LLMs for Metadata Enrichment and Question Answering in Open Data
Some Results and Lessons Learned

Lisa Wenige with Mark Markgraf, Lena Tohoff and Doreen Straß
Hochschule Merseburg

03.06.2026, LSWT 2026

HOCHSCHULE MERSEBURG

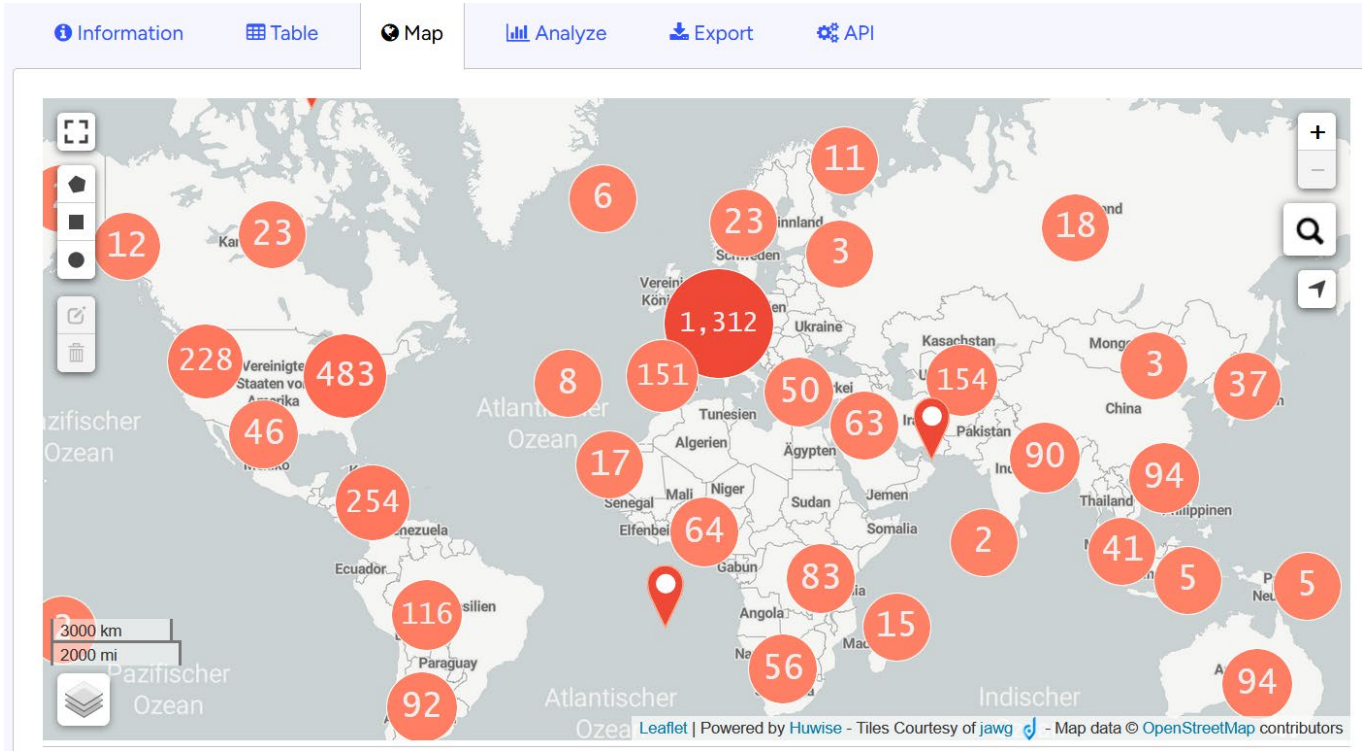
Open Data Question Answering - Motivation

Growing Significance of Open Government Data in recent years

- **Growing number of open data portals and datasets**

⇒ More than 3,700 data portals listed in Open Data Inception by Open Data Hub (compared to approx. 2,600 portals 10 years ago)

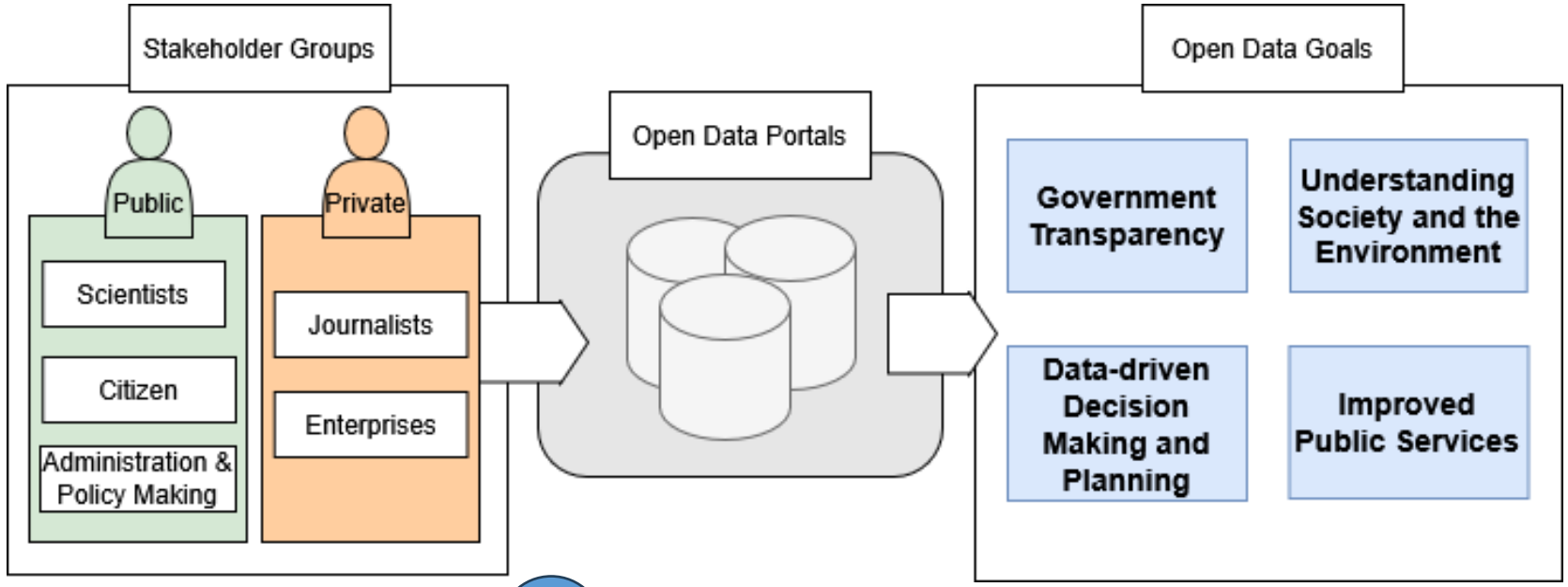
- **High Value Data (HVD) directive by the European Union** in 2023



Map of Open Data portals worldwide (https://opendatainception.io/)

