

Interoperable educational data: What if data is no longer available?

Kristine Hein^{*}, Katerina Kostadinovska^{*†}

^{*} Federal Institute for Vocational Education and Training (BIBB), Bonn, Germany

[†] TH Cologne, Germany

Email: kristine.hein@bibb.de, katerina.kostadinovska@bibb.de

Abstract—The integration of skill-centric educational data into data warehouses is a key enabler for data-driven analysis in education and labor market research. However, the efficacy of this process is hindered when underlying data sources undergo changes or become partially unavailable. This paper explores the question of how interoperable educational data can be maintained under such conditions. To this end, it uses the transition from the German platforms BERUFENET and Kursnet to meinNOW as a case study. We hereby propose a data engineering architecture that combines large-scale API-based extraction, relational data warehousing, and semantic enrichment using the European Skills, Competences, Qualifications and Occupations (ESCO) framework. The preservation of data continuity is of particular importance, and this is achieved through versioning, schema mapping, and the reconstruction of missing information. The findings indicate that while full continuity cannot be guaranteed due to structural and semantic changes in source systems, a consistent approximation of longitudinal data can be achieved. This necessitates the explicit management of schema evolution and uncertainty in reconstructed attributes. The paper emphasizes the necessity of critical reflection when interpreting longitudinal educational data derived from evolving sources.

I. INTRODUCTION

THE growing significance of lifelong learning and the rapid transformation of labor markets have heightened the necessity for interoperable educational data infrastructures [1], [2]. Educational offerings are increasingly evaluated not only in terms of institutional characteristics but also in terms of the skills they convey [3]. However, extant data sources such as Kursnet and meinNOW reveal heterogeneous and loosely structured representations that are not directly conducive to analytical processing.

The objective of this paper is to address the research question of how skill-centric data can be integrated as interoperable data into a data warehouse if data is no longer available. The work is grounded in a real-world data engineering project involving the large-scale extraction of educational provider and course data. Automated pipelines have been developed to extract raw data via APIs, transform it into structured formats, and persist it in relational models. The extraction process revealed constraints such as API pagination limits and performance bottlenecks, requiring adaptive filtering and parallelization strategies.

It is imperative to acknowledge the evolution of the core challenge from syntactic integration to semantic interoperability, a transition that is predicated on the solid technical

foundation that has been established, see for example [4]. To address this issue, the ESCO is utilized as a central reference model for skills. However, the evolution of ESCO and the subsequent deprecation of its portal introduce challenges for long-term system stability, see [5], [6]. To address these challenges, architectural considerations for versioning and decoupling must be implemented.

Additionally, BERUFENET, a platform run by the Federal Employment Agency in Germany, was replaced by meinNow. As Figure 1 shows, the data itself seems to be the same. However, there are several issues with data that is no longer available, which we will discuss in this paper. The main research question of this paper is: Can we use longitudinal online data if the data changes or becomes unavailable?

The remainder of this paper is organized as follows. In Section II, an examination is conducted of extant scholarship pertaining to ontology-based data integration, skill representations, and educational data warehousing. Section III delineates the methodological framework, encompassing the acquisition, transformation, and evaluation of data. In Section IV, the system architecture is presented, delineating its layered design and the integration of semantic components. Section V evaluates the approach with a particular focus on data continuity and the transition from BERUFENET to meinNOW. Finally, Section VI concludes the paper and discusses implications as well as directions for future research.

II. RELATED WORK

The integration of heterogeneous educational and labor market data into interoperable systems has been a subject of extensive discussion across numerous research domains, including data warehousing, semantic web technologies, and skill-based information systems, see [7], [8]. This section undertakes a review of the extant literature on the subject, focusing on three primary streams: ontology-based data integration in data warehouses, skill ontologies and labor market information systems, and recent advances in automated skill extraction and representation [9].

