Journal of Open Humanities Data

Open Public Vocational School Data for Germany

DATA PAPER

INES MARGARETE LOLL

ubiquity press

ABSTRACT

Public dual vocational schools in Germany as of 2025 were systematically modelled as individual Wikidata items linked to their offered training occupations and corresponding German occupational codes (KldB). Registries from all 16 federal states were harmonised and enriched with school website information. Each item includes geographical coordinates, training occupations, the person the school is named after (if applicable), multi-campus structures, and external identifiers. A total of 1,442 vocational school items are available on Wikidata and as a curated dataset on Zenodo. The vocational school dataset supports spatial analyses, including commuting patterns and regional coverage of training occupations, and it allows studies on school naming.

CORRESPONDING AUTHOR:

Ines Margarete Loll

Federal Institute for Vocational Education and Training, Bonn, Germany; Research Centre for Education and the Labour Market, University of Maastricht, Maastricht, Netherlands

ines.loll@maastrichtuniversity.nl

KEYWORDS:

Wikidata; vocational schools; vocational education and training

TO CITE THIS ARTICLE:

Loll, I. M. (2026). Open Public Vocational School Data for Germany. *Journal of Open Humanities Data*, 12: 25, pp. 1–6. DOI: https://doi.org/10.5334/johd.422

(1) OVERVIEW

REPOSITORY LOCATION

https://doi.org/10.5281/zenodo.17767966

CONTEXT

Working with German administrative education data often means handling datasets from 16 separate federal statistics or education offices. For vocational education research, data on vocational schools form a central empirical foundation. Each state publishes its school registry in a different format and level of detail. In some states, access to machine-readable registries of vocational schools requires payment. Furthermore, information on which training occupations are offered at each vocational school is not available in machine-readable form. However, combining geospatial school data with occupation-level information can provide valuable insights for a range of research and policy questions.

This project addresses this gap by making information about public vocational schools across all federal states findable, accessible, interoperable, and reusable (FAIR). To achieve this, school information was harmonised and made available in two forms. First, this information was added to the linked open data environment, Wikidata, which is FAIR (persistent Q-IDs, open interfaces, RDF export, CC0-licensed, machine-readable) by design (Vrandečić et al., 2023). Second, based on these Wikidata items, a curated dataset was extracted and published on Zenodo, with each school represented as one observation.

Vocational schools in this context were defined as part-time vocational schools (*Berufsschulen*) within the dual system, which provide the school-based component of apprenticeship training for occupations regulated under the Vocational Training Act (BBiG) or the Crafts and Trades Regulation Code (HwO). Full-time vocational education and training (VET) types (e.g., *Berufsfachschulen*, health schools, transition programmes) were excluded, following the distinction of Deissinger (2018) between part-time and full-time vocational schools in Germany. Vocational training centres (*Berufsbildungswerke*) for students with mental or physical disabilities were also excluded.

The dataset was created as part of Ines Loll's doctoral project and has not yet been used in another paper.

(2) METHOD

STEPS

First, official registries of public vocational schools were obtained from the statistical or educational authorities of all German federal states. These sources were available as Excel, CSV, or raw text files on the websites of the educational authorities. These datasets were harmonised and merged with information from jedeschule.de (snapshot April 2025). *Jedeschule.de* is an open data web scraping project initially launched by the Open Knowledge Foundation and BildungsCent, which collects and aggregates official school data (Code for Germany, 2025). All sources and data per federal state are uploaded to a GitHub repository[1] (see Supplementary File 1). During data preparation, for instance, geocoding, the jedeschule.de identifier, and aliases were added. Second, each state-level dataset was processed separately in OpenRefine (De Wilde & Verborgh, 2013) to reconcile schools against existing Wikidata items. OpenRefine was then used to generate QuickStatements for creating or enriching entities, followed by manual refinement. This workflow aligns with common practices in other disciplines (e.g., Braisher & Fitchett, 2025; Mihindukulasooriya et al., 2025; Thiery, 2022). The following information is provided for each school: school name, address, coordinates, federal state, training occupations, official website, and *jedeschuleID*. When applicable, additional details were included: the name of the person after whom the school is named, the year of foundation, and the existence of satellite campuses. The final dataset available on Zenodo includes information on 1,258 main campuses and 184 satellite campuses, with each school represented by a single row (observation). Missingness occurs mainly because 61.68% of schools are not named after a person, 17.60% lack a *jedeschuleID* (satellite campuses have no own ID and some schools in Bavaria are not covered), 11.16% share a website with their main campus, and 82.74% have

1 https://github.com/Bibroa/Vocational_School_Germany/.

no foundation year yet. As the project is ongoing, foundation years will be supplemented in future updates.

