Information systems** → **Data management systems**; **Information systems applications**; • **Applied computing** → **Enterprise computing**; • **Computing methodologies** → **Artificial intelligence**; **Machine learning**; • **Security and privacy** → **Database and storage security**.

Additional Key Words and Phrases: Data spaces, Skills Development, Skills Supply, Skills Demand, Skill Gaps

---

*Both authors contributed equally to this research.

Authors' Contact Information: Kostantinos Georgiou, konsgeor@csd.auth.gr, School of Informatics, Aristotle University of Thessaloniki, Greece; Rosaria Rossini, rosaria.rossini@eclipse-foundation.org; Marco Jahn, marco.jahn@eclipse-foundation.org, Eclipse Foundation Europe, Darmstad, Germany; Dimitrios Tsoukalas, tsoukj@iti.gr, Centre for Research and Technology Hellas, Greece; Vassilis Voulgarakis, vmvoulga@vilabs.eu, VILABS LTD, Cyprus; Sofia Tsekeridou, Sofia.TSEKERIDOU@netcompany.com, Netcompany-Intrasoft SA, Luxembourg; Mihaela Aluas, mihaela.aluas@clujit.ro, Asociatia Cluj IT, Cluj, Romania; Apostolos Vontas, avontas@vilabs.eu, VILABS LTD, Cyprus; Dionysios Kehagias, diok@iti.gr, Centre for Research and Technology Hellas, Greece; Nikolaos Mittas, nmittas@chem.duth.gr, School of Sciences, Democritus University of Thrace, Kavala, Greece; Apostolos Ampatzoglou, apostolos.ampatzoglou@gmail.com, School of Applied Informatics, University of Macedonia, Greece; Valia Kordoni, evangelia.kordoni@anglistik.hu-berlin.de, Humboldt-Universitaet Zu Berlin, Berlin, Germany; Alexander Chatzigeorgiou, achat@uom.edu.gr, School of Applied Informatics, University of Macedonia, Greece; Lefteris Angelis, lef@csd.auth.gr, School of Informatics, Aristotle University of Thessaloniki, Greece.

---

1 INTRODUCTION

Skills (both hard and soft) are an important concept in the operations and functions of the workforce. Possessing and acquiring the necessary skills and competences ensures that both prospective and current employees remain up to date with technological advancements and shifts in digital tools and evolving scientific IT, engineering and computer science fields such as AI, cybersecurity, data management, etc.. More often than not, companies and organizations that observe the latest trends in the technological landscape, invest on the upskilling and reskilling of their workforce, ensuring that they retain their competitive advantages and that talented employees are continuously honing their skillsets and advance from their current positions.

However, the process of upskilling and reskilling is frequently hindered by gaps in the skills that are in demand by employing organizations or employers and the skills that are actively supplied by the workforce as acquired primarily by higher education programs. This problem stems from fundamental transitions in the workplace, characterized by the principles of digital transformation, remote working, green skills and unexpected crises (e.g. the COVID-19 pandemic). In addition, the recent and skyrocketing growth of Large Language Models (LLMs) may lead to a number of positions becoming obsolete or requiring crucial restructuring to function properly.

Hence, it becomes clear that the labour market needs to adapt to the ongoing digital transformation, while monitoring and forecasting future developments. To achieve this, the provision, management and processing of relevant data play a crucial role in predicting trends in future market demands on advanced and digital skills, and recommending adaptations in current training/educational programs for timely and customized upskilling/reskilling of workforce. The rapid flow and sharing of information and data arms the workforce with access to learning pathways and leads to efficient skills development, bridging the growing gaps. To that end, the recent rise of DS offers enticing opportunities for linking data from different sources and investing in skills development.

In this paper the concept of a DS for skills supply and demand data is analyzed and presented via the context of SKILLAB[1], a HORIZON Europe funded project that aims at monitoring the European Labour Market, collecting data regarding skills supply and demand and synthesizing the accumulated knowledge to provide actionable insights, forecasts about trending and future skills, as well as practical implications for stakeholders and policymakers. The wealth of knowledge collected and extracted from SKILLAB could, potentially, be organized and offered as a DS that would enable companies and organizations to achieve effective upskilling, utilizing a full-fledged framework of data collection, data processing and data storage/management.