A central challenge in modern data integration lies in the heterogeneity of data sources and the lack of shared semantic structures [10]. Conventional data warehouse methodologies, while efficacious in the integration of structured data, are constrained in their capacity to incorporate semantically

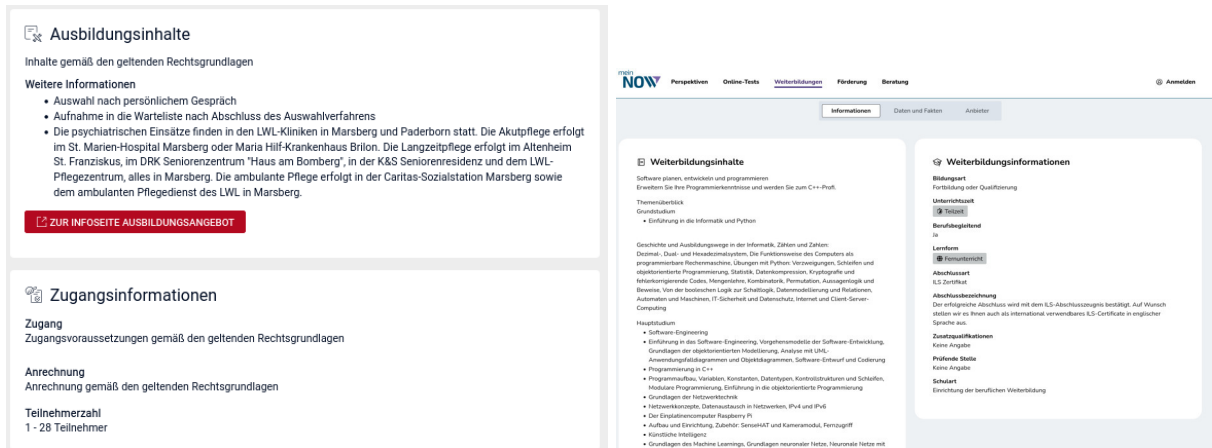


Fig. 1: Metadata as shown in Berufenet (left) and meinNow (right)

rich and unstructured data. In order to address this limitation, ontology-based approaches have been proposed as a means to enhance interoperability and semantic consistency. A systematic literature review indicates that ontologies are being used with increasing frequency throughout the data warehouse lifecycle, including in the stages of requirement analysis, dimensional modeling, and ETL processes. Their primary contributions are threefold: first, resolving semantic heterogeneity, second, enabling data integration, and third, improving analytical capabilities through semantic enrichment and inference mechanisms.

Ontology-based data warehouses augment conventional architectures by integrating formal knowledge representations. These approaches enable the alignment of heterogeneous schemas and facilitate the integration of distributed data sources into a unified analytical framework. In particular, ontologies function as intermediary layers that map source data to conceptual models, thereby supporting both syntactic and semantic interoperability. Furthermore, semantic OLAP approaches have been shown to enhance querying capabilities, wherein reasoning mechanisms can derive implicit knowledge from explicitly stored data, see [11], [12].

In the context of labor market and educational data, the necessity for semantic interoperability is particularly evident, see [13], [14]. Our previous works introduced the concept of linked labor market data and highlighted the challenges associated with integrating diverse datasets, including surveys, training data, and employment information. A salient observation is the absence of standardized structures for data collection and representation, which hinders integration efforts and curtails analytical potential. The paper emphasizes the role of ontologies such as ESCO and the German Labor Market Ontology (GLMO) as foundational components for achieving interoperability [10].

ESCO has become one of the most widely adopted ontologies in the European labor market domain, see [15]. It provides a multilingual classification of skills, competences, qualifications, and occupations, thereby establishing a common

vocabulary across institutions and countries. The objective of ESCO is to establish a connection between educational institutions and the workforce by facilitating the consistent articulation of skills and the facilitation of communication among relevant stakeholders. However, while ESCO offers a comprehensive taxonomy, its practical integration into data warehouse systems remains challenging, particularly in light of evolving data access mechanisms and structural updates.

