Chapter 2
How to Build a Knowledge Graph

Abstract This chapter outlines the state of the art of Knowledge Graph technologies by introducing the process of building a Knowledge Graph. We define the following major steps of an overall process model: (1) knowledge creation, (2) knowledge hosting, (3) knowledge curation, and (4) knowledge deployment. We demonstrate the methodology for the knowledge creation process that creates, extracts, and structures the fact base for a Knowledge Graph. We describe the process of knowledge collection, storage, and retrieval that implements established knowledge in a graph-based storage system. We analyze existing methods and tools to improve the quality of a large Knowledge Graph. For the Knowledge Curation process, we establish sub-steps, such as knowledge assessment, cleaning, and enrichment. For each of them, we determine various categories and dimensions that have been developed and described in the literature and identify tasks which can be applied (e.g., Knowledge Graph completion and correctness, error detection and correction, identifying and resolving duplicates). Finally, we describe the deployment process of a Knowledge Graph based on the following principles: findability, accessibility, interoperability, and reusability.

2.1 Introduction

According to Gómez-Pérez et al. (2017), Knowledge Graph technologies consist of:

- Knowledge representation and reasoning (languages, schema, and standard vocabularies)
- Knowledge storage (graph databases and repositories)
- Knowledge engineering (methodologies, editors and design patterns)
- Knowledge learning, including schema learning and population

This is true for knowledge-based systems. Knowledge Graph methods and techniques must additionally reflect the specific focus on vast amounts of instance data beyond any traditional knowledge base; see Schultz et al. (2012). We identify the following major steps (see Fig. 2.1) in an overall process model. A more detailed task model is provided in Fig. 2.2.

D. Fensel et al., *Knowledge Graphs*, https://doi.org/10.1007/978-3-030-37439-6_2

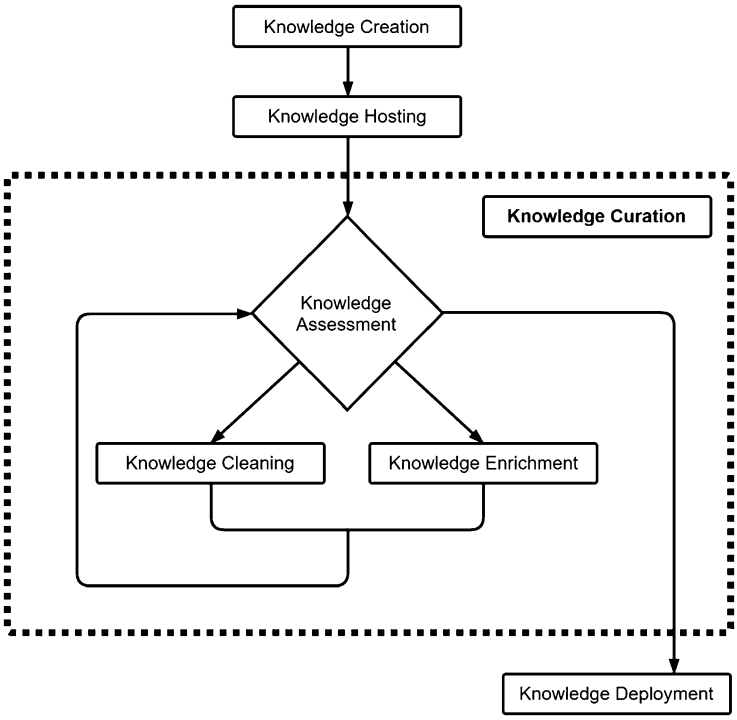

Fig. 2.1 A process model for Knowledge Graph generation

- A traditional knowledge acquisition (or maybe better-called knowledge engineering)[1] phase that establishes the core data for a Knowledge Graph (see Sect. 2.2).
- The process to implement this knowledge in a proper storage system such as a document or graph-based repository (see Sect. 2.3).
- The knowledge curation process (cf. Paulheim 2017) establishes large Knowledge Graphs of significant coverage and quality. As sub-steps of this curation process, we identify the following activities: knowledge assessment, cleaning, and enrichment (see Sect. 2.4).
- Finally, we need to deploy and apply such a Knowledge Graph (see Sect. 2.5).

Each of the mentioned steps is discussed in detail during the following subsections. Similar models can be found, for example, in Gawriljuk et al. (2016), Kejriwal et al. (2017), and Villazón-Terrazas et al. (2017).[2]

[1] A forgotten debate on whether knowledge is elicited or constructed.

[2] A framework for the large-scale integration of publicly available information on points of interest (POI)—that is highly relevant for the touristic area—is described in Athanasiou et al. (2019a, b). See more details on the data integration work bench at the SLIPO website http://slipo.eu/