The rest of the paper is organized as follows: In Section 2 we provide a brief overview of SKILLAB, emphasizing on its main objectives, expected outcomes and contributions as well as the information that is leveraged by its services. In Section 3 the relationship of SKILLAB with DS is described while we also analyze the SKILLAB architecture and the potential pathways it could be implemented as a DS or integrated in a DS. In Section 4, the pilot cases of SKILLAB are presented, in terms of their expected benefits, leveraged data and relationship with DS. Finally, in Section 5 some closing remarks are provided along with some interesting future work directions that could be implemented to link SKILLAB with DS in a more concise manner.

2 SKILLAB OVERVIEW

The SKILLAB project is an international project that comprises of 10 partners from 6 European countries (Greece, Romania, Germany, Cyprus, Luxemburg, Spain), which began in January 2024 and will last for three years, until December 2026. At its core, SKILLAB aims to deliver an efficient platform that will act as a crucial service for European citizens, industrial stakeholders, policymakers and educational institutions. The primary objective of SKILLAB is the exploitation of open and industrial data that concern both the European Labour Market but also the policies surrounding its development and evolution. The produced platform will be augmented by statistical methodologies, ML and explainability algorithms, while the research efforts will strive to include latest developments in ICT, such as LLMs [Zhao et al. 2023], Kolmogorov-Arnold networks [Liu et al. 2024] or Auto-Encoders [Liou et al. 2014; Tschannen et al. 2018; Zhai et al. 2018].

The methodologies and architecture of SKILLAB have been designed with flexibility and efficient data retrieval as the primary objective. Via the convergence of multiple sources and the possible development of a SKILLAB DS, the project aims at (a) utilizing statistical methodologies for the identification of skill gaps in the European Labour Market; (b) providing support to organizations and companies for the enhancement of hiring and HR policies, with special emphasis on employee training and skills development; (c) enabling HR recruiters to focus on specific sought-out skills and simultaneously; (d) offering multiple means and data to EU citizens, allowing them to enhance their skillsets, develop their skills and be incorporated in the ever-changing labour market.

As mentioned previously, the goals of SKILLAB will be achieved by designing and developing a smart hub that will be able to continuously monitor the skills of the labour market, bridge existing skill gaps and forecast future gaps that may emerge. The data collection pipelines of SKILLAB include the retrieval of data regarding the labour market (e.g job vacancies) and the overall state of skill demand, data on skills supply (e.g. online courses, user profiles) and available policies on skills development and upskilling from the European Union and other official bodies[2],[3]. The developed smart hub will be complemented with cutting-edge methodologies that will exploit statistical analysis, Machine Learning (ML), data mining and explainability techniques. The insights and outcomes of the analysis pipelines will be leveraged by other components of the platform in order to provide effective recommendations for Human Resource Management (HRM) departments, educational pathways for institutions and EU citizens, as well as practical policymaking directives for stakeholders.

The expected outcomes of SKILLAB can be traced in its contributions to the monitoring of the labour market, the provision of recommendations for upskilling and reskilling, as well as the application of data-driven methodologies for skill gaps identification. More specifically, the carefully designed data collection and data analysis procedures, in conjunction with the use of established European taxonomies for skills, are expected to simplify the multifaceted and often confusing landscape of the labour market by presenting to the end-user valuable and exploitable insights. Continuous monitoring will allow for detecting current and future trends and gaps in requested and processed skills. In addition, the platform can act as a stepping stone for recommendation tools to leverage its services and operations in demystifying the labour market and analyzing requested skills.

The forecasting and skills matching capabilities of SKILLAB, constitute an important aspect and contribution of the project, as job vacancies are frequently characterized by their requirements and features, which include specific skills. However, these skills are not always properly defined or concern broad concepts that may be irrelevant to the actual job requirements [McGuinness et al. 2018; Restrepo 2015; Vandeplas et al. 2019]. Thus, tracking emerging skills, digital or non-digital, and properly incorporating them to job descriptions can be a challenge for organizations, particularly during times of technological change. The involved parties (policymakers,

---

[1]https://skillab-project.eu/

[2]https://www.cedefop.europa.eu/en
[3]https://www.ela.europa.eu/en

higher educational institutions, and research centers, in addition to industrial stakeholders), should also be constantly perceptive and with a high understanding of shifts in industry and labour market needs, in order to invest in specific skillsets that ensure maximum profits and workforce engagement.