In addition to ESCO, several domain-specific ontologies have been proposed to address the representation of skills and educational content. The Skills and Recruitment Ontology (SARO) is an extension of the Employment and Skills Credit Ontology (ESCO) that facilitates the automation of skills extraction and normalization from unstructured text. This is achieved by establishing a link between job postings and skill concepts, as outlined in the work by [16]. In a similar manner, the EduCOR ontology furnishes a consolidated depiction of educational resources and facilitates customized learning and career guidance by correlating learning materials with labor market competencies, see [17]. These approaches demonstrate the increasing importance of ontology-driven representations in skill-based information systems.

Another significant research direction pertains to the automated extraction and standardization of skills from textual data. Early systems relied on rule-based or statistical methods, but recent advances leverage machine learning and deep learning techniques. For instance, systems for skill identification and normalization employ ESCO as a reference taxonomy and evaluate performance using precision and recall metrics derived from sampled data [18]. Recent approaches have incorporated transformer-based models and contrastive learning techniques to facilitate the mapping of unstructured text to ontology concepts. This integration has yielded substantial enhancements in scalability and accuracy, see [19].

Notwithstanding these advances, several challenges remain unresolved. Firstly, the alignment between heterogeneous data sources and standardized ontologies is often ambiguous and requires sophisticated matching techniques. Secondly, the dy-

dynamic nature of labor market data necessitates continuous updates to ontologies and mappings, thereby raising issues of versioning and consistency. Thirdly, the integration of semantic technologies into data warehouse architectures remains non-standardized, exhibiting a range of modeling, storage, and querying approaches.

In summary, an analysis of the extant literature reveals a clear trend towards the integration of ontologies and semantic technologies into data warehouse systems, particularly in domains characterized by heterogeneous and unstructured data. Information systems that are skill-based represent a research area that is evolving rapidly. Ontologies such as ESCO play a central role in enabling interoperability in this area. However, the practical realization of fully interoperable, skill-centric data warehouses remains an open challenge. This paper contributes to the extant research by proposing an architecture that combines large-scale data extraction with ontology-based semantic enrichment. It explicitly addresses the integration of evolving standards, such as ESCO, into a stable data warehouse environment.

III. METHODOLOGY

The present study employs a structured data engineering pipeline comprising four primary stages: The following four steps were taken in the research process: (a) data acquisition (scraping), (b) integration into a skill-centric data warehouse, (c) evaluation of the pipeline, and (d) analysis of challenges related to the underlying data schemas, particularly those of the meinNOW platform. The overarching objective is to facilitate large-scale extraction, transformation, and semantic enrichment of continuing education data for subsequent analytical use. We refer to Section V for a more detailed discussion of the data itself.

A. Data Acquisition

The acquisition of data is facilitated by the utilization of REST APIs, which are publicly accessible and underpin the Kursnet and meinNOW platforms. While both platforms offer user-friendly web interfaces, these interfaces depend on backend API endpoints that can be accessed directly via HTTP requests. Access to these endpoints necessitates the incorporation of specific HTTP headers, including API keys, for the purpose of authenticating requests.

The extraction process commences with querying the API endpoints, thereby retrieving data in JSON format. It has been determined that, given the substantial quantity of accessible records and the constraints imposed by the APIs, the retrieval of data in a single request is not a viable option. Instead, a paging and filtering strategy is employed to iteratively collect data while respecting request limits.

A significant challenge during this phase is the restriction on the maximum number of records returned per query. For instance, Kursnet is subject to limitations imposed by the API, which are defined by the number of pages and the number of results returned for each query. To address this issue, custom filtering mechanisms were developed to partition the data

space into subsets that remain within the API's constraints. These filters are systematically applied across providers and course offerings to maximize coverage.

The raw responses from the API are initially stored in a staging layer in JSON format. This approach guarantees the integrity of the original data and facilitates its reprocessing if necessary. Subsequently, the data undergoes a process of parsing and preparation for transformation into structured formats.

B. Integration into a Skill-Centric Data Warehouse

Subsequent to data acquisition, the raw JSON data is transformed into a unified relational schema suitable for analytical processing. The transformation process entails the mapping of heterogeneous source structures into a consistent data model, thereby representing core entities such as providers, courses, and associated metadata.

