

Why Linked Data?

Patterns for semantic data products

15.09.2026

Linked Data Meetup, Berner Fachhochschule

Emilie Boillat

Data Strategist, Cognizone



Data Product

An **autonomous**, **read-optimized**, **standardized** data unit containing at least one dataset, created for satisfying **user needs**

J. Majchrzak et al, *Data Mesh in Action* (2022)

Semantic technology...

...helps machines understand data.

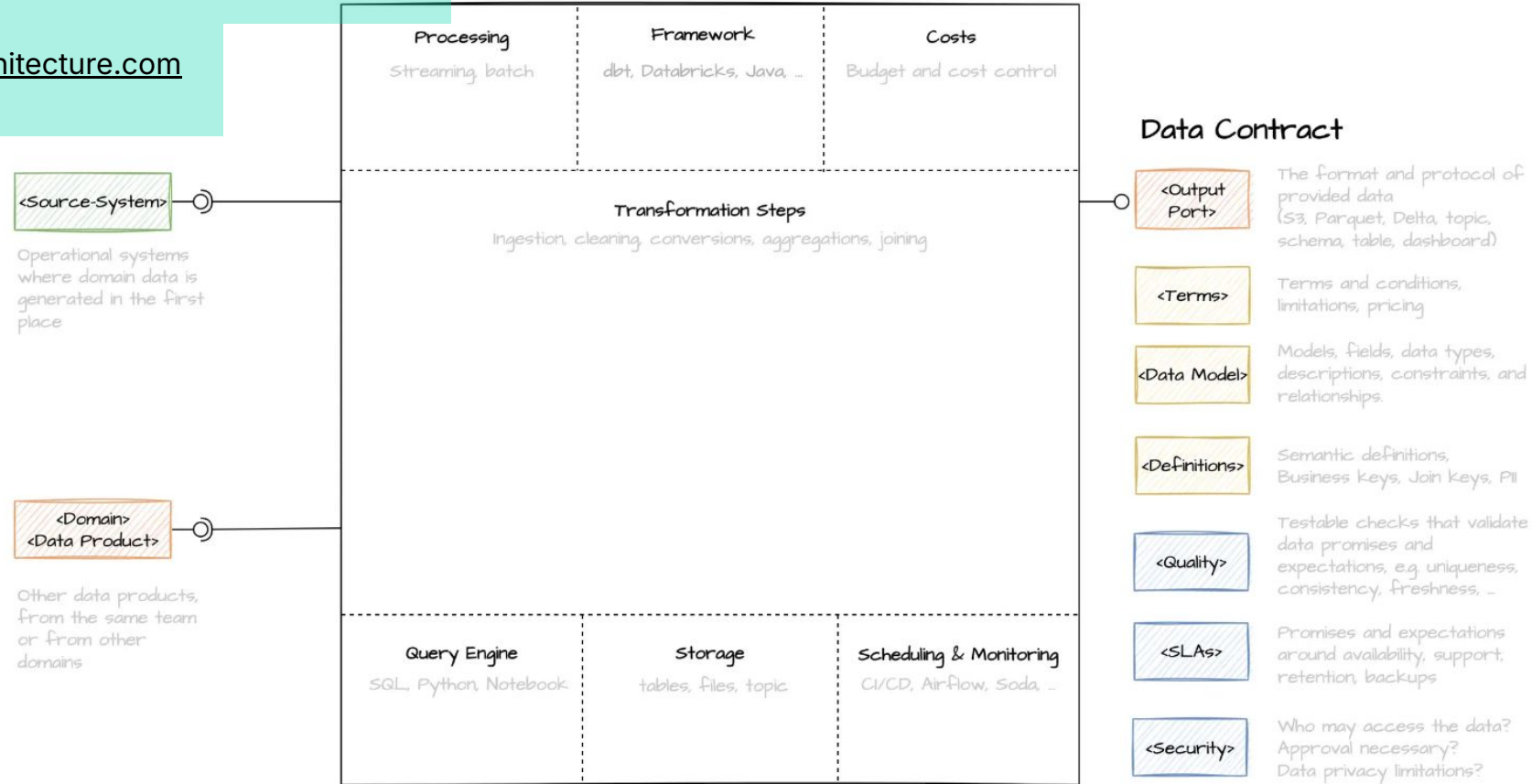
[It] encodes meanings separately from data and content files, and separately from application code.