Open Data Question Answering - Motivation

Current Dataset Access Limitations



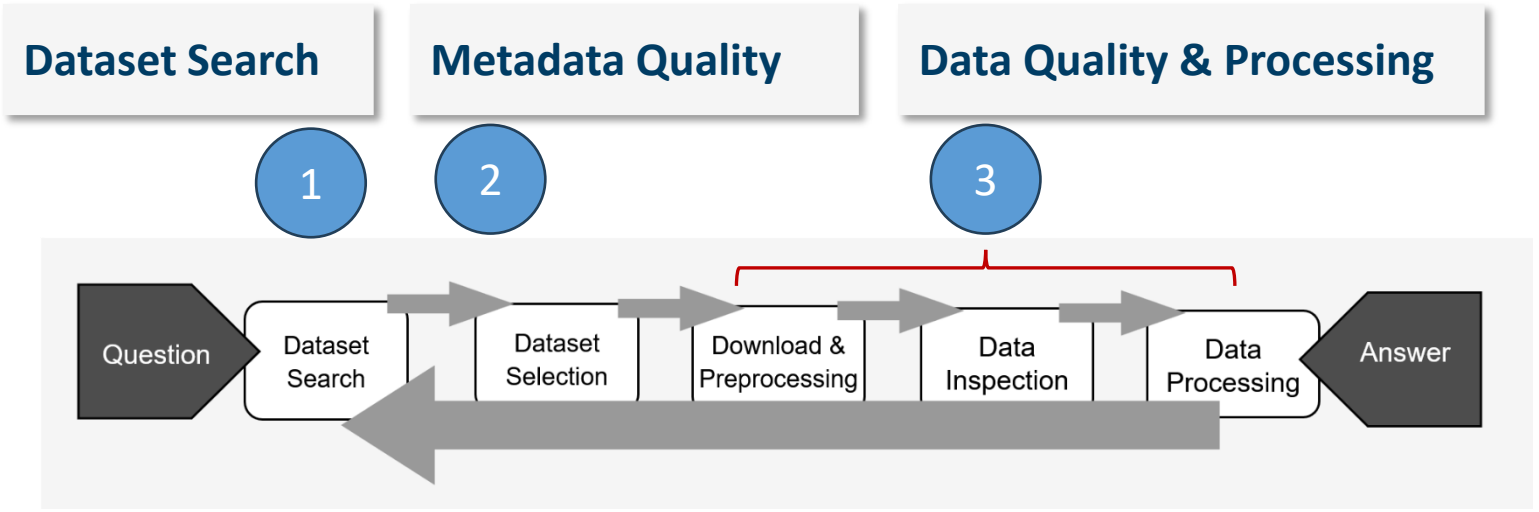
Limited FAIRness

1. **Dataset Search**
2. **Metadata Quality**
3. **Data Quality & Processing**

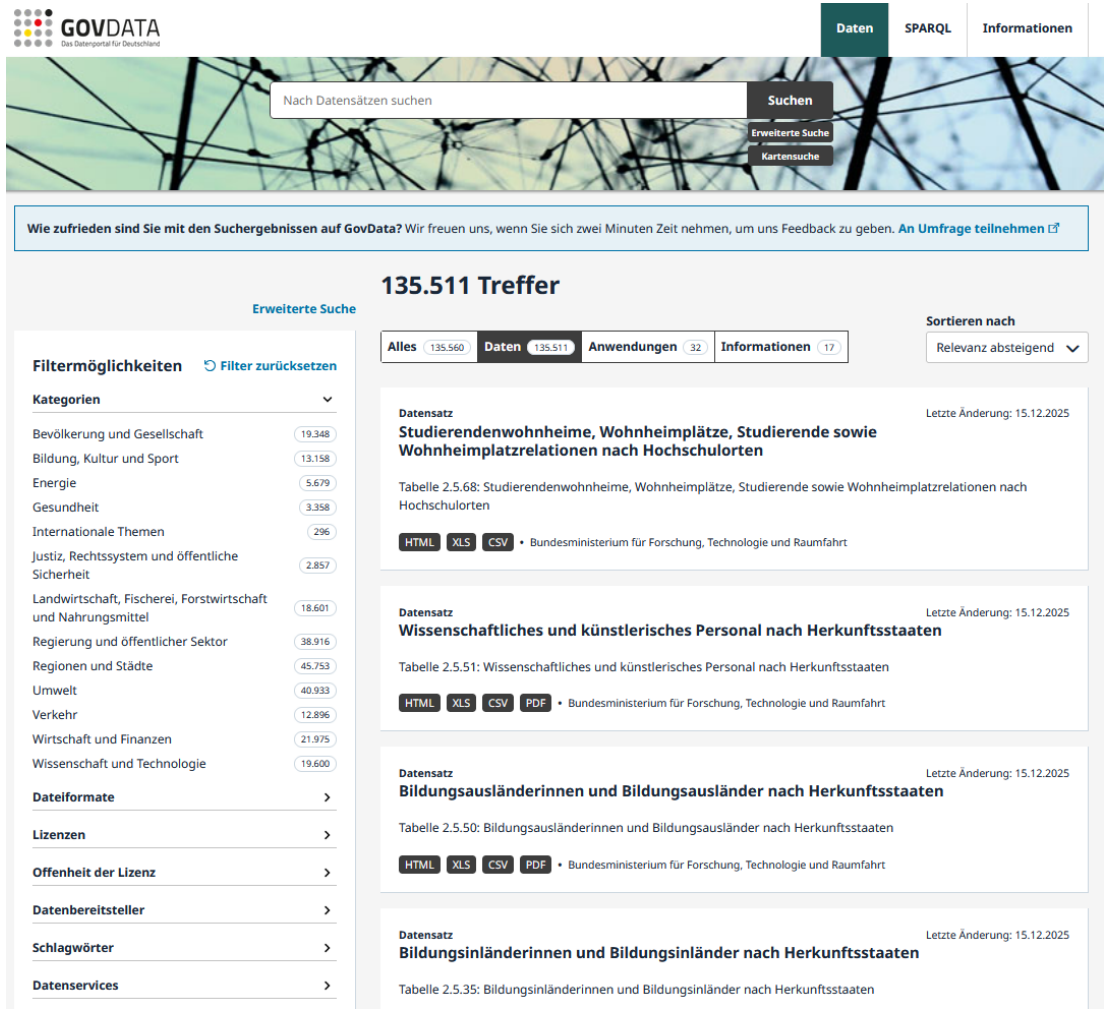
Open Data Question Answering (ODQA)

- The typical ODQA workflow is labor-intensive and mostly executed manually
- Dataset access limitations potentially hinder correct question answering at each workflow step

ODQA-Workflow



Open Data Question Answering – Challenges



https://www.govdata.de

1

Dataset Search

- Search interfaces are designed for expert users
- Sparse retrieval

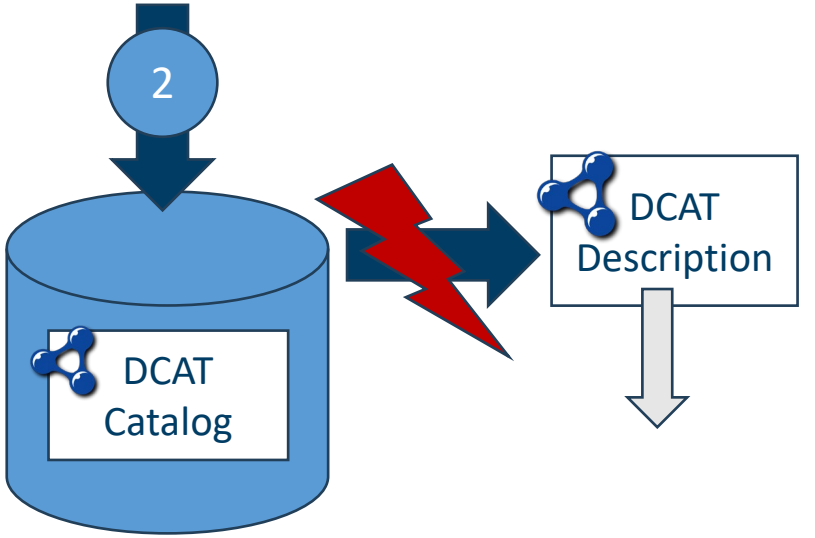
Open Data Question Answering – Challenges

How many places are there at the „Maja Bee" nursery in the city of Rostock?

Maja and Paul, Parents (Citizen)

