AIMWORKS: Template-Driven, Agentic Framework for FAIR Knowledge Graph Construction in Hydrogen Technologies

Mohammad J. Eslamibidgoli[1,2], Marjan Kohandani[1,2], Michael H. Eikerling[1,2], Kourosh Malek[1,2]

[1] Theory and Computation of Energy Materials (IET-3), Institute of Energy Technologies, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

[2] Centre for Advanced Simulation and Analytics (CASA), Simulation and Data Science Lab for Energy Materials (SDL-EM), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

Abstract

Hydrogen and electrochemical energy research produces rapidly evolving, heterogeneous outputs - protocols, instrument settings, conditioned performance metrics, and multi-scale materials descriptors - that are difficult to curate into FAIR, machine-actionable metadata [1]. Many “LLM-first” knowledge-graph pipelines rely on monolithic prompts and ad-hoc post-processing, which can lead to inconsistent terminology, unreliable unit handling, and missing provenance and dataset details [2]. We present AIMWORKS [3], a template-driven, agentic framework that improves FAIR metadata by design. AIMWORKS uses a stable core vocabulary and a curated library of reusable templates for common experimental patterns (measurements, processes, experimental context, instruments, metrics, and dataset/provenance blocks) [4]. Given a user’s natural-language research question, the system selects the most relevant templates, assembles them into a structured knowledge graph, and exports it in standard formats (RDF/JSON-LD). To ensure reliability, each template includes micro-level validation rules (SHACL) and the system applies deterministic checks and repairs to enforce consistent typing, represent conditioned metrics via a DataPoint pattern, normalise quantities and units using QUDT, and generate a complete dataset description (including title, license, and access information). Outputs integrate cleanly into downstream platforms such as Neo4j and institutional knowledge-graph infrastructures. Case studies from hydrogen technologies (polarization curves, impedance spectroscopy, durability protocols, and ionomer–catalyst-layer questions) show that the template-first approach improves metadata completeness and interoperability while reducing manual curation and providing a transparent trace from query to graph.

References

[1] Dreger, M., Eslamibidgoli, M. J., Eikerling, M. H., & Malek, K. (2023). Synergizing ontologies and graph databases for highly flexible materials-to-device workflow representations. Journal of Materials Informatics, 3(1), N-A

[2] Dreger, M., Malek, K., & Eikerling, M. (2025). Large language models for knowledge graph extraction from tables in materials science. Digital Discovery.

[3] https://meslamib3-aimworks4-streamlit-app7-2kyctx.streamlit.app/

[4] https://github.com/meslamib3/aimworks4/tree/main/templates