Wikipedia, *Semantic technology* (14.09.2026)

Data Product Canvas

CC BY datamesh-architecture.com
(by INNOQ)

Data Product Architecture



Classification

The archetype of the provided data
(source-aligned, aggregate, consumer-aligned)

Ubiquitous Language

Context-specific domain terminology (relevant for Data Product)

Follow-up

e.g. create a Data Contract YAML
(datacontract.com)
Date:

Data Product Canvas

About **well-fitting** rather than individually **great pieces**



Image: Wikidata

Data Product Canvas

About **well-fitting** rather than individually **great pieces**



Image: Romain Briaux



Image: Wikidata

Data Product Canvas

About **well-fitting** rather than individually **great pieces**



Image: Romain Briaux



Image: Wikidata



Image: Wikidata

Domain

Business or policy domain

Data product name

Name · base URI · version

Owner

Accountable for requirements, questions and fixes

Classification moved

What kind of data product is this?

Date

DD-MM-YYYY

Sources

Who supplies the data, and are they obliged or motivated to? Is it sensitive?

In which form does it arrive, and how often does it change?

Which shared code lists, registers or vocabularies does it rely on, and who controls them?

Data product architecture

How does the data become the shared, connected form?

What is checked before data is allowed in?

Where does the building and operation effort sit?

Data contract

Which things are described, in business terms, with which key facts?

Through which channels: web application, download, API, query endpoint?

What can users count on: checked on entry, stable identifiers, update rhythm, licence, free or paid?

Consumer need

Who needs this data?

What “job” do they need to do, why, and how often?

What are their current alternatives or workarounds?

Ubiquitous Language

Which shared vocabularies / schemas / dimensions are key for this product?

Who governs and maintains them?

Value new

To whose business goals or public mandate does this product contribute?

How does it contribute? Lowering costs or risks, raising efficiency, stakeholder satisfaction, revenue...

Why semantic technology here? What would not work without?

Domain

Business or policy domain

Data product name

Name · base URI · version

Owner

Accountable for requirements, questions and fixes

Classification

What kind of data product is this?

Date

DD-MM-YYY

Sources

Who supplies the data, and are they obliged or motivated to? Is it sensitive?

In which form does it arrive, and how often does it change?

Which shared code lists, registers or vocabularies does it rely on, and who controls them?

Data product architecture

How does the data become the shared, connected form?

What is checked before data is allowed in?

Where does the building and operation effort sit?

Data contract

Which things are described, in business terms, with which key facts?

Through which channels: web application, download, API, query endpoint?

What can users count on: checked on entry, stable identifiers, update rhythm, licence, free or paid?

Consumer need

Who needs this data?

What “job” do they need to do, why, and how often?

What are their current alternatives or workarounds?

Feasibility

Ubiquitous Language

Which shared vocabularies / schemas / dimensions are key for this product?

Who governs and maintains them?

Value

To whose business goals or public mandate does this product contribute?

How does it contribute? Lowering costs or risks, raising efficiency, stakeholder satisfaction, revenue...

Why semantic technology here? What would not work without?

Domain

Business or policy domain

Data product name

Name · base URI · version

Owner

Accountable for requirements, questions and fixes

Classification

What kind of data product is this?

Date

DD-MM-YYYY

Sources

Who supplies the data, and are they obliged or motivated to? Is it sensitive?

In which form does it arrive, and how often does it change?

Which shared code lists, registers or vocabularies does it rely on, and who controls them?

Data product architecture

How does the data become the shared, connected form?

What is checked before data is allowed in?

Where does the building and operation effort sit?

Data contract

Which things are described, in business terms, with which key facts?

Through which channels: web application, download, API, query endpoint?

What can users count on: checked on entry, stable identifiers, update rhythm, licence, free or paid?

Consumer need

Who needs this data?

What “job” do they need to do, why, and how often?

What are their current alternatives or workarounds?

Desirability

Ubiquitous Language

Which shared vocabularies / schemas / dimensions are key for this product?