The schema design follows a normalized relational structure, where entities are separated into distinct tables and linked via primary and foreign keys. Redundancies are minimized through normalization, ensuring data consistency and integrity. Key attributes include provider identifiers, course identifiers, temporal information, geographic data, and descriptive text fields.

Textual course descriptions play a pivotal role in the subsequent semantic enrichment phase and are consequently preserved in their original form during transformation. These descriptions are stored alongside structured attributes, enabling downstream natural language processing tasks.

The transformed data is subsequently loaded into a dimensional data warehouse model that has been optimized for analytical queries. This model is characterized by its adherence to a star schema, in which fact tables represent course offerings, and dimension tables represent providers, time, and other contextual attributes. The ETL (Extract, Transform, Load) pipeline has been demonstrated to support both initial full loads and potential incremental updates.

C. Evaluation

The evaluation of the system is conducted along several dimensions, including data completeness, structural consistency, semantic quality, and performance.

The completeness of the extracted dataset is assessed by comparing the number of retrieved records against the total number of records reported by the APIs. This includes evaluating the coverage of both providers and course offerings. The effectiveness of the filtering strategy is measured by the proportion of records successfully retrieved relative to the estimated total.

The validity of the transformation process is ascertained by confirming referential integrity between entities, thereby ensuring that relationships between providers and courses are accurately preserved. Additional validation checks include duplicate detection, schema conformity, and type consistency across attributes.

A fundamental objective of the system is the extraction and mapping of skill-related information from unstructured course descriptions. The evaluation of this component is centered on three key aspects: the accuracy of skill extraction, the correctness of mappings to standardized skill identifiers, and the management of ambiguous terms. Confidence scores are utilized to evaluate the reliability of mappings, and multiple candidate matches are permitted in instances of ambiguity.

The runtime performance of the pipeline is evaluated, with particular attention to the data acquisition phase. The extraction process is computationally intensive due to API rate limits and the need for multiple iterative requests. The identification of performance bottlenecks is facilitated by profiling methodologies, and strategies such as the parallelization of API requests are investigated to enhance throughput. However, hardware constraints, particularly limited memory capacity, can have a substantial impact on execution stability and runtime.

IV. SYSTEM ARCHITECTURE

The proposed architecture employs a layered approach that separates concerns of data ingestion, transformation, semantic processing, and analytics, see [10]. At the ingestion layer, API connectors are responsible for retrieving data from external platforms and storing it in a staging database. This layer has been engineered to exhibit resilience to alterations in API, thereby facilitating incremental updates. The transformation layer plays a pivotal role in converting raw data into a standardized relational schema, thereby ensuring data integrity and consistency. This encompasses entity resolution for providers and the normalization of course data. The design is such that the schema can accommodate changes in source structures without affecting downstream components. The semantic layer integrates ESCO as an external knowledge base. In order to mitigate dependencies on external services, ESCO data is locally replicated and versioned. A mapping service has been developed to link course descriptions to ESCO skills and maintain mapping histories to support reproducibility. The data warehouse layer implements a dimensional model that has been optimized for analytical queries. The system facilitates the aggregation of course offerings by skill, provider, region, and time.

A star schema is derived that places skills at the center of analytical operations. The central fact table, designated as FactCourseSkill, serves to record the interrelationship between courses and skills. Each record in the database corresponds to a particular course offering and a specific ESCO skill, along with pertinent metrics such as relevance score and occurrence frequency. The schema under consideration includes several conformed dimensions. The DimSkill dimension serves as a repository for ESCO skill identifiers, preferred labels, and hierarchical relationships. The DimCourse dimension is a comprehensive category encompassing educational offerings, including title, duration, and modality attributes. The DimProvider dimension is designed to capture information regarding educational institutions. Temporal aspects are addressed

through the utilization of a DimTime dimension, thereby facilitating longitudinal analyses.

