

AnNER and OnNER: Traceable Construction of Knowledge Graphs of Domain Entities from Text Sources

Umayer Reza, Torsten Hahmann

School of Computing and Information Science, University of Maine

{a.reza, torsten.hahmann}@maine.edu

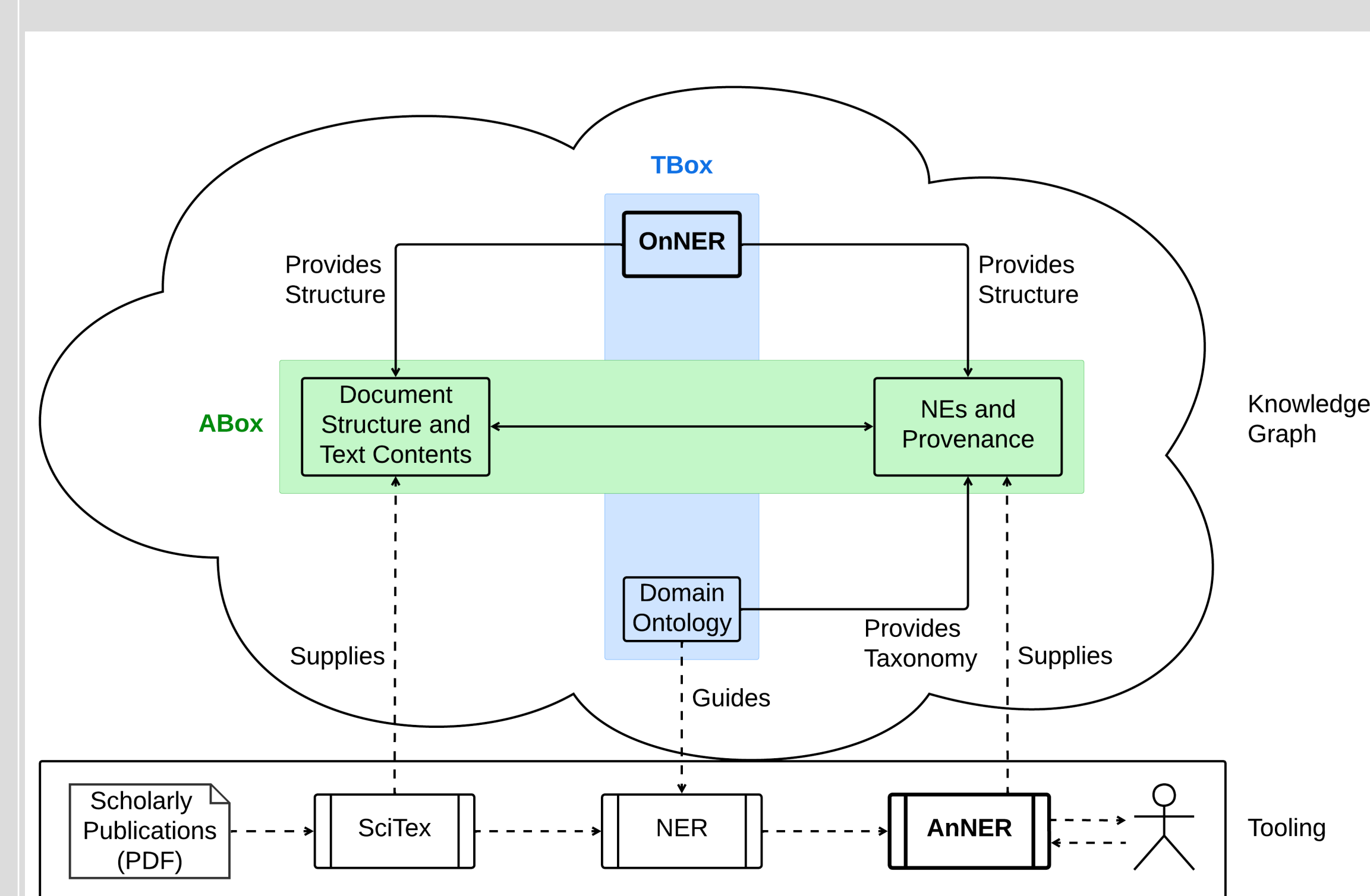
Background and Motivation

- Domain-specific terms are often buried in unstructured text, making it difficult to search, analyze, or reuse.
- Knowledge graphs help make this information accessible—but their trustworthiness depends on transparency about the included knowledge.
 - To trust a knowledge graph, users must know where each entity came from, how it was extracted, and who validated it.
 - However, NER alone—whether rule-based or LLM-based—still makes errors, misses domain nuances, and cannot ensure complete accuracy.
- A human-in-the-loop approach to the representation and linking of domain entities is essential, but existing tools lack:**
 - structured expert review capabilities,
 - complete provenance tracking of entities and the review process, and
 - clear links between recognized entities and their source text.

Contributions

- OnNER**, as an ontology, semantically represents domain entities extracted from text, the source and textual context of the entities, and their review/approval process.
- AnNER**, as an annotation review tool, includes full provenance tracking: annotator identity, review history, timestamps, and links to the exact textual context.
- CelloGraph**, as a queryable knowledge graph, is built on OnNER.