Nursery „Maja Bee" Rostock

2

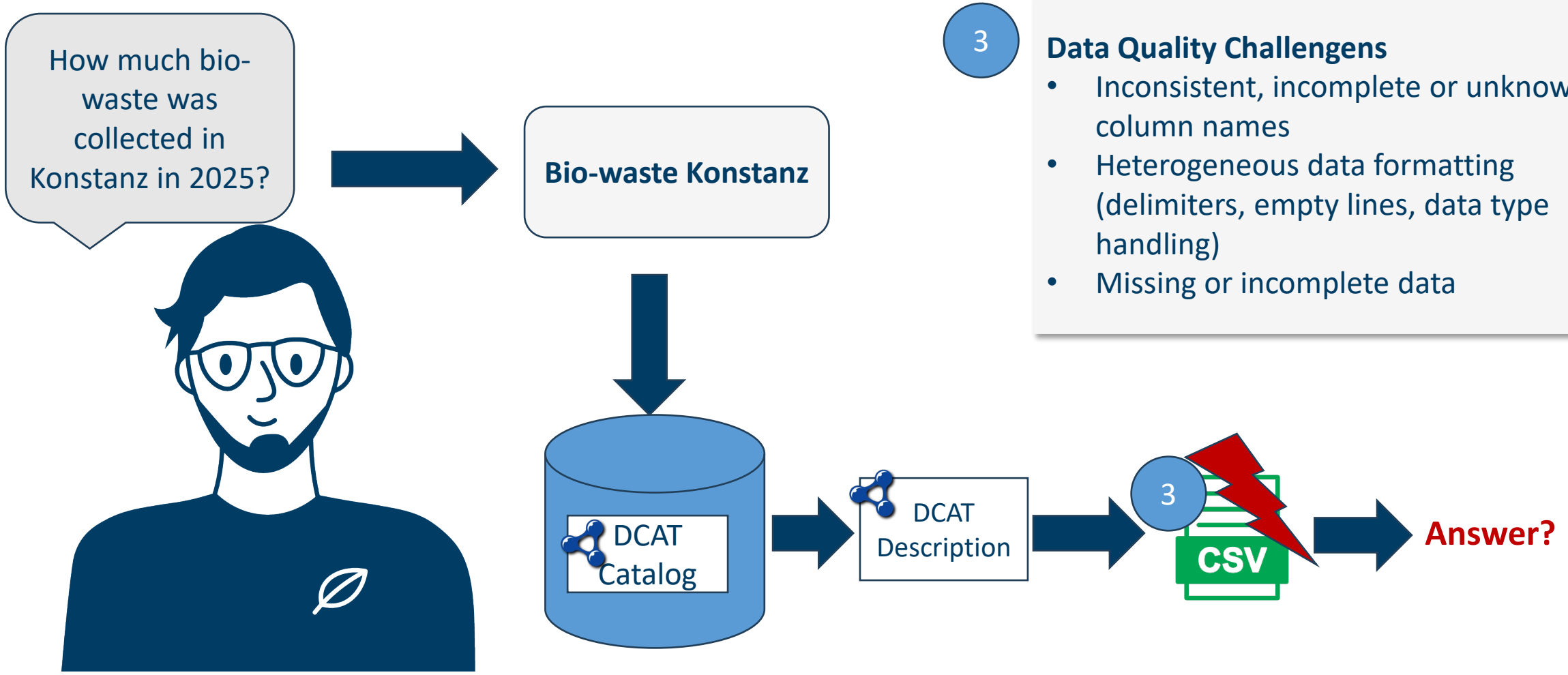


2

Metadata Challenges

- Incomplete metadata
- Limited semantic expressiveness
- Limited (semantic) web integration

Open Data Question Answering – Challenges

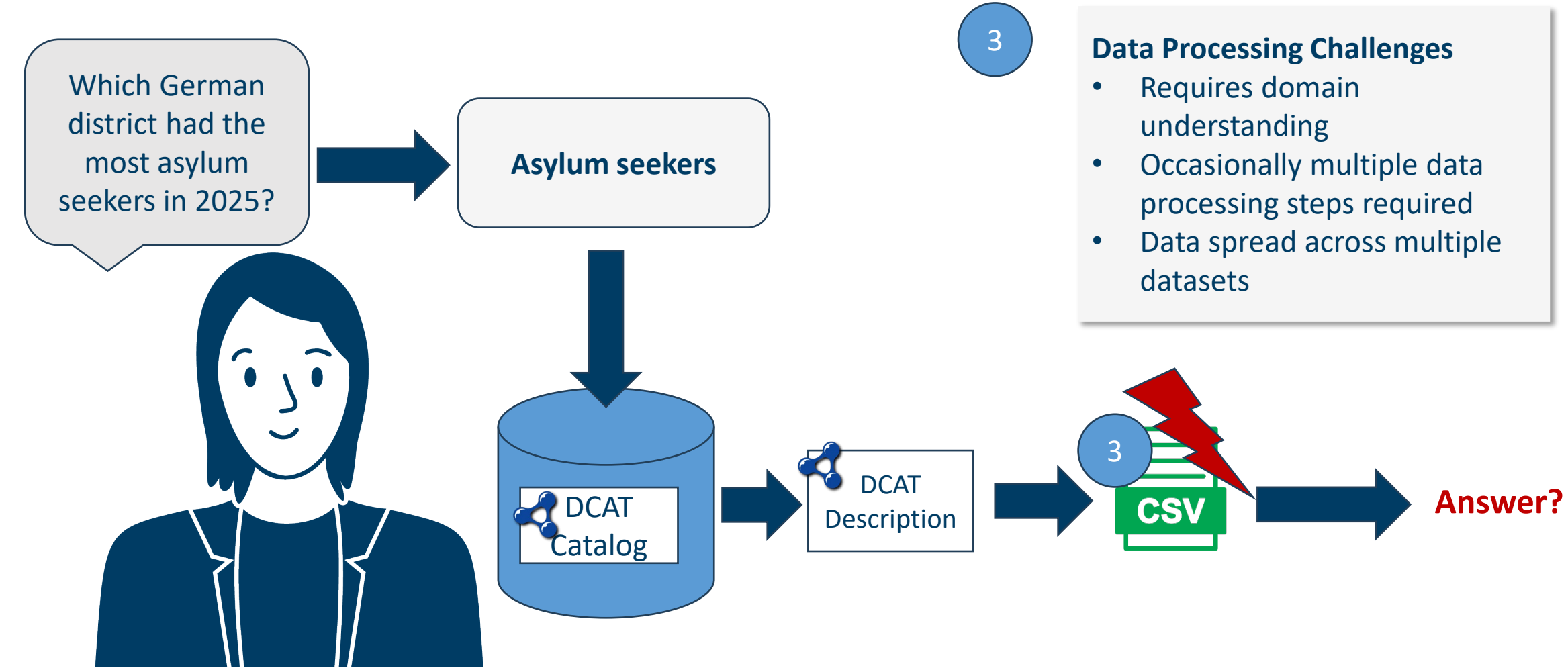


3

Data Quality Challengens

- Inconsistent, incomplete or unknown column names
- Heterogeneous data formatting (delimiters, empty lines, data type handling)
- Missing or incomplete data

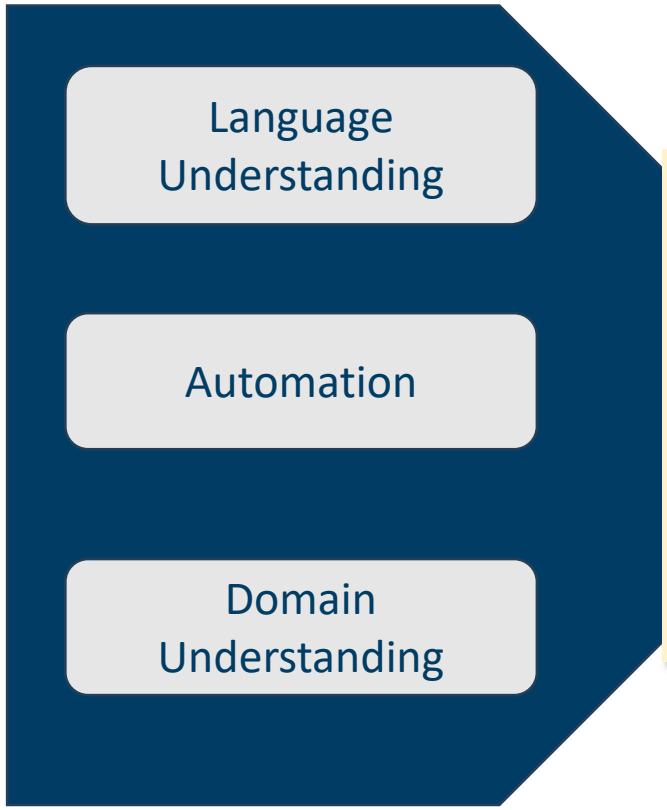
Open Data Question Answering – Challenges



Sandra, Refugee Policy and Finance Advisor, Federal Ministry of the Interior

Open Data Question Answering

Project 1:
Agentic Open Data Question Answering



Research Question
Can **LLM and agentic approaches** be feasibly **applied to open data question answering** (ODQA)?

