

## DARE25-045 - Annotating Meaning at Scale: Developing an Integrated Research Environment for Textual Data

### Abstract

TEKER – Annotating Meaning at Scale: Developing an Integrated Research Environment for Textual Data

The project TEKER addresses a central methodological gap in the humanities and social sciences: the lack of integrated, user-friendly tools that enable large-scale, semantically informed text annotation while retaining interpretive depth. Existing platforms either provide fine-grained qualitative annotation or scalable machine-learning functionality, but not both. TEKER will bridge this divide by creating an open-source, fully dockerized Python application that integrates transformer and Retrieval-Augmented Generation (RAG) pipelines with human-in-the-loop workflows. Researchers will be able to upload corpora, define custom tags, perform batch annotation with automatic Wikidata Q-ID suggestions, and refine results through an intuitive interface.

The project combines scalability with contextual awareness, aligning with FAIR data principles and supporting multi-layered analysis across temporal, spatial, and textual dimensions. Its direct users—historians, philologists, sociologists, anthropologists, and related scholars—will gain a powerful environment for transforming textual sources into structured, interoperable data.

TEKER will be developed over twelve months by a specialized developer and a student assistant under the PI's supervision. After public release via GitHub, the tool will be hosted within the University of Vienna's digital research infrastructure, contributing to its emerging core science facilities and national networks such as CLARIAH-AT and DHInfra.at. Dissemination will leverage the PI's international collaborations in digital humanities and interpretive social science, ensuring rapid uptake and continued co-development. Released under an open-source license and adhering to FAIR and Open Science standards, TEKER will offer a sustainable, interdisciplinary platform for annotating meaning at scale.

Scientific disciplines:

Classical studies (34%) | Digital humanities (33%) | Sociology of culture (33%)

Keywords:

Qualitative analysis Text-as-Data Batch-annotation

---

Principal Investigator: Till Hilmar

Institution: University of Vienna

---

Status: Contract in preparation

GrantID: 10.47379/DARE25045

---

Further links to the persons involved and to the project can be found under

<https://www.gmbh.wwtf.at/funding/programmes/ei/DARE25-045/>