

Why Mapping Rules are Important for Constructing your Knowledge Graph?



David Chaves-Fraga

✉ david.chaves@kuleuven.be

🐦 @dchavesf

A bit about myself...

PhD in Artificial Intelligence (2021)



“Knowledge Graph Construction from Heterogeneous Data Sources Exploiting Declarative Mapping Rules”

Co-chair W3C CG on Knowledge Graph Construction (2019-now)



Joint postdoctoral researcher (2022- now)

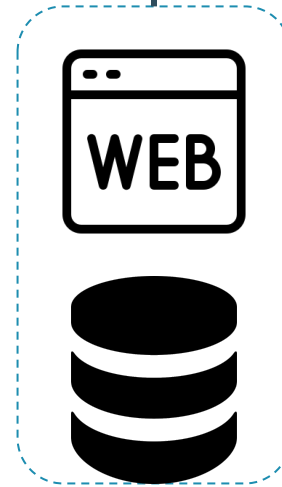
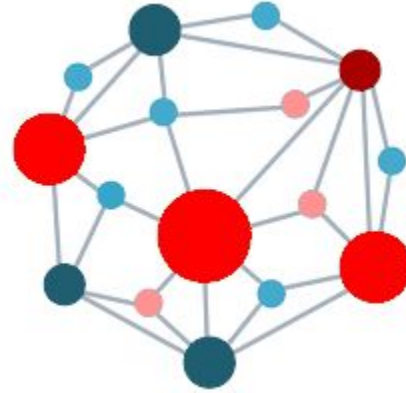
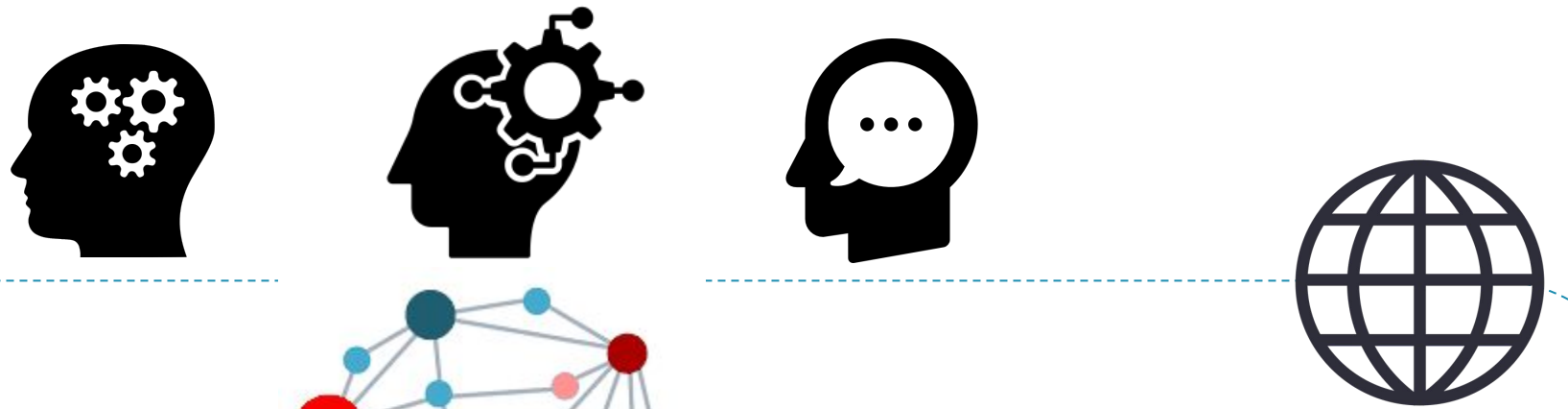


Workshop Organizer (2019-now)