Open Data Question Answering – Related Work

Project 1: Agentic Open Data Question Answering

Agentic Data Science

Previous studies have shown that data science tasks can be tackled with the help of agentic approaches (Sun et al.; Hong et al. 2024). **So far, agentic data science approaches have not been applied for the open data domain.**

Natural Language Interfaces for Open Data

Studies have predominantly focused on Natural Language interfaces for text data, where a RAG approach was chosen (Schelhorn, 2024; Kalyuzhnaya, 2025). **Natural language interfaces for question answering on tabular data from open data portals are missing.**

Open Data Question Answering – ODQA Benchmark

Project 1: Agentic Open Data Question Answering

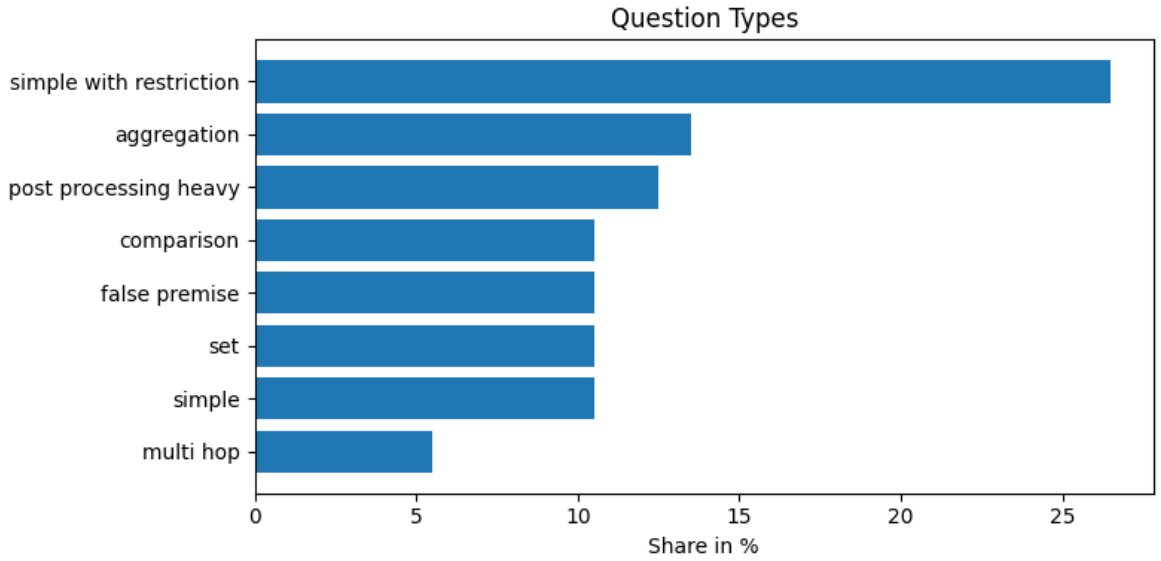
- **{en,de}-questions.csv**: Files containing the question–answer pairs in German and English.
- **sources.csv**: File including a unique mapping of each question id to corresponding source file(s)
- **data**: Directory containing the source files.
- **metadata**: Directory containing the DCAT metadata descriptions
- **govdata-catalog**: snippet of the GovData DCAT catalog (serialized in rdf/xml or turtle) as of August 2025

Statistic	Value
Total questions	200
DCAT themes	13
Task types	2
Question types	8
File types	2
Source files	220

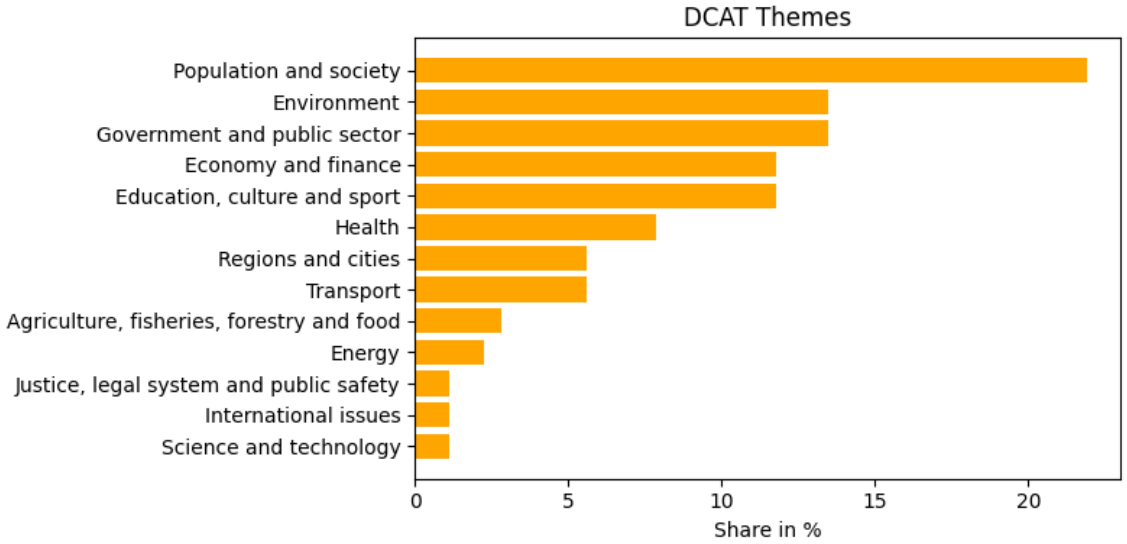
ODQA-Benchmark-Repository: https://github.com/OpenData-IQ/ODQA/tree/master/open-data-benchmark

Open Data Question Answering – ODQA Benchmark

Project 1:
Agentic Open Data Question Answering



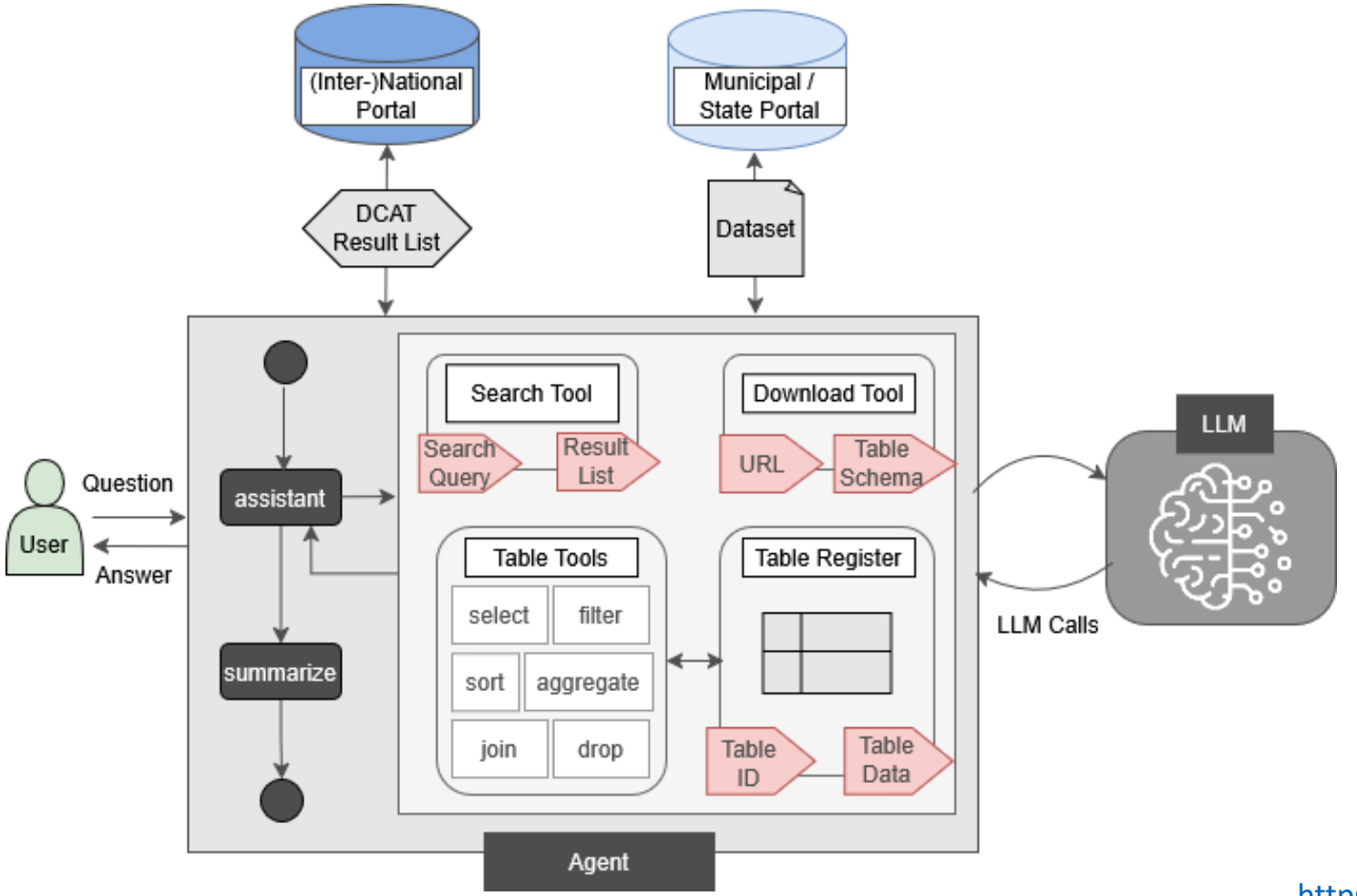
Question types come from the CRAG benchmark (Yang, 2024)



