Title: Enhancing Liked Data Trustability and Transparency through Knowledge-driven Data Ecosystems

Knowledge graphs (KGs) have momentum as expressive data structures to represent the convergence of data and knowledge spread across various data sources. Albeit coined by the research community for several decades, KGs play an increasingly relevant role in scientific and industrial areas. In particular, the rich amount of biomedical data in encyclopedic KGs like DBpedia and Wikidata, or domain-specific KGs (e.g., Bio2RDF or KnowLife) demonstrate the feasibility of integrating factual domain-specific knowledge following the Linked Data principles.

Data integrated into existing KGs are collected in heterogeneous formats or physically distributed over multiple sites. Years of research on semantic data management and knowledge engineering have paved the way for merging numerous data sources into a vast and distributed KG, the Linked Open Data (LOD) cloud. Despite accepting data management systems as crucial data processing tools for industrial and scientific database applications, the scenario is not necessarily the same for KG-driven solutions that demand analytics against LOD. Real-world applications require a complete understanding of all the decisions made during data management. Unfortunately, the absence of algorithmic methods to account for LOD transparency considerably affects trustability and prevents their full acceptance as reliable solutions for decision-making.

This talk will present the challenges faced at data integration, query processing, and knowledge engineering levels to empower the pipelines of KG creation with transparency. Solutions for query processing and data management over data integration systems represent the baselines. Moreover, we will explain the role of knowledge extraction, mapping languages, integrity constraints, and provenance towards data transparency and traceability. We will further discuss knowledge-driven data ecosystems as reference architectures to provide the foundations for transparent KG-driven frameworks to enhance trustability. We will illustrate our proposed approach potential in the context of the medical domain, where data transparency is crucial for building trustable solutions to support decision-making.