Who governs and maintains them?

Value

To whose business goals or public mandate does this product contribute?

How does it contribute? Lowering costs or risks, raising efficiency, stakeholder satisfaction, revenue...

Why semantic technology here? What would not work without?

Domain

General-purpose

Data product name

Wikidata

wikidata.org/entity

Owner

Wikidata community (volunteers + bots)
Wikimedia Foundation

Classification

Reference

Date

Sept. 2026

Sources

Volunteers, bots, GLAM institutions, tool uploads (OpenRefine, QuickStatements), sitelinks from Wikipedias:

- **Motivated**, not obliged.
- Continuous edits, 24/7; visible as a recent-changes stream.
- Open data. Some about living persons.

External identifiers: VIAF, GND, ISNI, ORCID, GeoNames, national registers (e.g. Swiss municipality numbers)...

Data product architecture

Wikibase stores statements → mapped to RDF (Wikibase ontology) → loaded into WDQS triplestore.

Enterprise: change stream (JSON) served from separate infrastructure.

Basic checks before entry. **Validation** is mostly social and **post hoc**.

Platform funded by donations; Content by volunteers; Commercial layer staffed and paid by Enterprise customers.

Integration effort → consumer (mapping Wikidata model into own schema)

Data contract

Creative Commons Zero license

Not checked on entry — constraints are reported, not enforced.

Free, Public: SPARQL / Linked Data interface Wikibase REST and Action APIs / weekly JSON/RDF dumps, recent-changes stream.

Free Enterprise account: on-demand API (50k calls/month), snapshots (30/month), Structured Contents.

Paid Enterprise: Realtime API, high-volume access, SLA, dedicated support, credibility and vandalism signals. Contract-priced.

Consumer need

Search engines & assistants: **Disambiguate entities**, in every language, realtime via API.

AI / LLM vendors: ground generations in **stable facts**; **resolve an entity** to a persistent ID. Snapshots or high-volume realtime API.

Libraries & GLAM: **Enrich catalogue records**. SPARQL / dumps, not latency-sensitive.

Ubiquitous Language

Wikidata properties and classes (community-governed). Q-IDs as lingua franca.

Wikibase RDF ontology (Wikimedia Deutschland)

schema.org for names and descriptions in RDF

skos:altLabel → **labels in 300+ languages**

thousands of **external-ID properties**

Value

One identifier hub = **entity reconciliation without bilateral agreements**

Saves you from parsing hundreds of Wikipedias

Paid consumers: **Reliable service** (fresh, support, uptime)

Donors, volunteers: Wikimedia Foundation's **free knowledge** mission

Domain

Rail transport infrastructure (EU)

Data product name

RINF+ Registers of Infrastructure

graph.data.era.europa.eu

Owner

European Union Agency for Railways

Classification

Aggregation

Date

Sept. 2026

Sources

~30 Infrastructure managers and **national registration entities (NREs)** of all EU and associated countries

- Maintain their own network data
- Via upload portal and API, with validation report
- Obligated by regulation
- Non-sensitive data
- XML or RDF uploads via portal + API
- Continuous, as networks change

Data product architecture

XML or RDF → RML mappings to RDF → values reconciled against codelists / URIs minted from national identifiers → topology and cross-border links, WKT geometry → merged into one European graph (GraphDB), indexed in Elasticsearch; named graph per submission; dumps per release.

Validation-first (SHACL, geometry, connectivity, completeness). Blocking errors reject the upload

No materialised reasoning, route traversal is computed by the application.

Integration effort → suppliers (legal obligation)

Data contract

Described: Operational points, sections of line, tracks, platforms, tunnels with their technical parameters

RINF+ web application (search, map, route compatibility check)

SPARQL endpoint / API / Zenodo (DOI) for further uses

Data quality checked and **enforced on entry**

Free public read access with fair-use limits

Consumer need

Railway undertakings:

- Can this vehicle run this route? Web application, before every new service.
- Machine access to international infrastructure information for their own applications (route book, timetabling, telematics,...)

Planners, analysts, researchers: cross-border routing, network statistics, completeness per member state. Occasional, via SPARQL endpoint and dumps.