The notable mismatch in the supplied skills of the workforce and the demanded skills of companies is rendered even more apparent when considering the role of educational institutes in this process. Modern curiculla are often unable to keep up with latest developments in technologies, offering skills that, though essential, cannot guarantee that graduates will get absorbed in the labour market. Hence, curiculla are in dire need of resturcturing in order to keep up with the ever growing digital transformation, which puts digital skills in the forefront as necessary prerequisites of the labour market. In addition, while striving to provide upskilling initiatives and programmes to their personell, companies and organizations struggle with the lack of effective guidelines which differ based on sectors, regions and context, thus ensuring that each workforce is trained according to the needs of key policymakers.

With all the aforementioned factors in mind, SKILLAB acts as a bridge between EU citizens, companies and educational institutions by offering an intelligent platform that can effectively forecast skill gaps and promote digitization. The services of SKILLAB will facilitate the journey of digital transition and trasnform organizations and educational institutions into hubs of updated knowledge that understand the labour market and move at the same pace, rather than lagging behind it. In addition, given that the core concept of SKILLAB relies on the efficient retrieval and processing of data on skills supply and demand, the possibility of creating a SKILLAB DS which will encompass all functionalities and datasets of the project will be examined thoroughly.

Overall, from a practical standpoint SKILLAB will deliver a flexible platform that will be able to forecast and predict skill shortages and will offer operations for:

- Monitoring the developments in skills demand and supply.
- Timely identification of present and future skill shortages that exist within an organization, but also in terms of a specific sector, country or timeframe.
- Discovery of individualized learning pathways that concern upskilling and reskilling.
- Formulation of hiring strategies along with actionable and insightful policies that be leveraged by HR departments and policymakers.
- Advancing the knowledge base of tertiary education and vocational training institutes by offering curiculla recommendations.
- Providing effective suggestions for policymaking on skills shortages and mismatches.
- Pinpointing the ideal person for a position, thus investing on job retainability and talent acquisition.

3 SKILLAB & DATA SPACES

In this Section, we will discuss the relationship of SKILLAB with DS and how can the architecture of the project be perceived as a DS. Firstly we discuss the concept of DS complemented by other initiatives on skills DS. Then we present the overall architecture with every component of the SKILLAB platform. Finally, we analyze potential ways that SKILLAB fits into this perspective while also outlining some potential ethical aspects of a SKILLAB DS.

3.1 Data Spaces - Relation with Skills

DS are an emerging concept that has been brought to the attention of the European Union and the international community as a novel way of data sharing, data access and information processing [DS book]. At their core, DS are secure environments that support data exchange between involved parties[4].

The technologies behind DS include many established and novel principles, such as blockchain, tokenomics, federated learning and semantic models [Designing DS]. The application of natural language processing, decentralized concepts and open science principles allow DS to gain flexibility and to improve data privacy, data trust and security. In this context, the SKILLAB project can be transformed into a DS by incorporating decentralized technologies, federated learning or open data sharing in its services, with emphasis on bridging skill gaps and mismatches.

In recent years, DS have been constantly in the forefront of the plans of the European Union, in areas such as health[5], digital economy[6], tourism[7] and the overall support of governmental procedures[8]. Particulary in the area of skills development, the European Union plans to launch a DS for Skills[9] to support open data sharing on skills, upskilling and reskilling, linking initiatives on skills development with the efficient exchange of data between multiple parties. Furthermore, to support the development of the DS For Skills, specialized projects have been launched, targeted at skills in different domains, including cybersecurity competence centers, health and business among others[10].

To that end, the development of a SKILLAB DS could benefit this effort and greatly improve the landscape of the labour market. SKILLAB as a DS could participate in these ongoing initiatives for open data sharing on skills development, under the principles of the European Union.

3.2 SKILLAB Architecture

In Figure 1, the architecture of SKILLAB is presented. The platform is designed with effective data sharing between components and between involved participants in mind. The goal of SKILLAB is to link EU citizens (the active and potential workforce) with potential employees, organizations, policymakers and stakeholders, hence the information provided by the platform is accessible from everyone to be viewed, analyzed, filtered and interpreted. However, the design of the platform ensures that data are anonymized and processed before entering the DS.

