

## Course Project (2025–26): Semantic Web Integration of Empirical Thesis Metadata

### Project

Undergraduate and postgraduate theses—primarily in the Social Sciences—often include **empirical studies** (quantitative or qualitative). However, institutional repositories (such as the HMU repository) typically do not store structured metadata describing these empirical components, making them difficult to search or query.

Based on the paper “[Enhancing Discoverability: A Metadata Framework for Empirical Research in Theses](#)”, a simplified XML schema was designed to capture metadata related to empirical studies. In addition, a web-based relational database application (MySQL) was developed to manage this information.

The objective of this project is to **transform the existing XML schema into a wellformed OWL ontology**, enabling the creation of a **Knowledge Graph (KG)** that represents real thesis samples containing empirical research.

Students participating in this project are expected to:

- Carefully design and implement the ontology using **Protégé**, mapping the 33 XML schema fields to appropriate OWL classes, properties, and relationships.
- Dublin Core Terms (<http://purl.org/dc/terms/>) and Schema.org (<https://schema.org/>) vocabularies should be used when is possible, instead of creating new terms
- Identify at least **30 real theses** from academic repositories that include empirical research, and generate the corresponding RDF triples (optionally using automated metadata extraction methods) to populate the Knowledge Graph.
- Deploy the Knowledge Graph in a **Virtuoso triple store**.
- Develop **at least 15 valid and meaningful SPARQL queries** that retrieve valuable insights from the Knowledge Graph.
- Design and implement a **web application** that demonstrates the system’s functionality by providing access to the predefined SPARQL queries and allowing users to construct and execute custom queries against the Knowledge Graph.
- Do literature review – prepare the paper with all the steps done during the project for a paper submission (Abstract, related work, System we created (description snapshots etc), discussion, conclusions)

This assignment will be carried out by **three postgraduate student teams**, who will collaboratively develop **three parallel implementations** of the same Semantic Web-based system for managing and querying empirical thesis metadata.

-