Gueulette Arnaud

Cloud Data Engineer | Data Platform & Governance Practitioner

Location: Brussels, Belgium | **Phone:** +32 488 69 25 31 | **Email:** arnaud_gueulette@outlook.fr

PROFESSIONAL SUMMARY

I design and scale modern data platforms by embedding governance, data quality, and operating models directly into engineering workflows, enabling organizations to treat governance as a driver of scale rather than a constraint.

PROFESSIONAL EXPERIENCE

Data Consultant, Datashift NV

Feb 2025 – Present

- **UCB** – Data Enablement Team
  Oct 2025 – Present
    - Supporting the **Data Quality team** in developing and maintaining **SQL-based data quality checks** across multiple domains. Improved the **efficiency, scalability, and maintainability** of rule development by introducing **structured templates**, **reusable query logic**, and **consistent validation standards**. Collaborated with data engineers and analysts to align business rules with technical implementations, enhancing data reliability and trust.
    - Leading an ongoing initiative to improve the **completeness**, **consistency**, and **integration** of product reference data across **Snowflake** and external systems. This includes analyzing multiple **Snowflake reference tables**, uncovering business logic behind manual data entry processes, resolving missing or inconsistent **product identifiers**, and collaborating with domain experts to validate **product catalog structures**. Designing and testing **scalable matching frameworks**—combining **exact**, **hierarchical**, and **fuzzy-matching techniques**—to maximise enrichment coverage between datasets and strengthen downstream **data quality**, **reporting accuracy**, and **governance**.
    - Currently leading a **proof-of-concept** to evaluate the **Collibra Model Context Protocol (MCP)** as a discovery and interaction layer for governance assets. The initiative focuses on improving **data discoverability**, **search relevance**, and **contextual access** to **business glossaries**, **data assets**, and **quality rules** by enabling **natural-language querying** across Collibra metadata. The POC explores **secure**, **role-aware access patterns** and integration with **engineering workflows** and tooling, with the objective of reducing friction between data consumers and governance platforms while increasing **adoption** and **trust** in governed data.

- **Scania CV** – Data Governance Initiative
  Feb 2025 – Present
    - Leading **cloud data engineering initiatives** to embed **data governance** and **data quality** practices directly into **engineering workflows** (governance by design) at Scania. Combining **functional ownership** with hands-on implementation across **Collibra** and **AWS**. Designed and deployed **end-to-end solutions** leveraging **Collibra Data Quality**, **metadata ingestion pipelines**, and **custom BPMN workflows**; built scalable AWS services (**Lambda**, **EventBridge**, **RDS**, **EKS**) to automate governance processes, enable **lineage tracking**, and improve the reliability of **data quality monitoring**.

- **Datashift NV** – Internal Initiatives
  Feb 2025 – Present
    - Developed an automated **Soda Agent** deployment framework on **AWS** using **Terraform** and **Terragrunt**, enabling **Infrastructure-as-Code** deployments in minutes through reusable, modular components.
    - Designed and implemented a **Data Quality & Governance** demo environment integrating **Soda**, **dbt**, **Airflow**, **Superset**, and **Python** to orchestrate automated validations and analytics pipelines.
    - Integrated **Collibra** as the governance backbone, linking **data quality metrics** and **lineage** to catalog assets to improve **ownership**, **traceability**, and **lifecycle visibility**.
    - Established **end-to-end observability** and **compliance-readiness** patterns, creating reusable frameworks and best practices for client engagements.

Data Consultant, Sopra Steria Benelux

Sept 2022 – Jan 2025

- **eu-LISA** – Schengen Information System (SIS)
  June 2024 – Dec 2024
    - Enhanced and maintained eu-LISA's **border-security systems** by developing, testing, and deploying improvements in **PREPROD**, ensuring **performance**, **reliability**, and **EU-standard compliance**. Collaborated with cross-functional teams to integrate changes and produced **technical documentation** to streamline future maintenance.

- **Eurocontrol** – ISOBAR
  March 2023 – Oct 2023
    - Designed and deployed a **multi-step flight-forecasting model** as an online **API endpoint** to help operators optimize **arrival/departure flows** under varying weather conditions, reducing manual workload and enabling **data-driven operational decisions**.

- **Sopra Steria Benelux** – Internal Initiatives
  Sept 2022 – Jan 2025
    - Researched and documented **AI-assisted development** tools (**GitHub Copilot**, **prompt engineering**) to support **SDLC** tasks, evaluating productivity gains and privacy implications.
    - Prototyped **Elasticsearch** integration into existing data systems to improve **search performance** and **query expressiveness**.