Information was primarily derived from the official registries. Where data were missing, data from *jedeschule.de* and school websites were used. The training occupations were manually collected from the schools' official websites, which were also cited as sources in the corresponding Wikidata items. Training occupations belonging only to the basic stage in construction, electrical, or metalworking (*Grundstufe Bau, Elektro,* or *Metall*) were not included unless the school also offered the corresponding specialisation stage (*Fachstufe*). All training occupations regulated under the BBiG/HwO framework that schools offered but were not yet represented in Wikidata were created as new items, totalling 116. This particularly applied to *Fachpraktiker*, theory-reduced vocational training occupations for people with disabilities (§66 BBiG/§42r HwO). Each training occupation was enriched with its 5-digit or 8-digit German official occupational code (KldB) (Bundesagentur für Arbeit, 2021; Kroll, 2024). For the English labels of all German training occupations, the official translations provided by the Federal Institute for Vocational Education and Training (2024) were used.

The school websites were also checked for satellite campuses, as most federal states and *jedeschule.de* did not provide this information in their datasets. Where applicable, these satellite campuses were created as separate Wikidata items and linked to the main campus.

The dataset uploaded to Zenodo was exported via the Wikidata Query Service using the query provided in Supplementary File 2. In post-processing, a Python script was used to merge the datasets returned by the two SPARQL queries, add the federal state, and remove training occupations without KldB codes (state-regulated, non-BBiG/HwO occupations).[2] Schools that did not offer any training occupation with a KldB code were also excluded. For the code, see Supplementary File 3 on GitHub.

QUALITY CONTROL

The school websites were cited as sources in the Wikidata items, referencing the training occupation offered. Information from the official registries was manually cross-checked with data from *jedeschule.de* and the respective school websites.

(3) DATASET DESCRIPTION

REPOSITORY NAME

Zenodo

OBJECT NAME

Public Dual Vocational Schools in Germany

FORMAT NAMES AND VERSIONS

CSV file

CREATION DATES

2025-04-01 – 2025-11-27.

DATASET CREATORS

Ines Loll created 1,150 Wikidata items (see Supplementary File 4) included in this dataset under the username *Weinessig*. Some items were created by Wikidata editors, who are not involved in this project. The systematic compilation, harmonisation, and provision of the dataset were carried out exclusively by the author.

LANGUAGE

English and German

2 Vocational colleges (*Berufskollegs*) in North Rhine-Westphalia are umbrella institutions that may also include full-time vocational schools (*Berufsfachschulen*); these can only be excluded during post-processing.

LICENSE

CC-BY-4.0

PUBLICATION DATE

2025-10-20, updated on 2025-11-28

(4) REUSE POTENTIAL

The dataset on Zenodo and its underlying Wikidata items form the first FAIR-aligned resource enabling nationwide spatial analyses of public dual vocational schools based on the training occupations they offer in Germany. By harmonising registry data from all 16 federal states and linking it with training occupation information, it supports a broad range of applications in research, policy, and education.

In the VET field, it enables spatial analyses, such as calculating commuting distances by training occupation. This could, for example, provide policymakers with evidence on how training occupations can be meaningfully reorganised during school restructuring processes to improve regional coverage. By documenting the people after whom vocational schools are named, the infrastructure also supports research into commemorative naming practices. Existing research on school naming (e.g., Engelmann & Weiand, 2024; Rusu, 2019) can be extended to the perspective of German vocational schools. These analyses are straightforward in Wikidata, since biographical information about the person to whom a school is dedicated can be queried directly via the linked item. In addition, the dataset can be combined with external sources like labour market statistics, demographic data, or regional economic indicators, allowing integrated studies that relate training provision to broader socio-economic factors. Its full representation in Wikidata ensures extensibility, allowing the community to add administrative data (e.g., number of students) and historical data (e.g., mergers or founding dates). Adding this information would enable longitudinal analyses of vocational school infrastructure and provide an empirical basis for examining how demographic change affects the spatial availability of vocational training, particularly in structurally weak regions, as discussed in the existing literature (e.g., Tonhäuser & Büker, 2014). Beyond research, the dataset can inform prospective apprentices about the location of the nearest vocational school. For example, an interactive map already available (see Supplementary File 5) allows users to filter vocational schools nationwide by training occupation (see Figure 1). However, it is important to note that