DCAT themes come from the EU vocabularies authority files

Open Data Question Answering – Agentic Setup

Project 1:
Agentic Open Data Question Answering



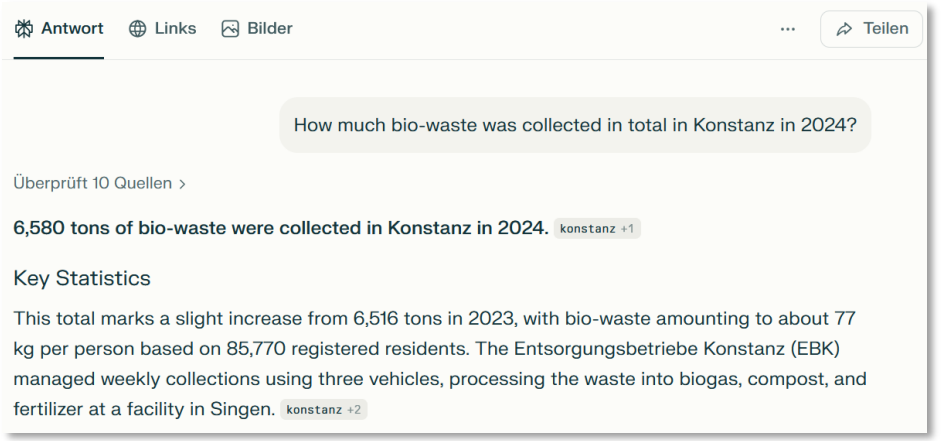
https://github.com/OpenData-IQ/ODQA

Open Data Question Answering – Experiments

Project 1: Agentic Open Data Question Answering

- **Testing the performance of the tool-calling LLM agent on the ODQA benchmark with different LLM models**
  - Higher-priced / proprietary: Claude Sonnet 3.7, GPT-5
  - Lower-priced / proprietary: Gemini 2.5 Flash, GPT-5 Mini
  - Open- Source models: Deepseek R1, Mistral Medium 3.1
- **Comparing agent performance with state-of-the-art retrieval-augmented generation systems (SOTA RAG),** such as GPT-4o-Search and Perplexity.AI (Perplexity Sonar)

SOTA-RAG-System Perplexity.AI



Open Data Question Answering – Experiments

Project 1: Agentic Open Data Question Answering

Model	Search Setup	Completion Rate	Success Rate	Conditional Accuracy	Tool Calls
Claude 3.7 Sonnet	LLM Agent	55.5	44.5	80.2	8.3 ± 3.5
Deepseek R1	LLM Agent	75.1	26.4	35.1	5.9 ± 2.9
Gemini 2.5 Flash	LLM Agent	**82.4**	40.7	49.4	3.7 ± 2.5
GPT-5	LLM Agent	58.5	57.0	**97.4**	8.4 ± 4.0
GPT-5 Mini	LLM Agent	64.0	55.0	85.9	7.5 ± 3.5
GPT-5 Mini (40)	LLM Agent	78.0	**66.5**	85.3	8.5 ± 6.0
Mistral Medium	LLM Agent	67.5	37.5	55.5	5.8 ± 3.6
GPT-4o Search	*SOTA RAG*	92.5	21.5	**23.2**	-
Perplexity Sonar	*SOTA RAG*	96.5	23.0	**23.8**	-

Open Data Question Answering – Results Summary

Project 1: Agentic Open Data Question Answering

Non-trivial accuracy results, but still not production ready

Public search / RAG systems should not be used for OQDA

Agentic approaches for ODQA need AI ready DCAT metadata

Addressing Metadata Challenges

Project 2: Automatic Metadata Enrichment with (L)LMs

Related Work

Authors	Models	Keywords `dcat:keyword`	Themes dcat:theme	Description dct:description	Additional fields	LLM Parameter	Few-Shot-Prompting	Prompting with example data	Evaluation with historic data and F1-Scoring (unless indicated otherwise)	Evaluation	User Interface
Kliimask & Nikiforova (2024) **Open Data Portal Estonia**	GPT-3.5-turbo, GPT-4	✅	❌	❌		❌	❌	✅	❌	✅ N=22 (Users) Ø 4.4 of 5	✅
Ahmed & Polini (2025) **European Open Data Portal**	(Fine-tuned) T5-small, T5-Large	✅	✅	❌	❌	❌	❌	❌	✅ Keywords: 0.45 (T5-small) 0.61 (T5-large) Themes: 0.63 (T5-small) 0.83 (T5-large)	❌	✅
Eger et al. (2025) **Open Data Portal Saxony**	Mistral-7B	✅	✅	✅	✅ Temporal and spatial data based on file names	❌	❌	✅	❌	✅ N=4 (Users), n=90 (Test runs) Approx. 87% correctly generated descriptions	✅

Addressing Metadata Challenges

Project 2: Automatic Metadata Enrichment with (L)LMs

Verwandte Arbeiten

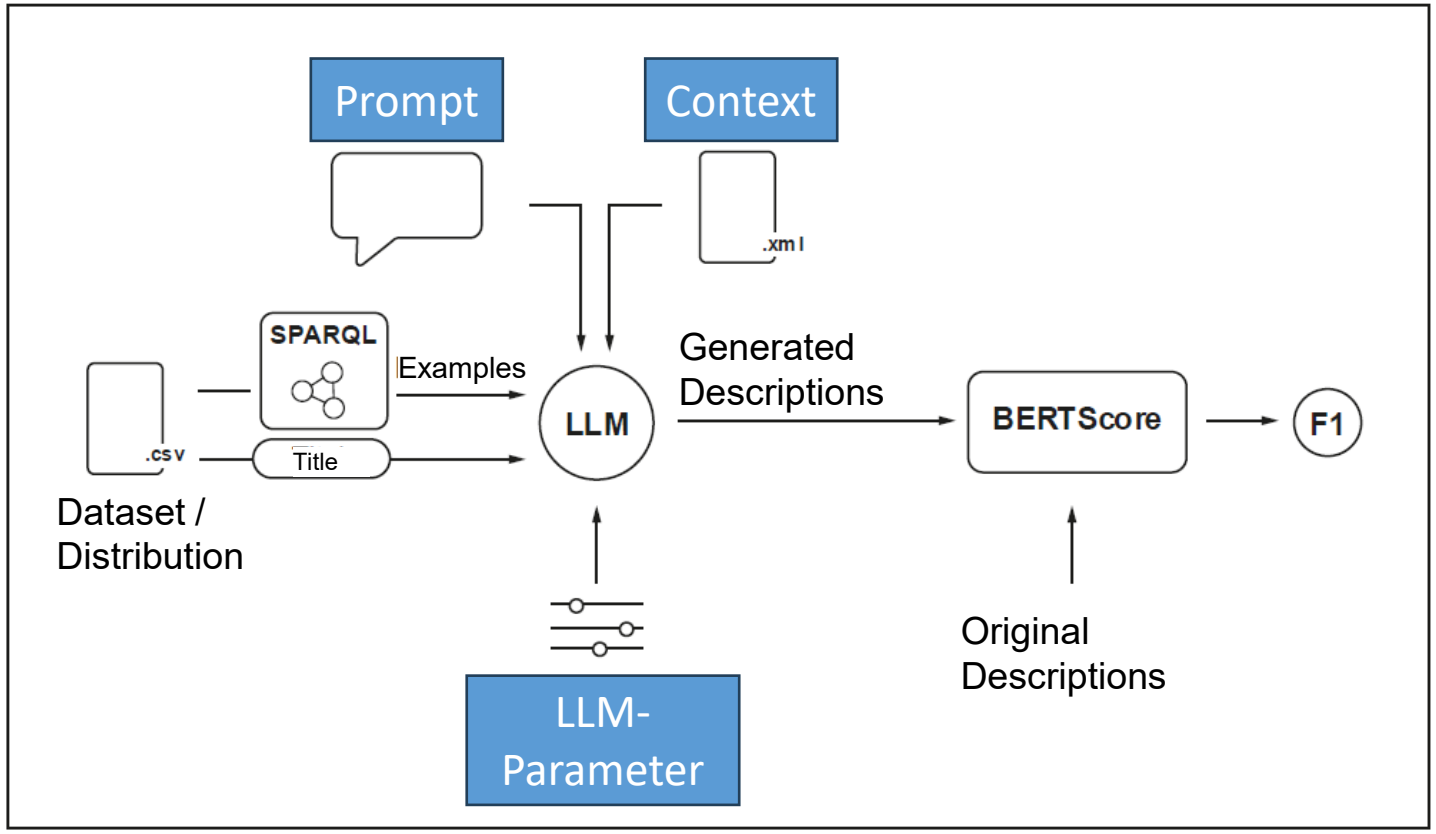
Authors	Models	Keywords `dcat:keyword`	Themes dcat:theme	Description dct:description	Additional fields	LLM Parameter	Prompting-Techniques (CoT, Few Shots)	Prompting with example data	Evaluation with historic data and F1-Scoring (unless indicated otherwise)	Evaluation	User Interface
Busch et al. 2025 **EUR-LEX, GIZ sector dataset**	T5, LLM (3B, 8B, 70B), Gemini (Flash, Pro, GPT (2, 4o, 4o-mini)	✅	✅	✅	✅ Temporal and spatial fields, creator, language	❌	✅ One-Shot vs. Few-Shot-Prompting	❌	✅ Beschreibung: 0.89 cosine similarity (Bertscore) with reference texts Themes: 0.55-0.79	❌	❌
Zhang et al., 2025, **Open Data Portal USA**	Llama 8B, GPT (4o-mini)	❌	❌	✅	❌	❌	✅	✅ Automatic generation of semantic profilea as preprocessing steps	✅ 0.71-0.89 NDCG@20 (best configuration)	✅ N=3 (Users) N= 148 (descriptions) 8.85 out of 10 for text quality (accuracy, readability)	❌
Markgraf, Mark, 2025 **Open Data Saxony-Anhalt**	GPT-4	(✅) Not quantitatively evaluated	❌	✅	❌	✅	✅	✅	✅ ca. 0.82 cosine similarity (Bertscore)	✅ N=7 (Users) SUS-Score=80.7	✅