Alternative: legacy RINF XML schema with manual checks.

Ubiquitous Language

ERA Ontology (~870 terms), **SKOS codelists** (~14,700 concepts), ERATV vehicle registers. Cross-organisation governance by ERA and national bodies. GitLab.

Partial re-use of DCAT / **RailTopoModel** / **GeoSPARQL/WKT**

National identifiers (governed by infrastructure managers and NREs)

Value

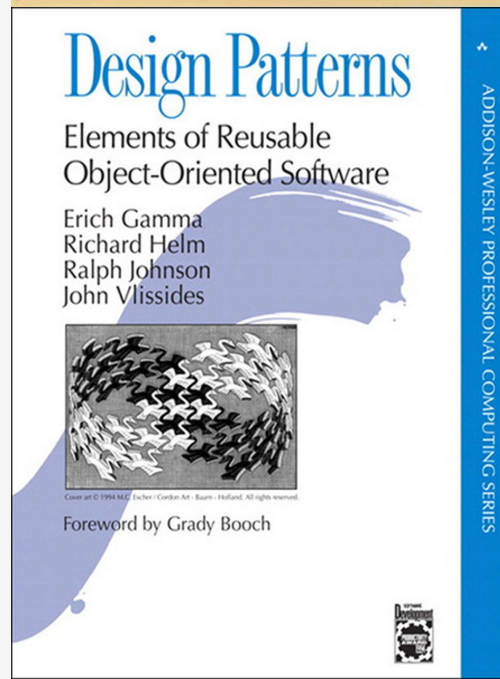
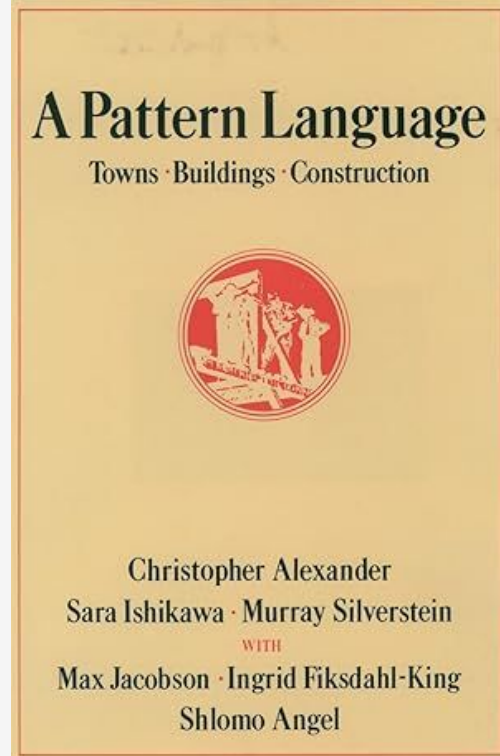
EU **interoperability mandate** (Reg. (EU) 2019/777, amended by 2023/1694)

No bilateral data exchange = **lower cost, lower legal risk**

Changes are cheap, no need for ERA re-mapping every time.

Design Pattern

A reusable **blueprint** to solve a common **problem**



Pattern 2: Aggregation

What: Collect data from many independent publishers and re-expose it, with the **mapping effort** pushed onto **each publisher** (not the consumer)

Why: **Near-zero integration cost:** one parser to build instead of one per source. Requires a **strong enough incentive** (regulation, SEO/GEO visibility, network effect) and a widely accepted schema.

[Schema.org](#)

Photos on Tripadvisor

Apple Maps

Freibad Marzili

Sports complex in Bern

76 Reviews

Directions

Call

Map

Marzilistrasse 29, Bern, CH 3005

+41 31 321 19 10

Tue 7-21h

What people say

Tripadvisor

Very large pool with clean, very cool water, perfect for going when it is hot. Very pretty view of the Federal Palace. I highly recommend it to you!!! [More on Tripadvisor](#)

Fearless61216758955

1 Sep 2026

[Deutsche Zentrale für Tourismus \(+ ODTA\)](#)

Datenbestand auf der Deutschlandkarte

Filtern Sie nach Kategorie oder suchen Sie einen Ort.

