

# Job Opportunity: Data Engineer

**Location:** Berlin, Germany (Hybrid/On-site)

**Employment Type:** Full-time

**Reports to:** Director, Data & Engineering

## About Wikirate International e.V.

Wikirate is an open data movement powered by a global community. We are a Berlin-based non-profit (e.V.) dedicated to making data about corporate impacts transparent and accessible to all. Our platform connects data on companies' environmental, social, and governance (ESG) performance, enabling research, advocacy, and better decision-making.

We believe that accessible, structured, and trustworthy data is essential to holding companies accountable for their environmental and social impacts. By building a transparent data infrastructure, we aim to empower researchers, civil society, and decision-makers worldwide.

If you care about building technology that serves the public good and want to strengthen the data infrastructure behind corporate accountability, we would love to hear from you.

## The Role

As a **Data Engineer**, you will help strengthen and evolve the data infrastructure behind the Wikirate platform.. Your work will focus on building and maintaining reliable ETL workflows, improving data quality, and refining the systems that ensure our datasets are accurate, consistent, and performant.

You will work primarily within our Postgres environment, writing and optimizing SQL queries, developing data pipelines, and improving entity resolution and deduplication logic. This role involves working with structured and semi-structured data from diverse sources and transforming it into well-modeled, reliable datasets that power the platform.

You will collaborate closely with our Software Developers and the Director of Data & Engineering to implement practical, maintainable solutions that directly support product functionality and long-term data integrity.

This role is ideal for someone who enjoys solving complex data problems, improving system performance, and working with real-world datasets. Your work will directly support greater company accountability by making company impact data more transparent, reliable, and accessible to researchers, advocates, and decision-makers worldwide.

## Key Responsibilities

- **Pipeline Development:** Design, build, and maintain ETL/ELT pipelines to ingest and transform structured, semi-structured (e.g. JSON, XML, CSV files) and occasionally unstructured data into reliable, well-modeled datasets within our Postgres environment.
- **SQL Optimization & Performance:** Write, review and optimize complex SQL queries to ensure high performance and maintainability. Improve existing queries, indexing strategies and data structures to support efficient execution and scalability.
- **Data Quality & Deduplication:** Develop and refine data cleaning, validation and deduplication logic to ensure consistency and integrity across companies, metrics and datasets.
- **Collaboration with Engineering:** Work closely with the **Software Development team** to ensure data structures and queries integrate effectively with application logic and frontend requirements. .
- **Data Standards & Consistency:** Contribute to maintaining clear data schemas and transformation standards to ensure consistency and interoperability across datasets.
- **Data Validation & Monitoring:** Implement checks and lightweight monitoring to detect inconsistencies, ingestion failures, and performance issues.
- **Documentation:** Document data models, transformation logic, and optimization decisions to support long-term clarity and maintainability.

## What We Are Looking For

- **Experience:** 2-4++ years of experience in Data Engineering or backend development with a strong data focus.
- **Technical Stack:**
  - Strong SQL skills are essential, including writing and optimizing complex queries.
  - Experience working directly with relational databases (ideally, PostgreSQL).
  - Strong Python skills, with experience building data processing or transformation workflows.
  - Experience building and maintaining ETL/ELT workflows.
  - Experience working with structured and semi-structured data (e.g. JSON, XML, CSV)
  - Experience implementing or improving data cleaning and deduplication logic; familiarity with similarity-based matching approaches (e.g., TF-IDF, blocking, threshold tuning) is highly valued.
  - Familiarity with Git and collaborative development workflows..
- **Mindset:**
  - You are a problem solver who enjoys optimizing systems and writing clean, maintainable code.
  - You are comfortable working with messy, real-world datasets and enjoy bringing structure to complexity.
  - You think critically about data quality, performance and matching logic, not just whether code “works”.
- **Communication:** You can explain complex data concepts to non-technical stakeholders and collaborate effectively with developers.

- **Language:** Fluency in English (our working language). German is a plus but not required.

## Nice to Have

- Familiarity with Docker and containerized development workflows.
- Exposure to entity resolution approaches.
- Familiarity with data validation or lightweight monitoring practices.
- Familiarity with data representation and visualization, with the ability to produce clear, accurate visualizations to support exploratory analysis and stakeholder communication.
- Experience exploring AI-assisted workflows or agent-based approaches for data extraction, classification, or cleanup in structured and semi-structured datasets.
- Interest in open data, sustainability, supply chains, or ESG standards.

## What We Offer

- **Purpose-Driven Work:** A chance to use your skills for a positive global impact in the non-profit sector.
- **Berlin Base:** A central office in Berlin with a flexible hybrid working culture.
- **Collaborative Team:** A flat hierarchy where your ideas are heard and you work directly with leadership.
- **Professional Growth:** Opportunities to attend conferences and upskill in the open-source community.
- **Work-Life Balance:** We offer 25 days of annual leave, and additional paid time off between Christmas and New Year.
- **Compensation:** €50,000-€55,000 gross annual salary, depending on experience and qualifications.

## How to Apply

Please send your CV and a brief cover letter explaining why you want to join Wikirate to [jobs@wikirate.org](mailto:jobs@wikirate.org).

*Wikirate International e.V. is an equal opportunity employer. We value diversity and welcome all applications regardless of gender, nationality, ethnic and social origin, religion/belief, disability, age, or sexual orientation and identity.*