Addressing Metadata Challenges

Project 2: Automatic Metadata Enrichment with (L)LMs

Test setup for LLM-based descriptions

`dct:description`



78 Test Cases
Open Data
Saxony Anhalt

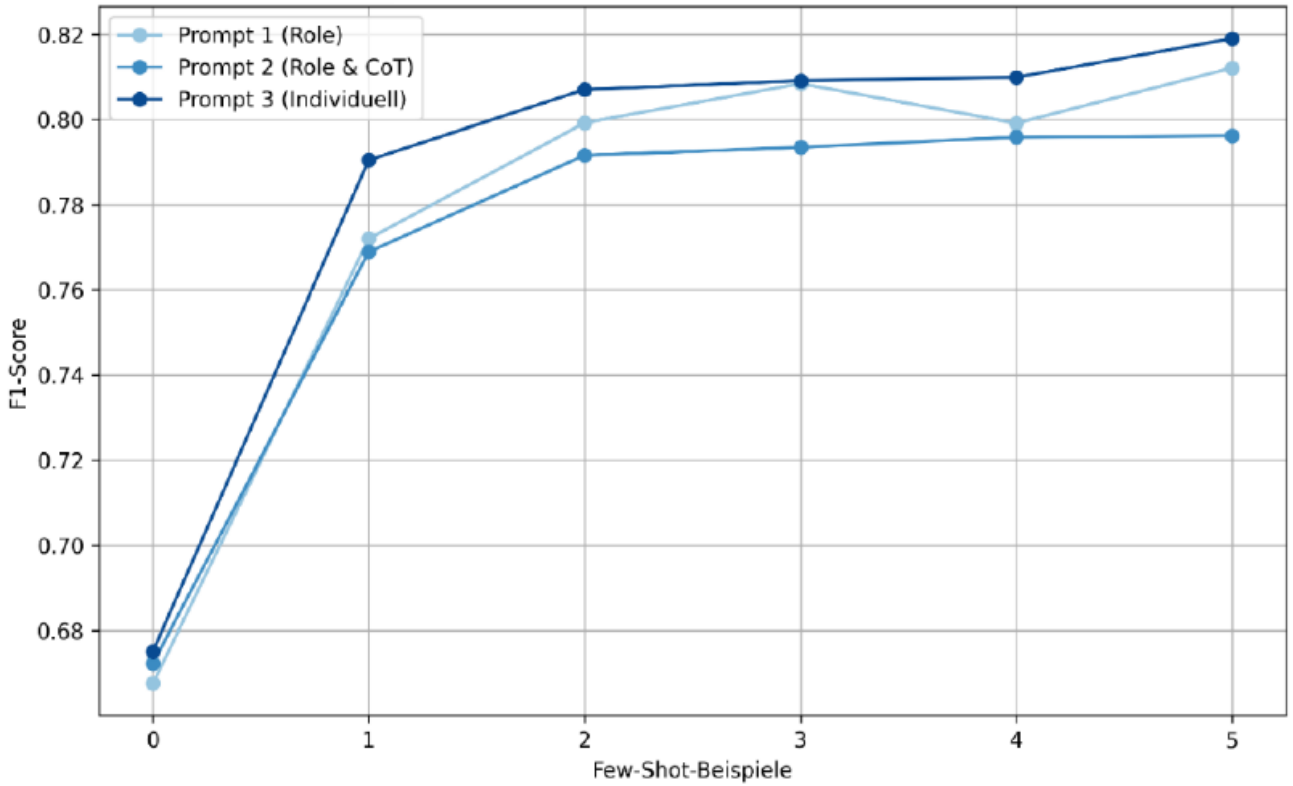
Source: Bachelor thesis, Markgraf, 2025

Addressing Metadata Challenges

Project 2: Automatic Metadata Enrichment with (L)LMs

Results of LLM-based descriptions

`dct:description`



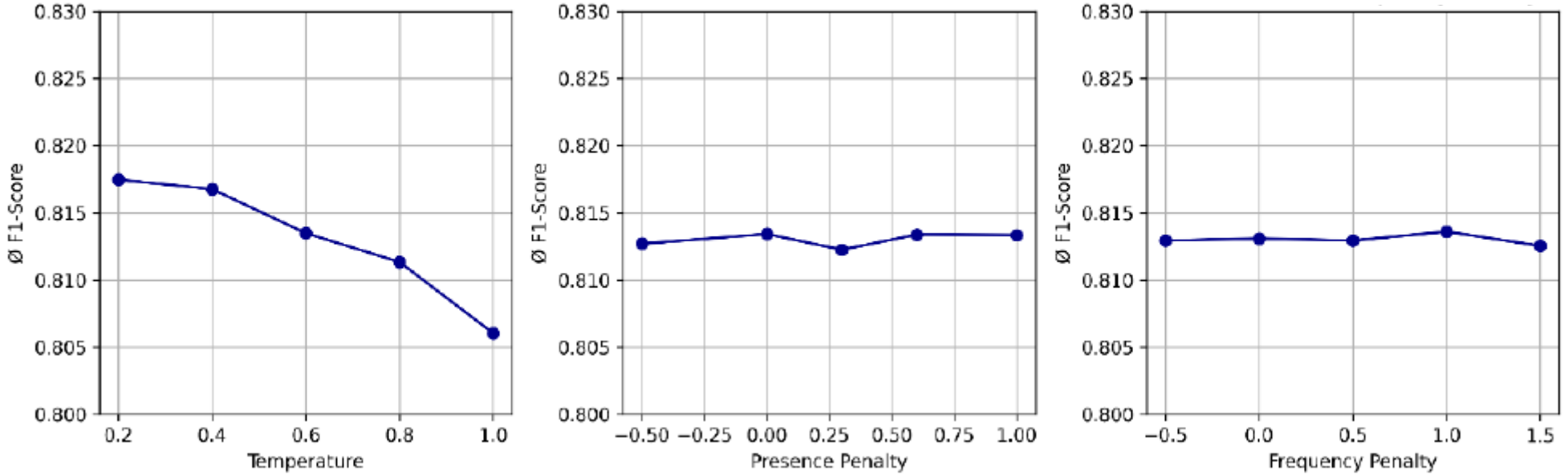
Source: Bachelor thesis, Markgraf, 2025

Addressing Metadata Challenges

Project 2: Automatic Metadata Enrichment with (L)LMs

Results of LLM-based descriptions

`dct:description`



Quelle: Markgraf, 2025

Addressing Metadata Challenges

Project 2: Automatic Metadata Enrichment with (L)LMs

Expert study with exploratory metadata editor
https://llm4opendata.streamlit.app

Research Question

What defines an effective open data metadata editor in the LLM era?