This subschema facilitates queries such as identifying the most frequently taught skills, analyzing skill coverage across providers, and detecting temporal trends in skill demand. By conceptualizing competencies as a pivotal analytical dimension, the schema facilitates cross-cutting analyses that would not be feasible in conventional course-centric models.

V. EVALUATION

The evaluation of the proposed system focuses on its ability to ensure continuity during the transition from the outdated BERUFENET platform to the new meinNow system. This includes addressing changes in data availability and metadata representation. In Table I we present the categorial data which is no longer available in meinNow. In Table II we give an overview about new and deprecated metadata. The analysis will examine how the system preserves the integrity of skill-centric educational data, adapts to evolving data models, and maintains usability for stakeholders who rely on historical and current datasets.

A primary focus of the evaluation is ensuring continuity of data and metadata between BERUFENET and meinNow. BERUFENET provided structured taxonomies and skill classifications that were critical for workforce planning and educational analytics. The deprecation of BERUFENET necessitated migrating its data to meinNow. While meinNow offers expanded access to course offerings, it introduced challenges in maintaining the same level of granularity and consistency. Certain BERUFENET categories and skills, such as "Wundmanagement" (wound care) and "Verbandtechnik" (bandaging techniques), were deprecated or reclassified in meinNow. This required the system to reconcile these changes through automated inference and manual curation. The absence of certain BERUFENET-specific fields, such as detailed occupational classifications, underscores the necessity of hybrid approaches that combine rule-based mappings with machine learning to ensure the continuity of skill representation.

Now, we will discuss some aspects in more detail.

A. Technical Aspects

A number of challenges were identified during the process of integrating data from the meinNOW platform. The primary challenge initially encountered pertained to the identification of a viable method for automated data extraction. Despite the platform's provision of a graphical user interface, it was observed that this interface is underpinned by API endpoints that can be accessed directly via HTTP requests, contingent upon the use of appropriate authentication headers. This finding facilitated efficient programmatic access to the underlying data.

Another significant challenge pertains to the limitations of APIs, particularly the restriction on the number of records returned per request. Addressing this issue necessitated an in-depth analysis of the API behavior and the development of filtering strategies that partition queries into smaller subsets,

TABLE I: Categories and Skills which are no longer available in meinNow

Name	Data	Content	Example
Category	Taxonomy	id+label	C 1410-10 C 1025-20-10-35
Skills	Taxonomy	id	030502-022 Computer-/IT-Grundlagen Wundmanagement, Verbandtechnik

TABLE II: Other metadata which are no longer available ($\rightarrow$), partially match ($\leftrightarrow$), or are new ($\leftarrow$) in meinNow

KURSNET		meinNow	
date	$\rightarrow$	beginn	
	$\rightarrow$	ende	
length	$\leftrightarrow$	dauer	Calculated
	$\leftarrow$	kostenWert	
	$\leftarrow$	kostenSort	
	$\leftarrow$	kostenWaehrung	
	$\leftarrow$	individuellerEinstieg	
	$\leftarrow$	veranstaltungsablaufe	
	$\leftarrow$	pruefendeStelle	
	$\leftarrow$	inhouseFirmenseminar	
	$\leftarrow$	berufsbegleitend	
	$\leftarrow$	praxisanteile	
	$\leftarrow$	digitaleBarrierefreiheit	
Anbieter	$\rightarrow$	ansprechpartner	[List]
schulart	$\leftrightarrow$		

each within the allowable limits. These filters were incorporated into the extraction logic and applied consistently across all data retrieval operations.

B. Data Aspects

In contrast to Kursnet, the meinNOW platform offers a relatively comprehensive and unrestricted list of providers. However, discrepancies in the structure of course-related data introduce additional complexity during transformation. These structural differences necessitate adaptations in the mapping logic and the unified schema to ensure compatibility across data sources.

As illustrated in Table II, not all data is discontinued; however, a greater proportion of structured information is available. A determination must be made regarding the specific data that is required and the manner in which it can be substituted. As previously mentioned, the extraction of ESCO-related data is a possibility. Nevertheless, the aforementioned process was contingent upon the utilization of skill data from BERUFENET; consequently, its direct application to meinNow is precluded.