At the center of the architecture are the SKILLAB Trackers, i.e., flexible services that collect multi-source data regarding the labour

---

[4]https://www.reach-incubator.eu/what-are-data-spaces/
[5]https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-spaceen
[6]https://internationaldataspaces.org/
[7]https://projects.eclipse.org/projects/technology.edc
[8]https://gaia-x.eu/
[9]https://www.skillsdataspace.eu/
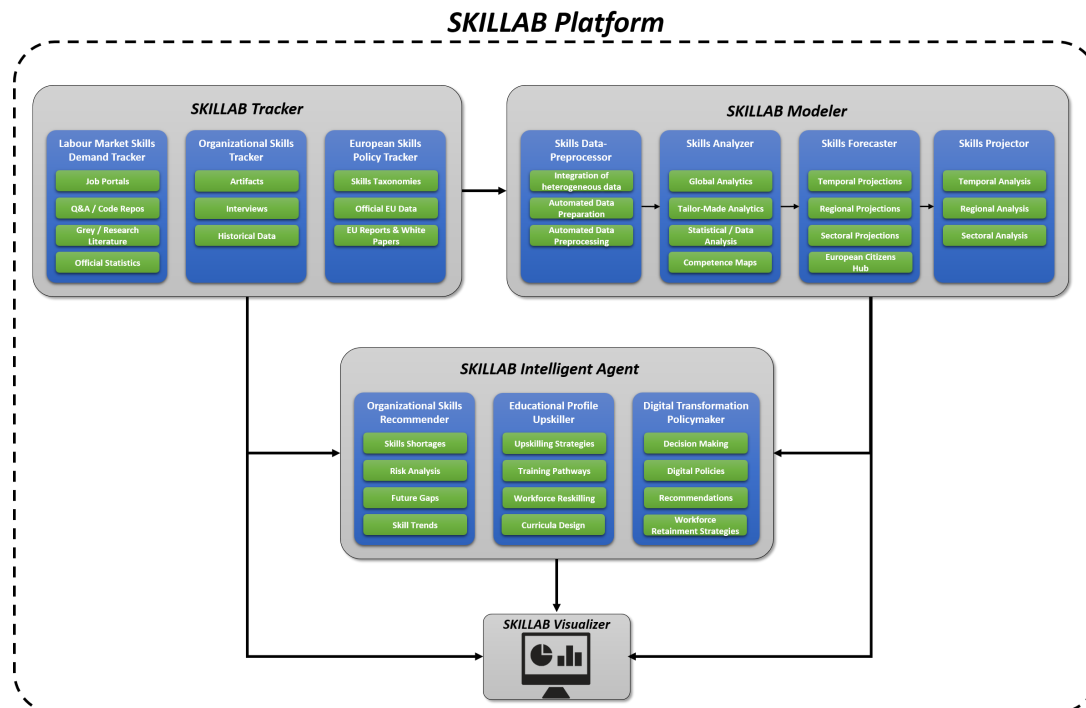[10]https://advancedskills.eu/specialised-projects/

Fig. 1. Overview of SKILLAB architecture

market. More specifically, the Labour Market Skills Demand Tracker is dedicated in collecting information directly relevant to skills demand and supply from job vacancies portals, portals of code and grey literature as well as official statistics dashboards[11], directly from European Union data. In addition, the Organizational Skills Tracker will collect intra-organizational data, ranging from artifacts and interviews to reports, evaluations, questionnaires and surveys. Finally, the European Skills Policy Tracker focuses on the governmental and policymaking aspect of skills development, gathering information from European portals and white papers. The entirety of the SKILLAB Tracker has the potential to operate as a DS, where data are shared between components and users of the platform have the ability to retrieve any data they want.

The information is then processed by the Skills Data-Preprocessor to be turned into intepretable structures and passed to the Skills Analyzer, where several statistical analyses are conducted, based on competence matrices and ML algorithms for forecasting and skills matching. Overall this component serves primarily as the backend service of SKILLAB and the corresponding DS, where operations on data are applied continuously, enhancing the extracted insights. Complementary to the Skills Analyzer, the Skills Forecaster operates as a predictor of future skill gaps, future trending competences and shifts in industry. Via the use of this component, SKILLAB is able to timely detect changes in the digital and skills landscape, inform the involved parties and recalibrate the data that are provided in the DS. Finally, the Skills Projector functions as a filtering pipeline

that allows the application of filters on the accumulated data and insights based on a specific sector, country or timeframe.