Alle

Ausflugsziele 0

Gastronomie 0

Naturattraktionen 0

Orte 0

Parks 0

Schwimmbäder 1K

Sehenswürdigkeiten 0

Tourist-Informationen 0

Unterkünfte 0

Veranstaltungsorte 0

Wellness & Beauty 0

[ERA RINF+ - Interoperable railway infrastructure register](#)

Operational Point	Name of operational point	Ty
Basel, St. Alban-Tor	Basel, St. Alban-Tor	
Zürich RB Limmattal	Zürich RB Limmattal	
Rothrist	Rothrist	
Massagno Nord	Massagno Nord	
Le Noirmont	Le Noirmont	
Dottikon Umspannanlage	Dottikon Umspannanlage	
Winterthur	Winterthur	
Altnau	Altnau	

15

Pattern 3: Open data publishing

What: Publish authoritative data — typically public-sector statistics, registries and reference lists — as standards-compliant linked open data with **persistent identifiers** and **machine-readable** catalog metadata.

Why: Third parties can **discover, understand, combine and reuse data** without bilateral agreements. The value is **reuse at scale** and **regulatory compliance** (open-data directives).

Noise pollution (BAFU) via Opendata.swiss / Visualize

Home > Organizations > Federal Office for the Environment > Traffic noise pollution

Traffic noise pollution

Last updated
August 4, 2025

Terms of use



Organization

Federal Office for the Environment
FOEN

Description

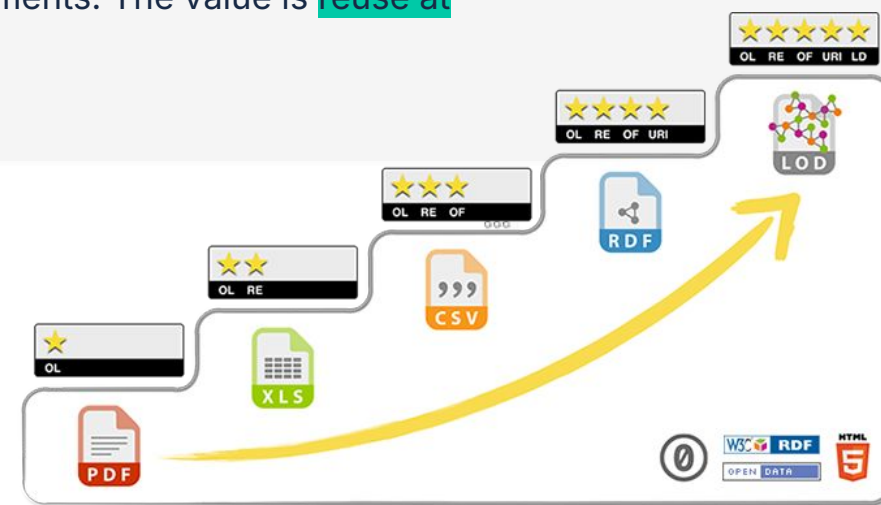
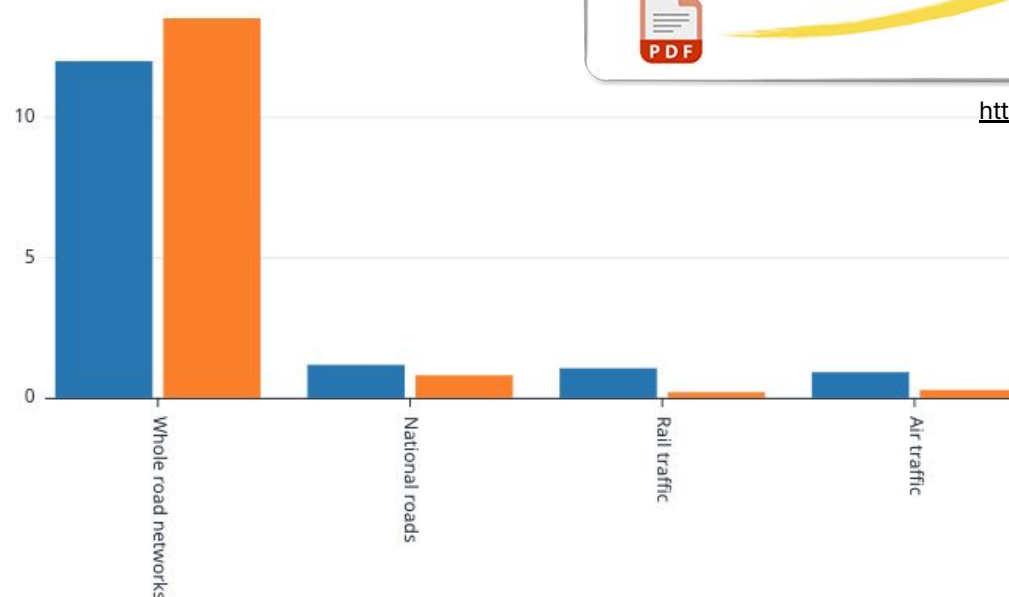
Percentage of the population, flats and buildings exposed to road, rail and air traffic noise during the day and night. The model calculations (sonBASE).