C. Continuing data?

A salient question that emerges from this study is whether a continuous and longitudinal dataset can be constructed in the presence of changing or partially unavailable source data. From a technical perspective, the proposed architecture demonstrates that continuity can be approximated through a combination of versioning, and semantic enrichment. Preserving raw data snapshots and maintaining historical mappings to

external ontologies, such as ESCO, facilitates the reconstruction of earlier states of the data and ensures the reproducibility of analytical results.

Nevertheless, the concept of continuity is not exclusively a technical concern; it is also a conceptual one. A comparison of BERUFENET and meinNow reveals that, even when initial observations suggest similarity, subsequent analysis reveals differences in underlying semantics, granularity, and classification schemes. Consequently, longitudinal analyses must account for these shifts explicitly. This includes the introduction of mapping layers that document transformations between schema versions, as well as the use of surrogate constructs to approximate deprecated attributes.

In consideration of data quality, the transition to meinNow engenders a mixture of advantages and disadvantages. Preliminary observations suggest that such reconstruction is feasible; however, it introduces uncertainty that must be quantified and communicated in downstream analyses. In summary, the findings indicate that continuous data integration is achievable under the assumption that the data remains constant. While the assurance of a fully stable longitudinal dataset remains elusive, the construction of a sufficiently consistent approximation is attainable through meticulous architectural design, explicit modeling of change, and transparent handling of uncertainty.

VI. CONCLUSION AND OUTLOOK

The objective of this paper is to address the question of whether interoperable, skill-centric educational data can be integrated into a data warehouse in situations where source data changes or becomes unavailable. The findings indicate that, from a technical perspective, the construction of a resilient data integration pipeline is indeed feasible. This pipeline needs to preserve a significant portion of the informational content across system transitions.

Conversely, from a social science perspective, the notion of a truly continuous dataset necessitates careful consideration. The modifications in data collection practices, classification schemes, and platform design imply that the observed data is not strictly comparable over time. Consequently, longitudinal analyses necessitate a deliberate examination of the evolving nature of the data and its underlying constructs. Rather than assuming continuity, future research should conceptualize such datasets as "constructed continuities," where consistency is achieved through methodological and technical interventions.

A number of fundamental insights can be extrapolated from the field of computer science. Firstly, it is imperative to decouple data ingestion from semantic interpretation to ensure system robustness. Secondly, versioning mechanisms for both data and ontologies are imperative in environments

characterized by evolving standards. Thirdly, the integration of heterogeneous data sources is significantly enhanced by ontology-based approaches; however, these must be complemented by flexible mapping and reconciliation strategies. Finally, scalability considerations, particularly in the context of API-based data acquisition, persist as a practical constraint that necessitates further optimization.

Beyond the technical contributions, this work underscores the necessity for a more comprehensive theoretical framework that can facilitate interoperability in educational data. A framework of this nature ought to incorporate perspectives from data engineering, semantic web technologies, and social science methodology. The objective of this study is to provide guidelines for handling data discontinuities, documenting transformations, and evaluating the validity of longitudinal analyses.

In the future, scholars should investigate the incorporation of further data sources, such as labor market data and learner trajectories, to enhance the analytical capacity of the proposed system. The incorporation of machine learning methods for dynamic skill extraction and alignment represents a promising direction for future research. Furthermore, the establishment of standardized guidelines—or "design principles"—for the development and maintenance of interoperable, skill-centric data infrastructures has the potential to contribute to the establishment of best practices in this emerging field.

The primary challenge lies not only in the construction of technically proficient systems, but also in the assurance that the resulting data facilitates meaningful and valid interpretations in the domains of educational and labor market research.