All the proccessed and accumulated data is then transferred and fed to the SKILLAB Intelligent Agent, which offers three basic components, concerning different stakeholders and involved parties. These components are individual in nature and serve different purposes. The Organizational Skills Recommender is targeted towards HR departments, organizations and stakeholders that wish to be informed on crucial skills, skill shortages, and gaps in order to formulate hiring strategies and improve workforce engagement and retainment. In contrast, the Educational Profile Upskiller is directed to EU citizens and educational institutions. Its main functions are the provision of educational pathways to citizens, the suggestion of plans for curiculla development, as well as the planning of strategies for upskilling programmes and initiatives. Finally, the Digital Transformation Policymaker is the core component for policymakers, offering practical and actionable insights regarding the development of policies concerning upskilling, reskilling, and the European labour market.

Data sharing could be feasible between the three components and between involved members (EU citizens and organizations), with organizations using the SKILLAB platform to explore data and insights, publicizing their requested competencies and looking for ideal candidates. On the other hand, EU citizens could benefit from the SKILLAB platform by interacting with organizations and their requested competencies, discovering learning pathways and being rapidly absorbed by the labour market. This concept of data sharing,

---

[11]https://ec.europa.eu/eurostat

common access to data, and communication between involved parties constitute the basics of a DS and will be thoroughly explored in the design of SKILLAB.

3.3 SKILLAB as a Data Space

Based on the presented architecture and the principles of open science and data sharing that it advocates, the SKILLAB project could be transformed into a DS, allowing access to its data, while also securing that the processed information is anonymized and ethical guidelines are followed.

Overall, the SKILLAB platform is designed to operate as a standalone framework that is able to be delivered on stakeholders, policymakers and institutions, with data retrieval and storage capabilities. Simultaneously, SKILLAB will be offered as a web service in EU citizens, with all mined information stored in a secure DS, where users will be able to access parts of data, extract insights and view selected outputs. In addition, the SKILLAB platform will be able to allow organizations and citizens to be linked based on requested and supplied skills, further enhancing the data sharing aspect and the principles of open data sharing of DS[12]. In addition, data interoperability and sovereignty are key components behind the design of SKILLAB, with heavy focus both on the interconnected nature of the retrieved information as well as the proper anonymization processes. This ensures that each citizen will be treated equally in this data sharing philosophy and will be offered the same information as everyone else.

As a broader concept, the SKILLAB ecosystem of pilots, collaborators, and supporting companies will continuously engage in information and data exchange, supporting the three pilot cases of the project. To that end, a DS between the SKILLAB consortium and collaborators is expected to be created, encouraging open data sharing within the developed components.

When examining the SKILLAB solution and its possible evolution into a DS from a more technical perspective, the modules and components should adhere to international standards and models that form a DS. More specifically, these international standards are designed to ensure that DS can be used by wider audiences, with interoperability in mind. To that end, reference architectures are introduced that define the design process of a DS. One of the most well known reference architectures is the International DS (IDS) model[13], where several layers are defined when designing a DS. Those layers are:

Business Layer: This layer is responsible for defining the roles of each participant in the DS, while also specifying the interactions between participants. Hence, it can be used for the formulation of interaction strategies and plans. In the context of SKILLAB, the DS and by extension the Business Layer would have four distinct participant roles (Citizen, Industrial, Educational Institution, Policymaker), with interactions between them for data exchange and sharing.

Functional Layer: This layer includes the functional requirements of the DS, which in the case of SKILLAB are translated to the functional requirements of the developed platform and service.

---

[12]https://data.europa.eu/en/publications/datastories/when-open-data-meets-data-spaces

[13]https://internationaldataspaces.org/offers/reference-architecture/

Based on the IDS model, this layer contains six requirement categories which reflect different aspects of the DS. The Trust pillar defines the access rights and capabilities of each role in the DS. Simultaneously, it provides user certification procedures. In SKILLAB, each role will have different capabilities and access rights, while a profiling process will ensure user certification.

The Security and Data Sovereignty pillar is dedicated to the definition of secure protocols for data processing, policies for usage and storage. This pillar will be implemented in SKILLAB as all data will be anonymized and stored in secure databases, while users will have access to them when using the platform.

The next pillar (Ecosystem of Data) refers to services that aim at the management and identification of different data sources, with specialized brokers and routines. The SKILLAB Tracker and Skills Preprocessor components of the platform will be dedicated to this purpose, with the first identifying multifaceted data from various sources and collecting them on regular based and the second deploying several scripts for preprocessing and storage. This procedure will be facilitated by Vocabularies containing skills hierarchies and semantic information that should be taken into account.

As far as Standardized Interoperability is concerned, SKILLAB will ensure that users utilizing the platform will be able to retrieve and view data from the storage components in fixed and standardized ways along with being able to connect with each other for data exchange. In addition, each user will be able to log in to the platform with his/her own credentials and use it on privatized sessions.

