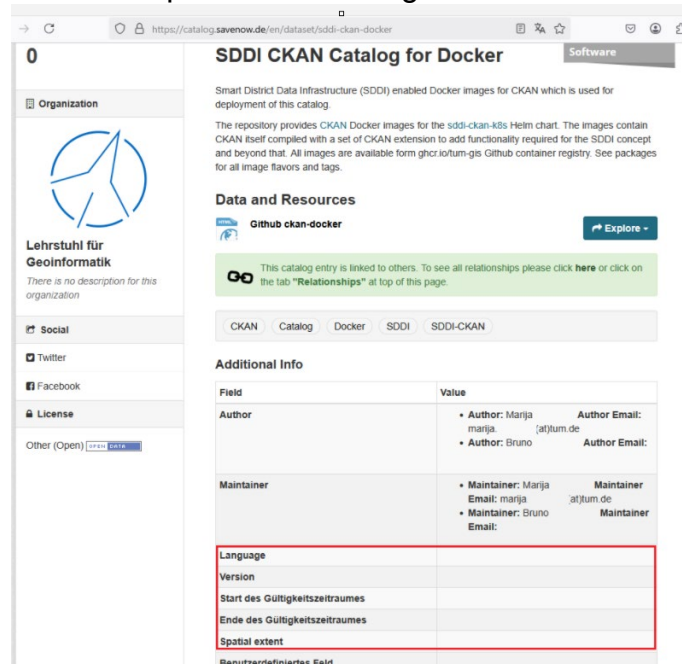


## Proposed topic for Master's thesis

# AI Methods for Evaluating and Improving Metadata Quality in Urban Data Catalogs

Metadata serve as the backbone in urban data infrastructures and are helping to make datasets discoverable, accessible, interoperable and reusable. As such, high-quality metadata is crucial in Urban Digital Twins. It provides structured descriptions that enable users and systems to search, retrieve, and integrate data efficiently across different stakeholders and systems. Metadata catalogs serve as centralized repositories that organize and standardize metadata entries, but the quality of metadata is often inconsistent across different catalog entries. Additionally, some entries are not described in a way that makes it easy to understand their contents without further investigation. This is leading to data misinterpretation, inefficiencies, and reduced usability.

This master's thesis will focus on evaluating and improving the quality of metadata based on AI tools that should be adopted and used for metadata analysis and for proposing changes to existing metadata sets. The study will be based on the Smart District Data Infrastructure (SDDI)<sup>1</sup> framework. The research will explore existing metadata quality frameworks and standards (e.g., FAIR principles, ISO 19157) and propose a structured methodology for assessing and improving metadata quality based on AI tools. The thesis will include a case study using a real-world SDDI metadata catalog and applying automated validation techniques. The evaluation should focus on key aspects of metadata quality, e.g., completeness (ensuring all required fields are fully defined), consistency (adherence to standardized formats and vocabularies), accuracy (alignment between metadata descriptions and the actual data), timeliness (regular updates and maintenance, particularly for dynamic data or highly important catalog entries), and interlinking (the extent to which metadata entries are reusable across systems). The final task is to identify common metadata quality issues, recommend improvements, and propose automated approaches for enhancing metadata management using AI tools (e.g., ChatGPT, Llama, DeepSeek).



Field	Value
Author	<ul style="list-style-type: none"> <li>Author: Marija (at)tum.de</li> <li>Author: Bruno</li> </ul>
Maintainer	<ul style="list-style-type: none"> <li>Maintainer: Marija (at)tum.de</li> <li>Maintainer: Bruno</li> </ul>
Language	
Version	
Start des Gültigkeitszeitraumes	
Ende des Gültigkeitszeitraumes	
Spatial extent	
Benutzerdefiniertes Feld	

Supervisors            Marija Knezevic, Dr. Andreas Donaubaueer,  
 Office                    0107, 0122,  
 Phone                    +49 89 289 22974, +49 89 289 22532  
 E-mail                    [marija.knezevic@tum.de](mailto:marija.knezevic@tum.de), [andreas.donaubaueer@tum.de](mailto:andreas.donaubaueer@tum.de)

<sup>1</sup> <https://www.asg.ed.tum.de/en/gis/projects/smart-district-data-infrastructure/>