ACKNOWLEDGMENTS

During the preparation of this manuscript, the authors used DeepL to check the language and grammar. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

REFERENCES

- [1] M. Udtha, K. Nomie, E. Yu, and J. Sanner, "Novel and emerging strategies for longitudinal data collection," *Journal of Nursing Scholarship*, vol. 47, no. 2, pp. 152–160, 2015.
- [2] S. I. Baek and Y. M. Kim, "Longitudinal analysis of online community dynamics," *Industrial Management & Data Systems*, vol. 115, no. 4, pp. 661–677, 2015.
- [3] A. Fischer and J. Dörpinghaus, "Web mining of online resources for german labor market research and education: Finding the ground truth?" *Knowledge*, vol. 4, no. 1, pp. 51–67, 2024.
- [4] J. Dörpinghaus, J. Binnewitt, D. Samray, and K. Hein, "Understanding informatics in continuing vocational education and training data in germany," *ACM Transactions on Computing Education*, vol. 24, no. 3, pp. 1–22, 2024.
- [5] J. Dörpinghaus, J. Binnewitt, S. Winnige, K. Hein, and K. Krüger, "Towards a german labor market ontology: Challenges and applications," *Applied Ontology*, vol. 18, no. 4, pp. 343–365, 2023.
- [6] D. Martić, A. Fischer, and J. Dörpinghaus, "Extending the german labor market ontology with online data," in *INFORMATIK 2024*. Gesellschaft für Informatik eV, 2024, pp. 2019–2030.
- [7] J. Dörpinghaus, D. Samray, and R. Helmrich, "Challenges of automated identification of access to education and training in germany," *Information*, vol. 14, no. 10, p. 524, 2023.
- [8] A. Dutt, M. A. Ismail, and T. Herawan, "A systematic review on educational data mining," *Ieee Access*, vol. 5, pp. 15 991–16 005, 2017.
- [9] S. K. Mohamad and Z. Tasir, "Educational data mining: A review," *Procedia-Social and Behavioral Sciences*, vol. 97, pp. 320–324, 2013.
- [10] K. Hein, "Linked labor market data: Towards a novel data housing strategy," in *2024 19th Conference on Computer Science and Intelligence Systems (FedCSIS)*. IEEE, 2024, pp. 355–362.
- [11] R. M. Chang, R. J. Kauffman, and Y. Kwon, "Understanding the paradigm shift to computational social science in the presence of big data," *Decision support systems*, vol. 63, pp. 67–80, 2014.
- [12] I. Moalla, A. Nabli, L. Bouzguenda, and M. Hammami, "Data warehouse design approaches from social media: review and comparison," *Social Network Analysis and Mining*, vol. 7, no. 1, p. 5, 2017.
- [13] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert systems with applications*, vol. 33, no. 1, pp. 135–146, 2007.
- [14] J. Dörpinghaus and M. Tiemann, "Vocational education and training data in twitter: Making german twitter data interoperable," *Proceedings of the Association for Information Science and Technology*, vol. 60, no. 1, pp. 946–948, 2023.
- [15] European Commission, "Esco – european skills, competences, qualifications and occupations," <https://esco.ec.europa.eu/>, 2023, accessed: 2026-03-25.
- [16] E. M. Sibarani, S. Scerri, C. Morales, S. Auer, and D. Collarana, "Ontology-guided job market demand analysis: a cross-sectional study for the data science field," in *Proceedings of the 13th international conference on semantic systems*, 2017, pp. 25–32.
- [17] E. Ilkou, H. Abu-Rasheed, M. Tavakoli, S. Hakimov, G. Kismihók, S. Auer, and W. Nejdl, "Educor: An educational and career-oriented recommendation ontology," in *International Semantic Web Conference*. Springer, 2021, pp. 546–562.
- [18] S. M. Pryima, O. V. Strokan, J. V. Rogushina, A. Y. Gladun, and A. Mozgovenko, "Ontology-based methods and tools for validation of non-formal learning outcomes," in *UkrPROG*, 2020, pp. 50–60.
- [19] S. Criscuolo, D. Spoladore, A. Mahroo, E. Conti, B. Volta, F. Isgro, P. Arpaia, and M. Sacco, "Towards machine learning-based ontology mapping to bridge o* net and esco skills," in *2025 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRINE)*. IEEE, 2025, pp. 299–304.