Defining a Knowledge Graph Development Process Through a Systematic Review

GYTĖ TAMAŠAUSKAITĖ and PAUL GROTH, University of Amsterdam, Netherlands

Knowledge graphs are widely used in industry and studied within the academic community. However, the models applied in the development of knowledge graphs vary. Analysing and providing a synthesis of the commonly used approaches to knowledge graph development would provide researchers and practitioners a better understanding of the overall process and methods involved. Hence, this article aims at defining the overall process of knowledge graph development and its key constituent steps. For this purpose, a systematic review and a conceptual analysis of the literature was conducted. The resulting process was compared to case studies to evaluate its applicability. The proposed process suggests a unified approach and provides guidance for both researchers and practitioners when constructing and managing knowledge graphs.

CCS Concepts: • **Software and its engineering** → **Software development process management**; • **Computing methodologies** → *Ontology engineering*; *Semantic networks*; • **Information systems** → *Semantic web description languages*; *Information integration;*

Additional Key Words and Phrases: Knowledge graphs, knowledge graph construction, development process semantic network, information integration

ACM Reference format:
Gytė Tamašauskaitė and Paul Groth. 2023. Defining a Knowledge Graph Development Process Through a Systematic Review. *ACM Trans. Softw. Eng. Methodol.* 32, 1, Article 27 (February 2023), 40 pages.
https://doi.org/10.1145/3522586

1 INTRODUCTION

Knowledge graphs—graph-structured knowledge bases [57]—are widely employed to represent structured knowledge and perform a variety of AI driven tasks in the context of diverse, dynamic, and large-scale data [32, 87]. Given this increasing adoption, there is a need for guidance on knowledge graph development that would assist researchers, developers, and engineers in the process of creating and maintaining knowledge graphs [9]. While there are descriptions of methods for knowledge graph development [37, 80], that outline the necessary steps to take in order to develop a knowledge graph, these methods vary per article and there is a lack of a *global view* of the development of these software artifacts.

While generally applicable development processes exist in such areas as software development [3], ontology construction [26], and knowledge engineering [64]; it is unclear to what extent these existing theories can be directly applied to knowledge graph development, due to the complex combination of data and software used for their construction. Indeed, from a software engineering

Authors' address: G. Tamašauskaitė and P. Groth, University of Amsterdam, Faculty of Science, Postbus 94323 1090 GH, Amsterdam, Netherlands; emails: gyte.tama@gmail.com, p.t.groth@uva.nl.

1049-331X/2023/02-ART27
https://doi.org/10.1145/3522586