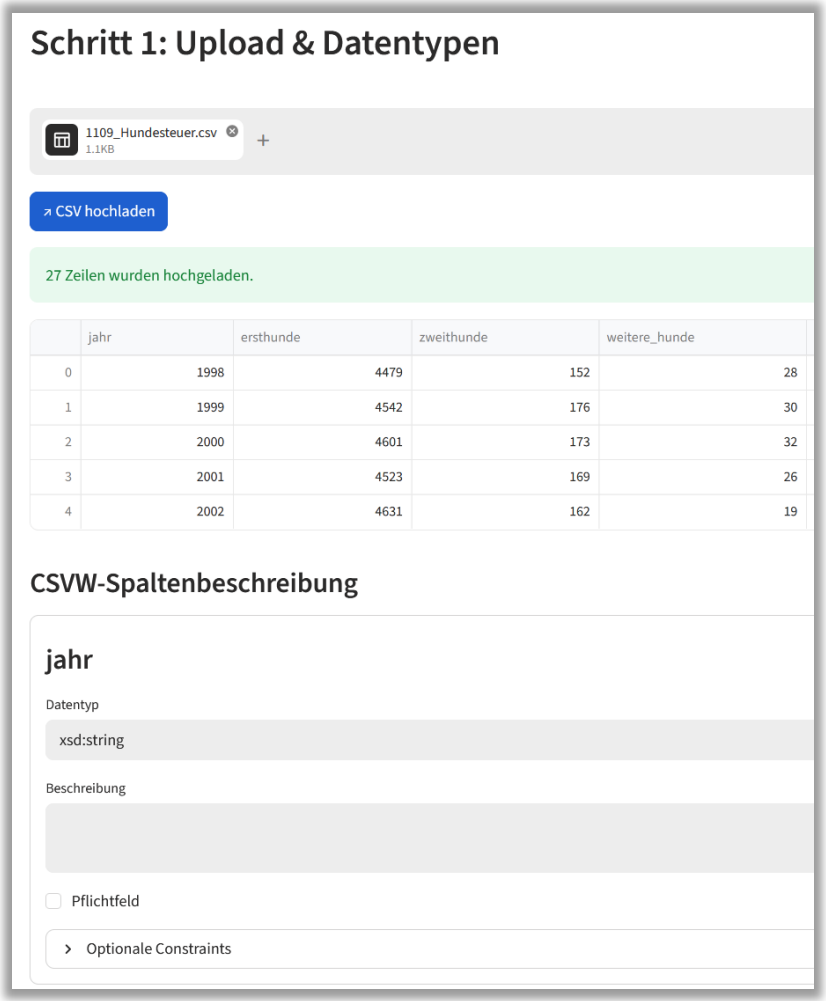
- Process
- Features
- Human vs. LLM

Addressing Metadata Challenges

Project 2: Automatic Metadata Enrichment with (L)LMs

Step 1:
Upload & Data Types

Expert study with exploratory metadata editor
https://llm4opendata.streamlit.app



Desired Features

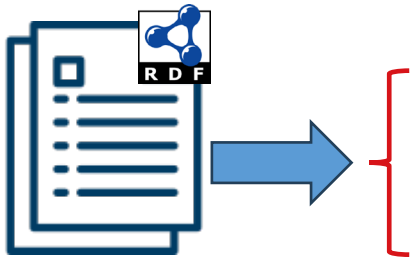
- Data Upload
- Schema Description
- Data Quality Assurance (planned)

Addressing Metadata Challenges

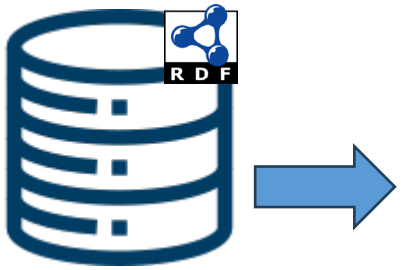
Project 2: Automatic Metadata Enrichment with (L)LMs

Step 2:
Metadata Fields &
LLM Configuration

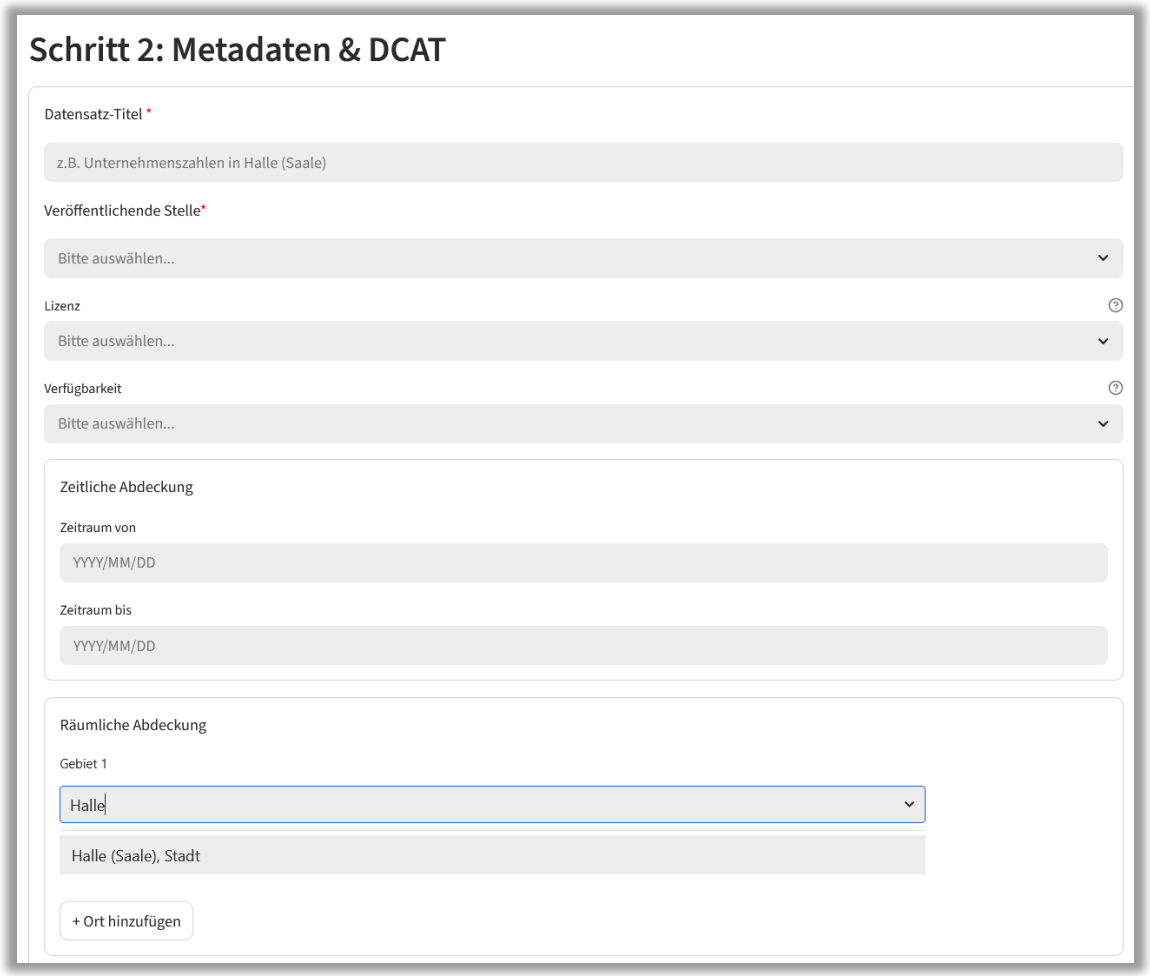
Expert study with exploratory metadata editor
https://llm4opendata.streamlit.app



URI List



Search Index



Desired Features

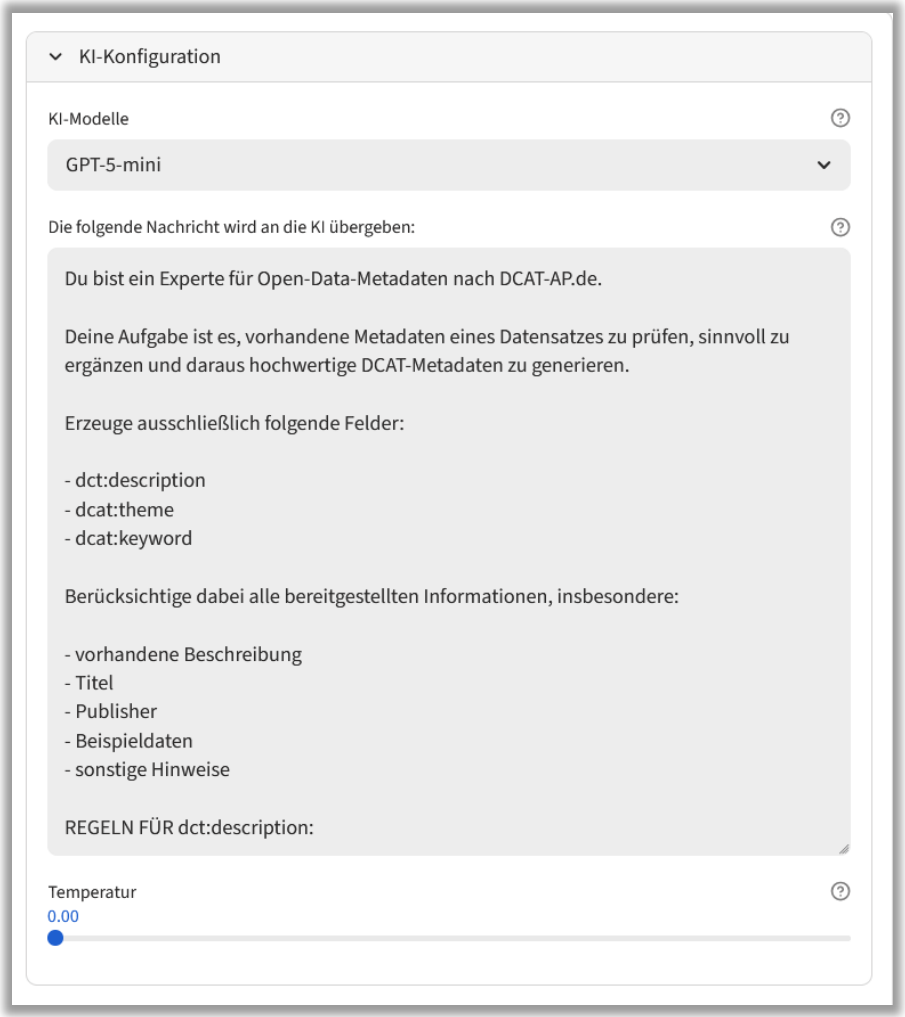
- Mandatory and recommended DCAT fields with online documentation
- URI-based term suggestions