- Built an **end-to-end pipeline** to extract and process **PDF-based energy-consumption invoices** into the **SAS IBM Envizi ESG Suite**, improving **ESG reporting** and internal analytics.
- Delivered a **PoC data system** surfacing energy-invoice **KPIs** to support **sustainability tracking** and cost-efficiency decisions.

EDUCATION

Master's Degree in Data Science, [Université Catholique de Louvain](#)

2020 – 2023

- Master's Degree in **Data Science** with specialization in **Statistics**.
  GPA: 3.0

Bachelor's Degree in Business Engineering, [Université Catholique de Louvain](#)

2017 – 2020

- Bachelor's Degree in **Business Engineering** with specialization in **Computer Science** and **Operational Research**.
  GPA: 3.0

TECHNICAL SKILLS

Programming & Scripting
- Python
- SQL
- Linux/Bash
- Java

Data Engineering & Pipelines
- Airflow
- pandas/pySpark
- dbt
- DVC
- MLflow

Data Systems
- PostgreSQL
- Microsoft SQL Server
- MySQL
- MongoDB
- Elasticsearch
- Snowflake
- Dremio

Libraries & Frameworks
- Scikit-learn
- PyTorch
- TensorFlow
- SQLAlchemy
- Pydantic
- PyTest
- Flask
- FastAPI
- ...

Cloud Platforms
- AWS
  - S3
  - EC2
  - RDS
  - Glue
  - Athena
  - Lambda
  - Lake Formation
  - EventBridge
  - EKS
  - CloudWatch
  - ...
- Azure
  - Data Lake
  - Data Factory
  - ML Studio

Dev, Infrastructure & Ops
- Terraform
- Terragrunt
- Docker
- Rancher
- GitLab CI
- Github Actions
- Helm

Data Governance, Quality & BI
- Collibra
  - Catalog
  - Lineage
  - Workflow
  - Integration
  - Collibra DQ
- Soda
- Great Expectations
- Power BI
- Grafana
- Superset
- Kibana

Other Tools
- Kafka
- Debezium
- Power Automate
- ...

SOFT SKILLS

Critical Thinking

Problem Solving

Continuous Learning

Agile & Iterative Delivery

Teamwork & Collaboration

Business-Driven Mindset

Languages: French (C2), English (C1), Dutch (B1)

CERTIFICATIONS

Collibra

- [Collibra Solution Architect Certification](#)
  - **Collibra Solution Architect** certification demonstrating expertise in data governance platform design and implementation.

AWS

- [AWS Certified Developer – Associate (DVA-C02)](#)
  - **AWS Certified Developer – Associate** demonstrating proficiency in developing and maintaining applications on AWS.
- [AWS Certified Solutions Architect – Associate (SAA-C03)](#)
  - **AWS Certified Solutions Architect – Associate** demonstrating ability to design distributed systems on AWS.
- [AWS Certified CloudOps Engineer – Associate (SOA-C03)](#)
  - **AWS Certified CloudOps Engineer – Associate** demonstrating expertise in cloud operations and system reliability.
- [AWS Certified Data Engineer – Associate (DEA-C01)](#)
  - **AWS Certified Data Engineer – Associate** demonstrating proficiency in designing and implementing data engineering solutions on AWS.

Soda

- [Soda Certified: Cloud Fundamentals](#)
  - **Soda Certified: Cloud Fundamentals** demonstrating proficiency in cloud-based data quality fundamentals.

SAS

- [SAS Certified Specialist: Base Programming Using SAS 9.4](#)
  - **SAS Certified Specialist: Base Programming Using SAS 9.4** demonstrating proficiency in SAS base programming fundamentals.

Microsoft

- [Microsoft Certified: Azure AI Fundamentals](#)
  - **Microsoft Certified: Azure AI Fundamentals** demonstrating foundational knowledge of AI concepts and Azure AI services.
- [Microsoft Certified: Azure Data Fundamentals](#)
  - **Microsoft Certified: Azure Data Fundamentals** demonstrating foundational knowledge of core data concepts and Azure data services.
- [Microsoft Certified: Azure Fundamentals](#)
  - **Microsoft Certified: Azure Fundamentals** demonstrating foundational knowledge of cloud services and Azure core concepts.

OTHER ACHIEVEMENTS

Hackathon

- [EU Defence Innovation Scheme (EUDIS) – Hackathon 2024](#)
  - Team EUnify – 1st (Belgium) – 3rd (Europe)