noise-pollution-noise-environmental-monitoring-sonbase-database-road-traffic-noise-rail-traffic-noise road-traffic-noise rail-traffic-noise air-traffic

Traffic noise pollution

Show Filters +

Percentage (%)



<https://5stardata.info/>

Pattern 4: Enterprise 360° View

What: Unify fragmented data describing the **same business entity** (customer, asset, product, supplier) across silos into a single, query-able 360° view — continuously refreshed and contextually rich

Why: Staff can ask **cross-domain questions** and run **consistent processes** on top of **one shared model**.

BKW Smartmeter Knowledge Graph

BKW Suon Datenhütte

Home > Meter > EL520411128

Meter

Shows details of a particular meter. Multiple records with the same manufacturer serial number are combined here.

Tenant: BKW State:	Utility Serial number: 10486793, 10486893	Manufacturer Serial number: 20411128	SAP Serial number: 301408583, 301408683	OnePortal Identifiers: Mx10486793
--------------------------	--	---	--	--------------------------------------

Metering Point:

in SAP ISU

Quick search

Group columns

45 entries

> equipmentNr	property	value
> 301408583 (23)	23 values	
> 301408683 (23)	23 values	

in MDM

Quick search

Group columns

90 entries

> mdmID	property	value
> 64850 (45)	45 values	
> 70270 (45)	45 values	

Swissgrid Grid infrastructure Knowledge Graph (in progress)

Grid 2026 with connection capacity - Load flexibility 0 %

Key:

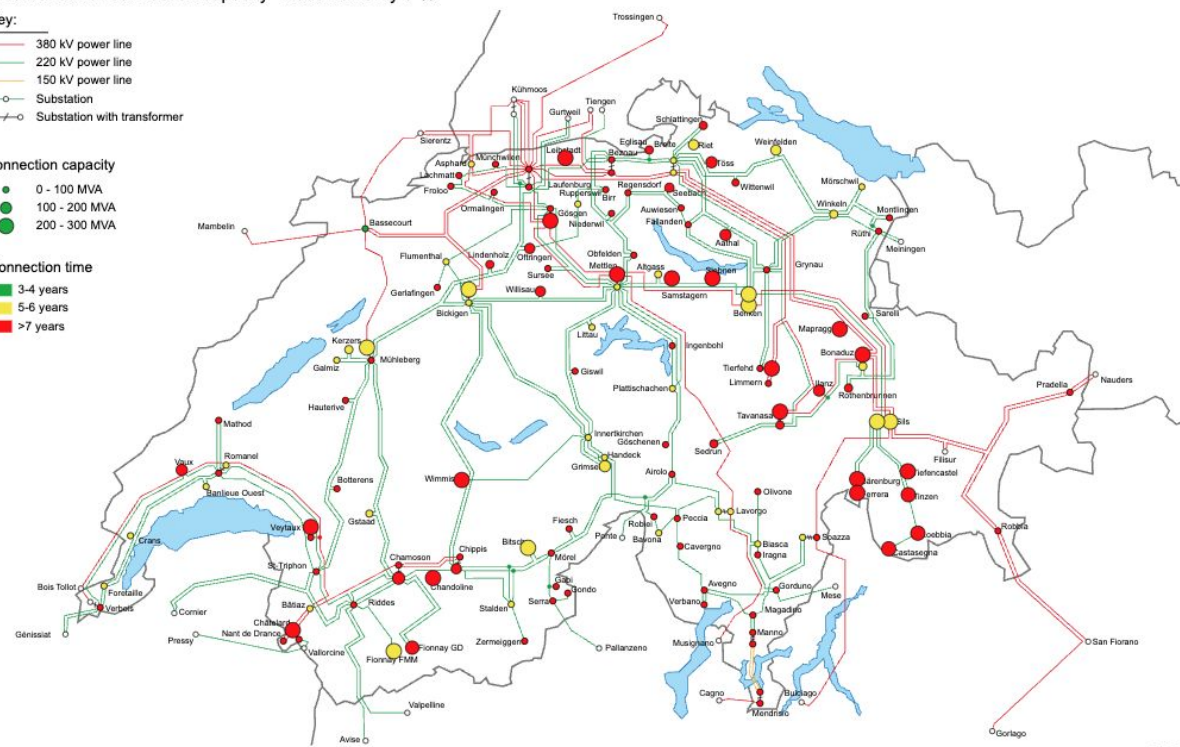
- 380 kV power line
- 220 kV power line
- 150 kV power line
- Substation
- Substation with transformer