Knowledge Graph System: High-Level Approach

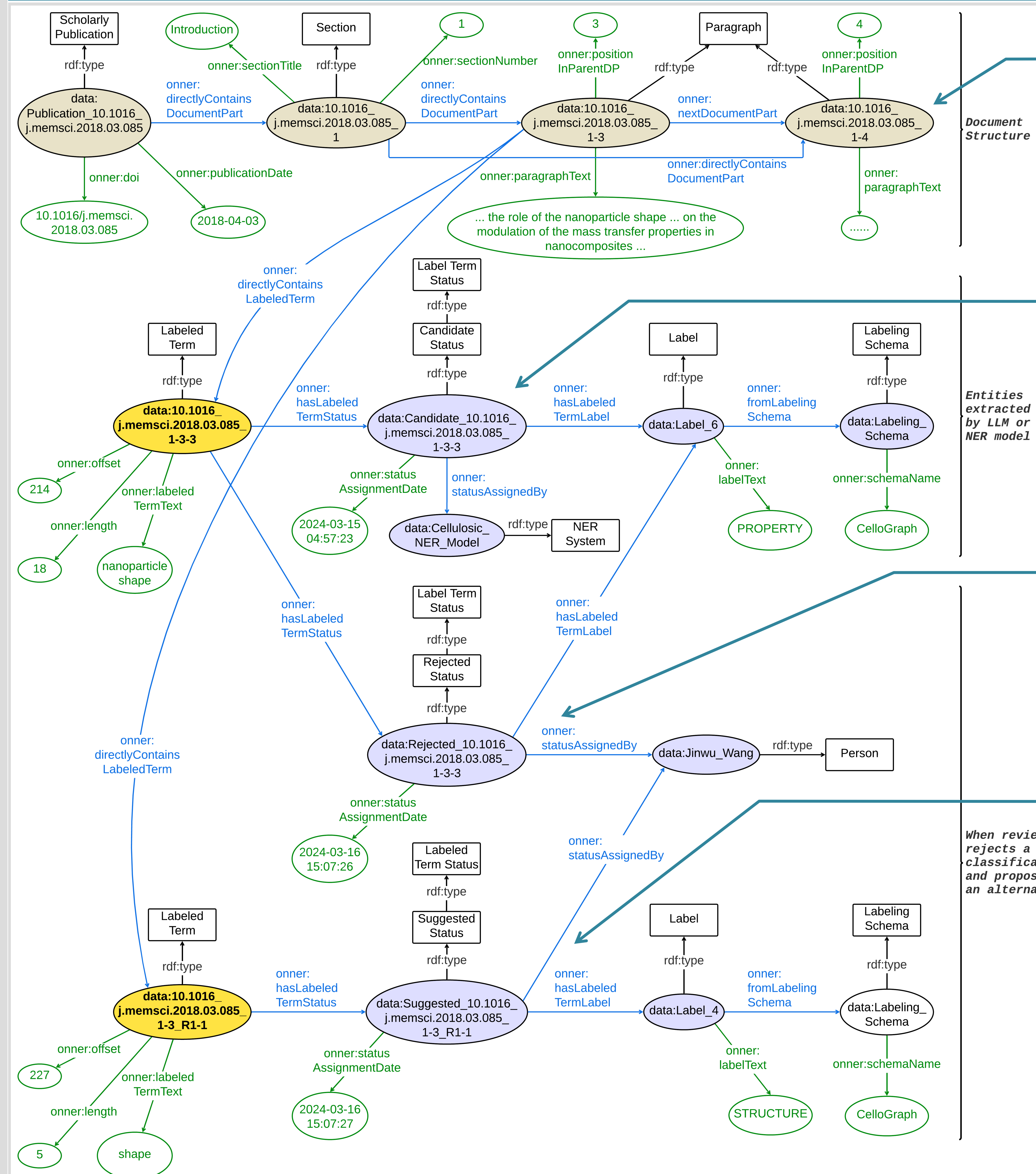


- Document structure is preserved during SciTex processing
- NER-identified entities are enriched and reviewed in AnNER
- Document structure, text content, and reviewed entities populate OnNER
- Identified domain entities can be linked to terms (classes) from an existing domain ontology or help extend a domain ontology

Acknowledgement:

This research is supported in part by the U.S. Department of Agriculture: Forest Service, Project 20-JV-1111124-055, USDA Agricultural Research Service (ARS), Project 0204-41510-001-98S, National Institute of Food and Agriculture (NIFA), Award 2021-67022-34366.

Entity Provenance Tracking in the Constructed Knowledge Graph



Every document, its sections, the paragraphs within those sections, and the entities within those paragraphs are each assigned a unique identifier.

All entities extracted by NER model are initially assigned the *Candidate* status.

An entity's status is updated to *Accepted/Rejected* after a review by a domain expert in AnNER.

Alternatively, reviewers can suggest changes to the span, or to the label assigned to an entity, resulting in an entity with the *Suggested* status.

References:

- U. Reza, N. Pease, J. Wang, T. Hahmann (2025). AnNER: Supporting Efficient Entity Annotation and Review Workflow for Knowledge Graph Construction, K-CAP '25.
- U. Reza, X. Zhang, T. Hahmann (2024). OnNER: An Ontology for Semantic Representation of Named Entities in Scholarly Publications, FOIS '24.