



How Can Locally Deployed Large Language Models Improve Dataset Discovery in Open Government Data Portals?

Tobias Siebenlist^a and Jennifer Gnyp^a

^a *Rhine-Waal University of Applied Sciences, Kamp-Lintfort, Germany*

1. Introduction

Open government data portals constitute critical infrastructures for democratic transparency and public access to information. Their effectiveness, however, depends fundamentally on users' ability to discover relevant datasets through keyword-based search, faceted filtering, and exploratory browsing. Prior research has shown that metadata in open data portals is often incomplete, inconsistent, and lacking semantic depth, which directly limits search result quality and hampers navigation (Neumaier et al., 2016; Wentzel et al., 2023). Comparable challenges around AI-assisted metadata quality are emerging in adjacent domains such as research information systems (Azeroual et al., 2025), underscoring the broader relevance of automated approaches to metadata enrichment beyond open government data alone.

2. Approach

This contribution investigates whether locally deployed, open-weight large language models (LLMs) can address this challenge by automatically generating natural-language descriptions that enrich existing structured metadata. Building on prior conceptual work (Siebenlist & Gnyp, 2025), the approach takes as input existing metadata fields (title, existing description, DCAT-AP theme/tags), the dataset's structural schema (column names and types), and representative content samples (Eger et al., 2025; Zhang et al., 2026), and produces descriptions at three levels of linguistic complexity (standard, plain language, and easy language) using a locally hosted, open-weight model (Qwen 3.5, 9B parameters, served via Ollama). Local deployment ensures data sovereignty, institutional control, and regulatory compliance – key requirements for public sector data infrastructures – and shares its technical foundation with a companion system described in Siebenlist (2026, forthcoming).

3. Evaluation

We evaluated the approach on 50 datasets sampled from GovData.de, stratified by domain (e.g., transport, environment, education, finance), structural complexity, and existing metadata quality, yielding 150 generated descriptions (50 datasets × 3 complexity levels). Description quality was assessed along four dimensions directly relevant to search and discovery support: semantic completeness (topical, temporal, and spatial coverage), accuracy relative to the underlying data, accessibility for non-expert searchers, and practical usefulness for different search strategies (keyword search, faceted filtering, exploratory browsing). All 150 descriptions were scored by an automated LLM-judge (Prometheus 2; Kim et al., 2024); a stratified subset of 15 descriptions was additionally scored by a domain-expert evaluator as a reliability check against the automated assessment.

SEASON 2026, September 15–17, 2026, Hamburg, Germany

EMAIL: tobias.siebenlist@hochschule-rhein-waal.de (T. Siebenlist); jennifer.gnyp@hochschule-rhein-waal.de (J. Gnyp)

ORCID: 0000-0001-9435-910X (T. Siebenlist); 0009-0007-4812-9910 (J. Gnyp)

DOI: <https://doi.org/10.48441/4427.3728>



© 2026 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

4. Results and Implications

Expert ratings indicate that the generated descriptions achieve high semantic completeness and accuracy, supporting close matching between user queries and dataset content. Automated and expert scores diverge substantially on several dimensions, particularly for the plain- and easy-language variants, underscoring that automated LLM-based quality assurance alone is not yet sufficient to certify description quality for citizen-facing discovery interfaces.

Beyond the technical findings, the results carry concrete implications for the design of search and discovery systems in open government data portals:

1. Faceted search and category filters should draw on semantic categories extracted from generated descriptions rather than relying solely on raw curated metadata fields.
2. Search indices should incorporate LLM-generated descriptions as an additional, weighted searchable field that complements – rather than replaces – curated metadata.
3. Description variants of different linguistic complexity can be surfaced contextually depending on search mode and user expertise, e.g., plain-language descriptions in exploratory browsing interfaces aimed at non-expert citizens.
4. Automated LLM-based evaluation is valuable for scaling metadata enrichment across large portal collections, but should be complemented by periodic human validation before deployment, since automated judges can systematically misjudge accessible-language variants that are particularly relevant for citizen-facing discovery.

In summary, the contribution provides:

- an approach for automated, locally deployed metadata enrichment across multiple description complexity levels;
- an empirical evaluation combining automated and expert human assessment of description quality on 50 real-world datasets from a German open government data portal;
- concrete implications for the design of search and discovery interfaces in open government data portals.

5. References

Azeroual, O., Störl, U., Rothfritz, L., Schöpfel, J., Herb, U., & Borges, G. (2025). Datenmanagement 4.0: Künstliche Intelligenz als Treiber für innovative Forschungsinformationssysteme. *Information – Wissenschaft & Praxis*, 76(1), 32–42. <https://doi.org/10.1515/iwp-2024-2048>

Eger, B.-L., Ullmann, F., Dinter, B., & Gluchowski, P. (2025). Mit Open-Source-LLMs zu aussagekräftigen Metadaten: Sprachmodelle als Schlüssel für nutzerfreundliche Open-Data-Portale. *HMD Praxis der Wirtschaftsinformatik*, 62, 1112–1126. <https://doi.org/10.1365/s40702-025-01197-1>

Kim, S., Suk, J., Longpre, S., Lin, B. Y., Shin, J., Welleck, S., Neubig, G., Lee, M., Lee, K., & Seo, M. (2024). Prometheus 2: An open source language model specialized in evaluating other language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 4334–4353). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.248>

Neumaier, S., Umbrich, J., & Polleres, A. (2016). Automated quality assessment of metadata across open data portals. *Journal of Data and Information Quality*, 8(1), Article 2. <https://doi.org/10.1145/2964909>

Siebenlist, T., & Gnyp, J. (2025). Enhancing Open Data Findability through Generative AI: Conceptual approaches to improving structured data discovery. In *Search Engines and Society Annual Conference (SEASON 2025)*, Hamburg, Germany. <https://doi.org/10.5281/zenodo.17156309>



Siebenlist, T. (2026). Generating Multi-Level Accessible Dataset Descriptions for German Open Data Portals Using Local Language Models. In *Proceedings of Ongoing Research, Workshops, and Posters at the International Conference EGOV-CeDEM-ePart 2026*. TU Delft OPEN Publishing. (forthcoming)

Wentzel, B., Kirstein, F., Jastrow, T., Sturm, R., Peters, M., & Schimmler, S. (2023). An extensive methodology and framework for quality assessment of DCAT-AP datasets. In I. Lindgren et al. (Eds.), *Electronic Government: 22nd IFIP WG 8.5 International Conference, EGOV 2023* (pp. 262–278). Springer. https://doi.org/10.1007/978-3-031-41138-0_17

Zhang, H., Liu, Y., Santos, A. S. R., Hung, W.-L., & Freire, J. (2026). AutoDDG: Automated dataset description generation using large language models. *Proceedings of the ACM on Management of Data*, 4(1), Article 12, 1–27. <https://doi.org/10.1145/3786626>