connection capacity

- 0 - 100 MVA
- 100 - 200 MVA
- 200 - 300 MVA

Connection time

- 3-4 years
- 5-6 years
- >7 years



Pattern 5: Grounding LLMs

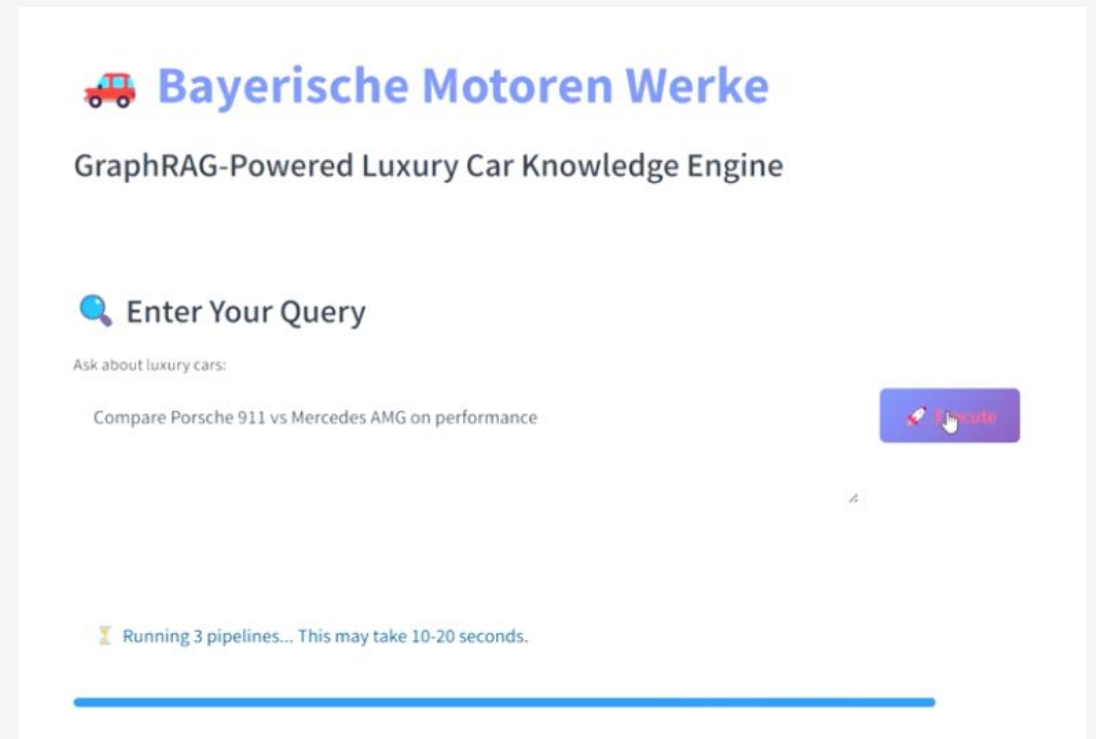
What: Ground LLM answers in a knowledge graph extracted from (or augmenting) the source corpus, instead of relying on plain vector retrieval alone

Why: Produces answers that are more accurate, traceable, and able to traverse relationships that span multiple documents.