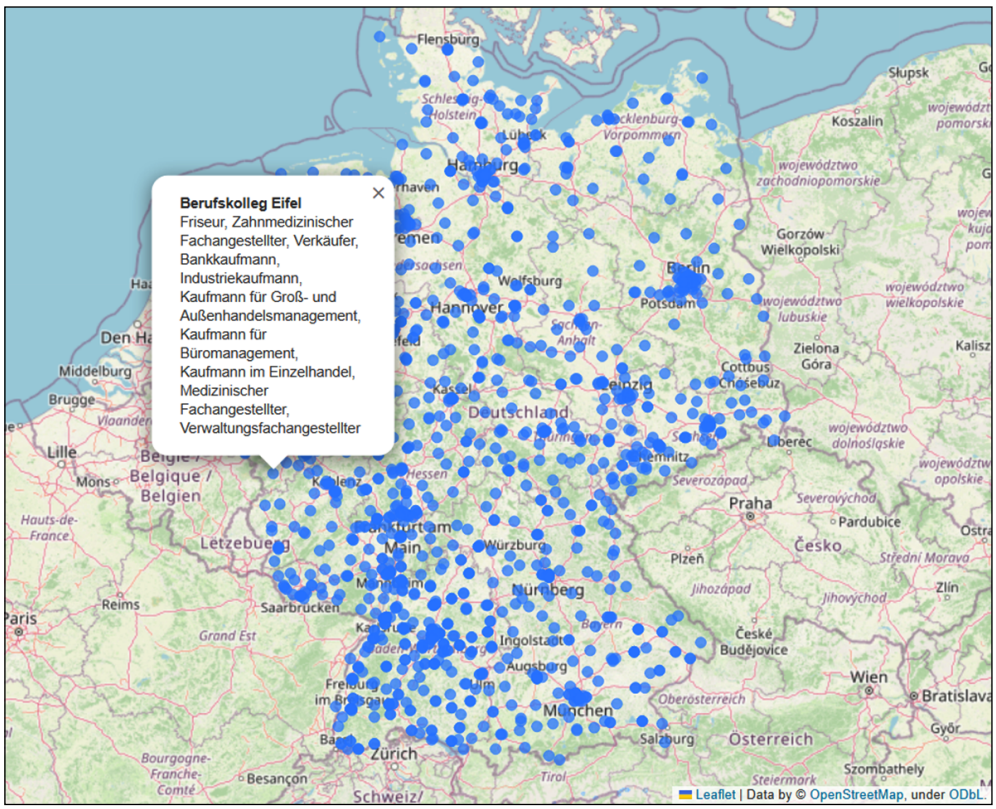

Figure 1 An Interactive Filtering Tool for all Public Vocational Schools with the Training Occupations Offered.

Note. Each dot represents one vocational school or its satellite campus. One click displays all the available training occupations.

in the German VET system, apprentices do not choose their vocational school freely; instead, training companies register them with the school that is typically closest to the workplace. Nevertheless, the information can still be valuable when individuals consider entering a rare occupation ("Splitterberufe"), in which apprentices may have to go to vocational school in another federal state.

LIMITATIONS

The dataset and the Wikidata items reflect the status as of November 30, 2025, and require ongoing maintenance to capture changes, including school closures, mergers, and updates to training occupations or their codes. In addition, the links stored as sources in Wikidata may become outdated when the school website changes. In some cases involving satellite campuses, it is not clear where the training occupations are taught. The next step is to contact schools to find out.

ADDITIONAL FILES

The additional files for this article can be found as follows:

- **Sources.** Overview of all data sources used to create the dataset. DOI: https://doi.org/10.5334/johd.422.s1
- **Queries.** SPARQL queries used to download the German and English datasets from Wikidata. DOI: https://doi.org/10.5334/johd.422.s2
- **Post-Processing.** Python script for data post-processing and generation of the interactive map. DOI: https://doi.org/10.5334/johd.422.s3
- **User_Contributions.** List of Wikidata items created by the author. DOI: https://doi.org/10.5334/johd.422.s4
- **Public_schools_map.** Interactive map to explore the training occupations offered by public dual vocational schools in Germany. DOI: https://doi.org/10.5334/johd.422.s5

Note. Zenodo refers to the GitHub folder with the supplementary files.

ACKNOWLEDGEMENTS

I want to thank the Wikidata community, my doctoral supervisors Tim Huijts and Katarina Weßling, Horizon Europe for funding, and the IAB-BIBB-ROA workshop participants for their valuable feedback.