Addressing Metadata Challenges

Project 2: Automatic Metadata Enrichment with (L)LMs

Step 2:
Metadata Fields

Expert study with exploratory metadata editor
https://llm4opendata.streamlit.app



Desired Features

LLM Configuration

- Model selection
- System prompt
- LLM parameter

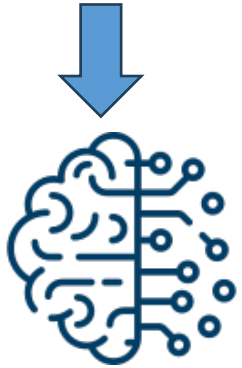
Addressing Metadata Challenges

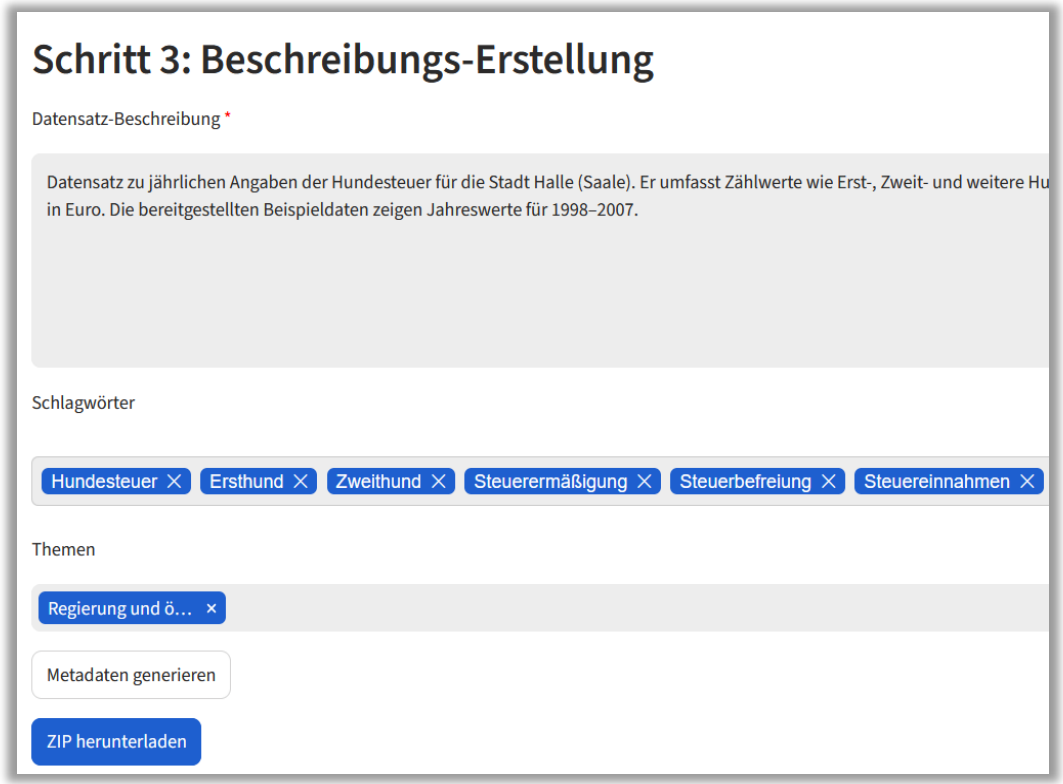
Project 2: Automatic Metadata Enrichment with (L)LMs

Step 3:
LLM-based metadata enhancement

Expert study with exploratory metadata editor
https://llm4opendata.streamlit.app

LLM Configuration
+ System Prompt
+ Example Data
+ initial metadata
+ similar datasets (planned)





Desired Features

- Metadata fields with LLM-based suggestion (editable)
- SHACL-based DCAT validation (planned)
- Download or publication of DCAT+ schema description + file(s)

Addressing Metadata Challenges

Project 2: Automatic Metadata Enrichment with (L)LMs

Summary (so far)

LLMs already produce good results when generating metadata

`dct:description` should be generated based on other metadata fields

Human-in-the loop approach needed

References

Work from HS Merseburg

- Markgraf, Mark: Automatisierte Generierung von Metadatenbeschreibungen für Open-Data-Kataloge am Beispiel des Landes Sachsen-Anhalt unter Einsatz von Large Language Models (LLM), Bachelorarbeit, Hochschule Merseburg, 2025 (Betreuerin: Lisa Wenige)
- Wenige, Lisa, Lena Tohoff, Doreen Straß. ODQA - A Starting Point for Agentic Open Data Question Answering. 19th International Conference on Metadata and Semantics Research (MTSR 2025) (in press)

Literature

- Ahmed, Umair, and Andrea Polini. "Enhancing open data findability: Fine-tuning llms (t5) for metadata generation." *Conference on Digital Government Research*. Vol. 26. 2025.
- Busch, Lennart, Daniel Tebernum, and Gissel Velarde. "Exploring llm capabilities in extracting dcat-compatible metadata for data cataloging." *arXiv preprint arXiv:2507.05282* (2025).
- Eger, Björn-Lennart, et al. "Mit Open-Source-LLMs zu aussagekräftigen Metadaten: Sprachmodelle als Schlüssel für nutzerfreundliche Open-Data-Portale." *HMD Praxis der Wirtschaftsinformatik* 62.5 (2025): 1112-1126.
- Hong, S., Lin, Y., Liu, B., Liu, B., Wu, B., Zhang, C., et al.: Data Interpreter: An LLM Agent For Data Science (Oct 2024), http://arxiv.org/abs/2402.
- Kalyuzhnaya, A., Mityagin, S., Lutsenko, E., Getmanov, A., Aksenkin, Y., Fatkhiev, K., et al.: LLM Agents for Smart City Management: Enhancing Decision Support Through Multi-Agent AI Systems. Smart Cities 8(1), 19 (Feb 2025), https://www.mdpi.com/2624-6511/8/1/19, publisher: Multidisciplinary Digital Publishing Institute
- Kliimask, Kevin, and Anastasija Nikiforova. "TAGIFY: LLM-powered tagging interface for improved data findability on OGD portals." *2024 Fifth International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*. IEEE, 2024.
- Schelhorn, T.C., Gnewuch, U., Maedche, A.: Designing a Large Language Model Based Open Data Assistant for Effective Use. In: Mandviwalla, M., Söllner, M., Tuunanen, T. (eds.) Design Science Research for a Resilient Future. pp. 398–411. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-61175-9_27
- Sun, M., Han, R., Jiang, B., Qi, H., Sun, D., Yuan, Y., et al.: A Survey on Large Language Model-based Agents for Statistics and Data Science (Dec 2024), http://arxiv.org/abs/2412.14222
- Yang, Xiao, et al. "Crag-comprehensive rag benchmark." *Advances in Neural Information Processing Systems* 37 (2024): 10470-10490.

Pictures

- Personas, Lena Tohoff
- Icons - Flaticon

Thank you for your attention!

Questions?

Kontakt

Lisa Wenige

Hochschule Merseburg
Eberhard-Leibnitz-Strasse 2
06217 Merseburg

lisa.wenige@hs-merseburg.de

www.hs-merseburg.de