3. Inversion: from knowledge graphs to raw data

Knowledge graphs are gaining traction nowadays and more and more companies use them, such as Amazon (https://www.amazon.science/tag/knowledge-graphs), Bosch (https://iswc2022.semanticweb.org/wp-content/uploads/2022/11/978-3-031-19433-7_45.pdf), IKEA (https://medium.com/flat-pack-tech/ikeas-knowledge-graph-and-why-it-has-three-layers-a38fca436349), Facebook (https://developers.facebook.com/docs/sharing/opengraph/), Google (https://developers.google.com/knowledge-graph), LinkedIn (https://engineering.linkedin.com/blog/2016/10/building-the-linkedin-knowledge-graph), SIEMENS (https://indico.cern.ch/event/669648/contributions/2838194/attachments/1581790/2499984/CERN_Open_Lab_Technical_Workshop_-_SIEMENS_AG_-_FISHKIN_-_11-01-2018.pdf), Zalando (https://engineering.zalando.com/posts/2021/07/knowledge-graph-master-data-mdm.html), etc. Most knowledge graphs are nowadays constructed from other heterogeneous data sources, such as tables in relational databases, data in XML files or in JSON format derived from a Web API. While the construction of knowledge graphs from heterogeneous data was theoroughly investigated so far, the inverse, namely constructing raw data from knowledge graphs is not explored so far.

goal and method

In this master thesis, we will investigate how we can reuse the same rules that define how to construct knowledge graphs from raw data to construct raw data from knowledge graphs. RML (https://rml.io/specs/rml/) is a broadly used mapping language to construct knowledge graphs from heterogeneous data. However, RML does not support the inverse process, i.e. the construction of raw data in CSV, XML or JSON from knowledge graphs. We will investigate

- how we can leverage RML to construct raw data from heterogeneous data
- how we can extend an existing system (https://rml.io/implementation-report/ and https://doi.org/10.1016/j.websem.2022.100753) or create a new system to construct raw data from knowledge graphs

students: 1