FUNDING STATEMENT

This research project was funded under the Horizon Europe Skills2Capabilities (Grant agreement ID: 101094758).

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR CONTRIBUTIONS

Ines Loll: Conceptualisation, Data curation, Investigation, Methodology, Project administration, Validation, Visualisation, Writing – original draft; Writing – review & editing.

AUTHOR AFFILIATIONS

Ines Margarete Loll orcid.org/0009-0001-1855-7224
Federal Institute for Vocational Education and Training, Bonn, Germany; Research Centre for Education and the Labour Market, University of Maastricht, Maastricht, Netherlands

REFERENCES

Braisher, T., & **Fitchett, D.** (2025). The New Zealand thesis project: Connecting a nation's dissertations using Wikidata. *Journal of Librarianship and Scholarly Communication*, *13*(1). https://doi.org/10.31274/jlsc.18295

Bundesagentur für Arbeit. (2021). *Klassifikation der Berufe 2010 – überarbeitete Fassung 2020 Band 1: Systematischer und alphabetischer Teil mit Erläuterungen*. Bundesagentur für Arbeit. https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/KldB2010-Fassung2020-Nav.html

Code for Germany. (2025). *Jedeschule* [Dataset]. https://jedeschule.codefor.de/csv-data/jedeschule-data-2025-04-12.csv

De Wilde, M., & **Verborgh, R.** (2013). *Using OpenRefine* (1st ed). Packt Publishing.

Deissinger, T. (2018). Problems and challenges of full-time and school-based VET in Germany. In J. Gallacher & F. Reeve (Hrsg.), *New Frontiers for College Education: International perspectives* (1. Aufl.). Routledge. https://doi.org/10.4324/9781315142593

Engelmann, S., & **Weiand, K.** (2024). Waldorf, Montessori und Pestalozzi-Hype? – Schulnamen im Spiegel der Geschichte der Pädagogik. *Berichte zur Wissenschaftsgeschichte*, *47*(1–2), 27–45. https://doi.org/10.1002/bewi.202300020

Federal Institute for Vocational Education and Training. (2024). *List of training occupations within the dual system*. Federal Institute for Vocational Education and Training. https://www.govet.international/dokumente/pdf/Liste%20der%20deutschen%20Ausbildungsberufe%20DE-EN.pdf

Kroll, S. (2024). *Erhebungsberufe der Berufsbildungsstatistik sowie Berufsschlüssel der KldB 1992 und der KldB 2010*. Bundesinstitut für Berufsbildung. https://www.bibb.de/dokumente/xls/dazubi_berufsliste-berufsschluessel-kldb1992-und-kldb2010.xlsx

Mihindukulasooriya, N., Tiwari, S., Dobriy, D., Nielsen, F. Å., Chhetri, T. R., & **Polleres, A.** (2025). Scholarly Wikidata: Population and exploration of conference data in Wikidata using LLMs. In M. Alam, M. Rospocher, M. Van Erp, L. Hollink, & G. A. Gesese (Hrsg.), *Knowledge Engineering and Knowledge Management* (Bd. 15370, S. 243–259). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-77792-9_15

Rusu, M. S. (2019). Mapping the political toponymy of educational namescapes: A quantitative analysis of Romanian school names. *Political Geography*, *72*, 87–98. https://doi.org/10.1016/j.polgeo.2019.04.007

Thiery, F. (2022). SPARQLing Ogham – Irische Ogham-Steine als Linked Open Data [HTML,XML,PDF]. *Fabrikation von Erkenntnis – Experimente in den Digital Humanities. Hg. von Manuel Burghardt, Lisa Dieckmann*, 5). https://doi.org/10.17175/SB005_010

Tonhäuser, C., & **Büker, L.** (2014). Ansätze für die künftige Entwicklung berufsbildender Schulen vor dem Hintergrund der Herausforderungen der demografischen Entwicklung. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *110*(3), 374–394. https://doi.org/10.25162/zbw-2014-0023

Vrandečić, D., Pintscher, L., & **Krötzsch, M.** (2023). Wikidata: The making of. *Companion Proceedings of the ACM Web Conference 2023*, 615–624. https://doi.org/10.1145/3543873.3585579

TO CITE THIS ARTICLE:
Loll, I. M. (2026). Open Public Vocational School Data for Germany. *Journal of Open Humanities Data*, 12: 25, pp. 1–6. DOI: https://doi.org/10.5334/johd.422

Submitted: 20 October 2025
Accepted: 02 December 2025
Published: 09 February 2026

COPYRIGHT:
© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.