The SKILLAB platform is highly relevant with the next pillar (Value Adding Apps), which involves the development and deployment of services for data processing, transformation and analysis as well as the implenentation of Data Apps that contain interfaces, data analytics capabilities and dependencies. SKILLAB will implement the SKILLAB Modeler and SKILLAB Intelligent Agent components which, along with the SKILLAB Visualizer will offer advanced data analytics, forecasts, projections and processing. In addition, all users of the platform will be able to install or use these components as part of the platform, in a unified ecosystem of Data Apps.

The final layer (Data Markets) involves the potential exploitation of the SKILLAB DS for commercial purposes along with the legal aspects of using such a DS. While SKILLAB is a research project, an exploitation plan will be developed during the proposal's lifecycle that will thoroughly explore these concepts and formulate the best strategies to act upon them.

Information Layer: The Information layer involves the definition of Vocabulary Units that define the data sources used by the DS, the representation of the DS architecture in appropriate formats (e.g. UML diagrams) as well as the portrayal of ontologies and relationships between data in programming languages, queries or entities. The concept behind this layer is that users are able to access the Information Layer through Data Apps, understand the inherent relationships between the data and the common Vocabulary. In SKILLAB, the storage components will have carefully designed architecture schemas that aim at optimal data management while querying mechanisms will be present in every component and module (SKILLAB Tracker, SKILLAB Modeler, SKILLAB Intelligent Agent, SKILLAB Visualizer), showcasing the structure of the

backend services and encouraging users to pose different queries, according to their needs.

Process Layer: The process Layer specifies the interactions between the components of the DS, in terms of data sharing, data analysis with Data Apps, exchange of information and granting of access to users as well as searching capabilities. In that sense, the components of SKILLAB will frequently exchange and share data between one another for efficient analysis and forecasting, while users will be able to search, filter and narrow down the collected and processed data. These processed will be clearly defined in the SKILLAB Requirements and Architecture, which are deliverables of the project.

System Layer: This is a holistic layer, encompassing the services of the aforementioned layers. It is comprised of connective routines between participants, brokers for metadata and data exchange, resource management and analysis modules, as well as processing components. The entirety of the SKILLAB platform operating as a unified and individualized service and DS constitutes the System Layer of the project. Based on the definition, it will contain interactions between different roles, data collection routines from the SKILLAB Tracker, analysis and processing modules from the SKILLAB Modeler and practical insights from the SKILLAB Intelligent Agent. These services altogether form a fully functional DS for skills demand and supply analysis, with concurrent communication between involved parties.

Finally, based on the IDS model and the structure of SKILLAB as a DS, the project primarily undertakes the role of (a) **the connector**, with the purpose of providing technical components and services required for interested participants and parties (e.g. data providers, consumers, intermediaries) to join the DS;(b)**the Data App**, which includes applications deployed inside the connector to facilitate data processing and analysis, via the SKILLAB Tracker, SKILLAB Modeler and SKILLAB Intelligent Agent;(c)**the Services** which refer to software running in a connector, which in the context of SKILLAB DS includes all interfaces, routines, modules and components that form the SKILLAB platform and ensure its proper operation.

While SKILLAB as a DS has many practical applications and uses, the project could also contribute to the expansion of the IDS model, for a better understanding of DS in terms of the skills that are required to develop them. A potential avenue to achieve this would be the identification of the core technologies used to develop the IDS modules and leverage the ESCO taxonomy to extract the respective skills. This way, the SKILLAB DS could provide, apart from a functional service for interested parties, an overview of DS skills and their matching to IDS modules, allowing for rapid input on the subject by users of the SKILLAB platform.

3.4 Ethical Considerations of the SKILLAB Data Space

Beyond the design of SKILLAB as a DS, which favours open data sharing and exchange, some ethical aspects must be considered, as the exchanged data of SKILLAB may refer to well known entities and organizations. Some ethical considerations that will be considered and outlined in the design of the SKILLAB DS and subsequent deliverables are:

Data Ownership: In the SKILLAB DS, data ownership is a sensitive issue, as the exchange of data between involved parties in a shared space may create conflicts. Hence, any DS that will be designed should properly define which data belong to the SKILLAB DS as an outcome of its data collection and retrieval modules and which data can be freely accesed and exchanged from the participants defined in the Business Layer.