AI4LUX



BMW PoC



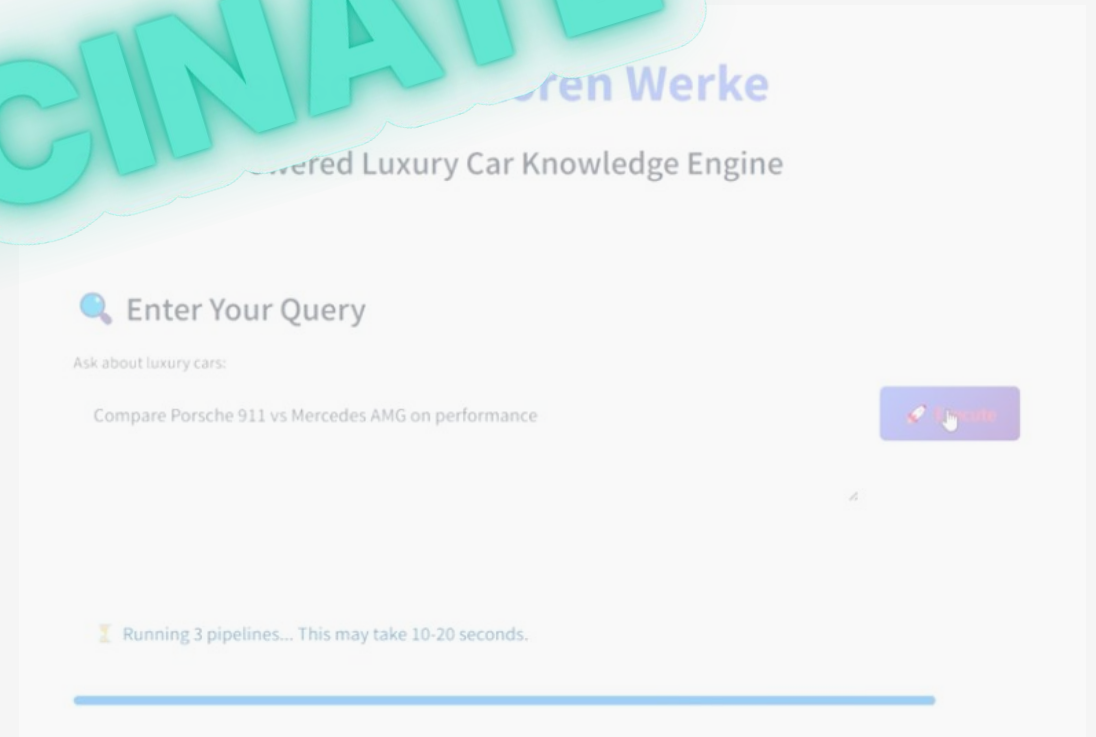
Pattern 5: Grounding LLMs

What: Ground LLM answers in a knowledge graph extracted from the source corpus, instead of relying on plain vector retrieval alone

Why: Produces answers that are more accurate and relationships that span multiple documents.

**MY KNOWLEDGE
GRAPH DOESN'T
HALLUCINATE**

cogni.zone



Towards a library of patterns and example

Why?

Drive **adoption**. Linked data gains value the more it is... well, linked.

Inspiration for your own data products

→ Can you contribute?

- Canvas structure
- Patterns
- Concrete examples of data products?



www.cogni.zone



emilie.boillat@cogni.zone



<https://www.linkedin.com/in/emilie-boillat/>

- Av. Arnaud Fraiteur 15-23
1050 Bruxelles, Belgium

- Route de la Riviala 4
1754 Avry-sur-Matran,
Switzerland

- Char. Trikoupi 18, 5th Floor
106 79 Athens, Greece

Any questions?

**MY KNOWLEDGE
GRAPH DOESN'T
HALLUCINATE**

cogni.zone