Transparency: SKILLAB as a project and DS plans to adhere to European Union guidelines for FAIR (Findable - Accessible - Interoperable - Reusable) data practices as well as transparency. The handling of the retrieved data and the training processes of the algorithms will be fully described and explained, while the implemented strategies for data collection, secure storage and analysis will be outlined in detail. Finally, the coding practices and training procedures of the data retrieval and data analysis NLP-ML-LLM algorithms must be showcased to ensure that the DS operates in an ethical manner.

Privacy: As SKILLAB is a project that retrieves data on skills demand and supply, the data storage components will be equipped with anonymization modules that convert the retrieved data into private structures that do not give away sensitive information. Overall, the storage components will be designed to handle data in secure ways which ensure that no data leakage will take place during the implementation of the project and subsequent DS.

Intention: The SKILLAB DS must carefully describe the intents of data use by the various components and collaborating partners, the involved parties of the DS and the potential uses after the end of the project. This will allow for transparency and clarity on data management and data storage.

Ethical Use of Algorithms: The Data Apps of the SKILLAB DS will exploit several Natural Language Processing (NLP) and ML algorithms. Hence, the design of SKILLAB should incorporate ethical guidelines on the writing, training and exploitation of these algorithms, which should be evaluated by humans regarding their outcomes, unbiased predictions and ethical use of data.

4 PILOT CASES AND DATA

SKILLAB is a platform comprised by multiple interconnected components, which rely on exchanging data, knowledge, insights, projections and feedback both between them but also between involved parties. The enviosioned platform and its potential to be expanded into a DS offers multiple different benefits to the target groups that will mainly use it, while the prospect of open data sharing, in accordance with EU principles is a favourable outcome. In order to properly showcase the potential of SKILLAB and its usage by multiple parties, as well as its connection with DS, three pilot cases have been designed. These pilot cases are implemented by three partners of the SKILLAB consortium (VILABS LTD - VIL, Eclipse Foundation Europe - ECL, Asociatia Cluj IT - CLUJ) and offer different objectives, challenges and expected outcomes from the SKILLAB platform.

In the next subsections the proposed use cases, the type of data that the pilot provides and exploits from the SKILLAB project and the components are presented, while some indicative suggestions

on how a SKILLAB DS could be leveraged in each pilot case are also provided.

4.1 VIL Use Case

This pilot case, led by VIL, aims to provide a proof of concept regarding the SKILLAB's platform societal impact in what regards bridging the regulatory, labour and education decision-making with the data-driven monitoring and forecasting of labour market by the SKILLAB's solutions. The main concept of the pilot case is skills anticipation processes and how SKILLAB can facilitate the timely identification of present and future skill gaps and requirements of the labour market, and effectively share its data with public and private sector bodies to prompt greater alignment of skills match within the labour market.

To this end, VIL will act as the liaison between SKILLAB project, and its envisioned DS, and the interested stakeholders, who are also identified as key data sources regarding the skills demand and supply, in two countries, that is Greece and Cyprus. Being based strongly within the innovation ecosystems of these countries, VIL will seek to collect quantitative and qualitative data on the labour market intelligence and what kind of structural procedures are employed in the two pilot sites. Bringing such data back to the SKILLAB platform, along with a socio-economic impact assessment enabled by SKILLAB solutions, VIL, in this pilot case, will map skill mismatches and put forward systematic recommendations for adopting policies and education plans to upskilling, reskilling and overall digital transition of the economy.

The development of a DS can greatly aid the design, integration and deployment of this pilot case. By offering an open space for data sharing, with full anonymity, the agencies of Greece and Cyprus will be able to effectively analyze the insights of SKILLAB, under the guidance of VIL. Hence, the process of decision making will be greatly accelerated and the first outcomes will be delivered in a short timeframe.

4.2 ECL Use Case

The ECL pilot case focuses on the skills reflected in open-source communities and the knowledge regarding the required and supplied skills. This knowledge is currently decentralized and unstructure, while also available in multiple forms (job portals, conferences, coding repositories). This multifaceted nature of the data creates inherent difficulties in extracting skillsets and trending skills. It is thus apparent that ECL and similar companies would greatly benefit from a DS where these types of data would be structured, organized and shared among involved participants and parties for evaluation, skills extraction and community building.

The usage of a SKILLAB DS regarding open-source skills supply and demand would thus accelerate the identification of key skills in open-source projects, the discovery of promising developers and personnel and their subsequent hiring. The potential of applying the principles of DS in this concept constitutes a great added value for open-source foundations and further complements the benefits of open science.

4.3 CLUJ IT Use Case

The final pilot case, concerning CLUJ, is directly targeted towards personell hiring, talent acquisition and the journey towards digital transformation. Currently, the CLUJ IT ecosystem encounters problems in discovering, explaining and pinpointing which skills are necessary for their personnel to be familiar with digital technologies and the basic notions of digital transformation. This inherent difficulty prevents HR departments to issue proper job descriptions and formulate necessary policies on upskilling and workforce training. In addition, these problems are also transferred in ideal candidate identification as well as performance evaluation. Finally, the lack of data on skillsets, competency mapping and skill gaps prevent the efficient organization of workshops and seminars aiming at enhancing the skills of the workforce.

The deployment of this pilot case could leverage the SKILLAB DS by safely accesing the insights provided by the analysis components of the platform and deriving useful recommendation from the SKILLAB Intelligent Agent. In addition, CLUJ could establish internal data sharing between its own ecosystem, external collaborators and the other pilot parrners, establishing building blocks for skills development, brainstorming on novel ideas for training initiatives and strategic planning on hiring for HR departments and education plans for workshops and seminars. Finally, a DS of this magnitude could link the CLUJ ecosystem with ideal candidates for specific positions, encouraging talen acquisition and workforce development.

5 CONCLUSION & FUTURE WORK

In the present paper, we present the SKILLAB platform, a HORIZON Europe project that aims at monitoring the European Labour market and offer practical insights on skill gaps, forecast future skill mismatches and provide reccomendations on HR departments, educational pathways to EU citizens and institutions and policy plans for policymakers. The expected outcomes of SKILLAB concern multiple parties, ranging from academia, industry leaders and governmental representatives to European citizens and policymakers. The deployed methodologies will utilize cutting edge technologies along with established concepts and will be targeted towards effective skills matching and forecasting both in a generic context as well as in specific sectors, regions or timeframes. The mined information that will be used for analysis is comprised of an extensive database that contains job vacancies, user profiles, courses, white papers and academic literature, along with official European Union data, indicators and policy papers.

The objectives of SKILLAB render it an efficient and promising solution to the skill gaps and shortages problem, while offering a flexible tool for workforce development, upskilling and skills development. Simultaneously, SKILLAB supports the transition towards digitalization by pinpointing necessary skills in this direction.

As data play a crucial role in all SKILLAB operations, the concept of DS can be harmoniously combined in the envisioned platform by offering opportunities for safe data sharing and data exchange between involved parties both in a project-wise level (e.g. between pilot partners) but also between different target groups. The wealth

of open information contained in SKILLAB can potentially be transformed into a fully functional skills supply and demand DS that will usher skills development into a new era of maturity and prowess.

Some interesting future work that can be undertaken for the design of the SKILLAB DS and the contributions of SKILLAB to the undesrtanding of DS are the careful examination of the IDS model and its integration in the project deliverables, the matching of DS skills with the ESCO taxonomy and the possibility of extending the DS to include more roles, services and Data Apps.

ACKNOWLEDGMENTS

REFERENCES

Cheng-Yuan Liou, Wei-Chen Cheng, Jiun-Wei Liou, and Daw-Ran Liou. 2014. Autoencoder for words. *Neurocomputing* 139 (2014), 84–96.

Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark. 2024. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756* (2024).

Seamus McGuinness, Konstantinos Pouliakas, and Paul Redmond. 2018. Skills mismatch: Concepts, measurement and policy approaches. *Journal of Economic Surveys* 32, 4 (2018), 985–1015.

Pascual Restrepo. 2015. Skill mismatch and structural unemployment. *Massachusetts Institute of Technology Job Market Paper* 13, 9 (2015), 66–94.

Michael Tschannen, Olivier Bachem, and Mario Lucic. 2018. Recent advances in autoencoder-based representation learning. *arXiv preprint arXiv:1812.05069* (2018).

Anneleen Vandeplas, Anna Thum-Thysen, et al. 2019. *Skills mismatch & productivity in the EU.* Publications Office of the European Union Luxembourg.

Junhai Zhai, Sufang Zhang, Junfen Chen, and Qiang He. 2018. Autoencoder and its various variants. In *2018 IEEE international conference on systems, man, and cybernetics (SMC).* IEEE, 415–419.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023).