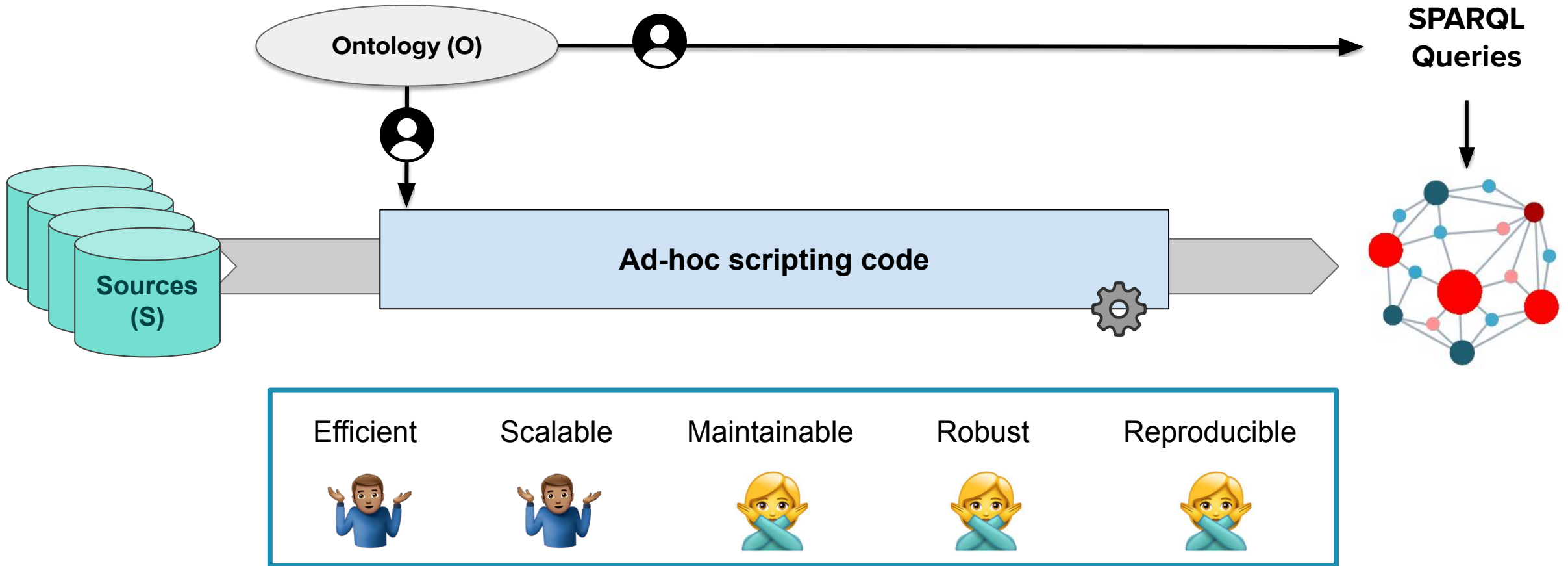


Introduction

- Efficient
- Scalable
- Maintainable
- Robust
- Reproducible

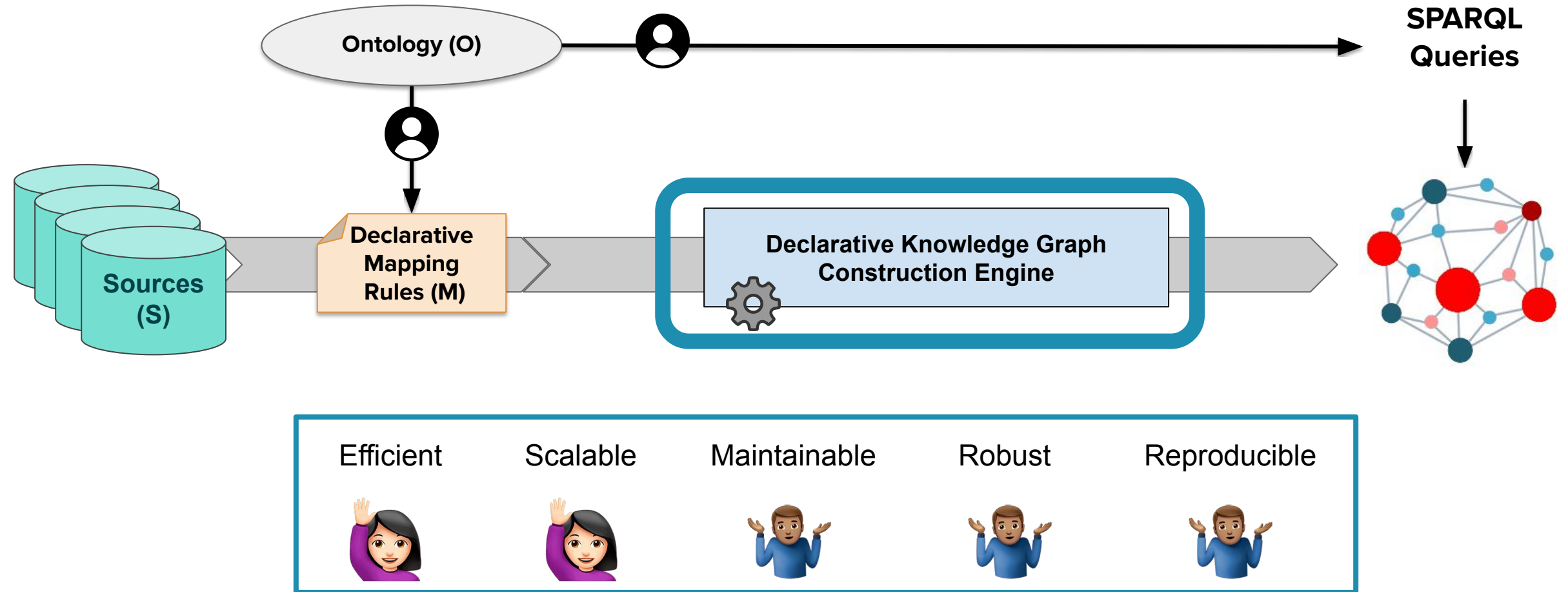


Knowledge Graph Construction: Scripting-based

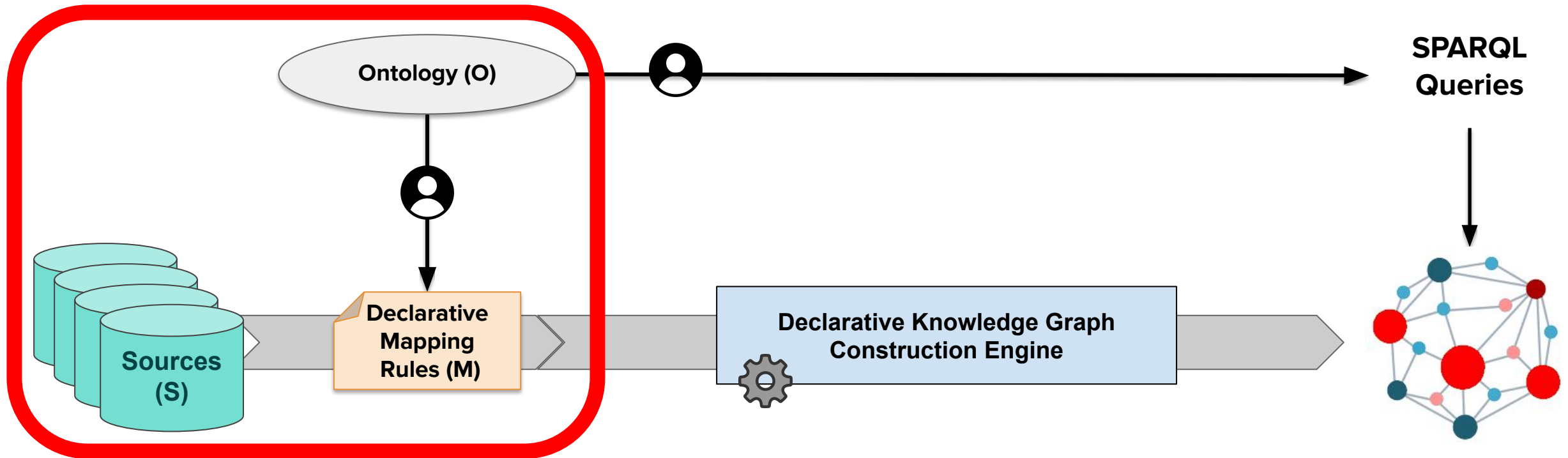


Knowledge Graph Construction

Knowledge Graph Construction = Data Integration System (DIS) = <S, M, O>



KG Construction with Mapping Rules



Knowledge Graph Construction: Scenarios



Partner/company wants to create a KG with its own data...

Scenario 1

No vocab/ontology
Domain expert w/o technical skills

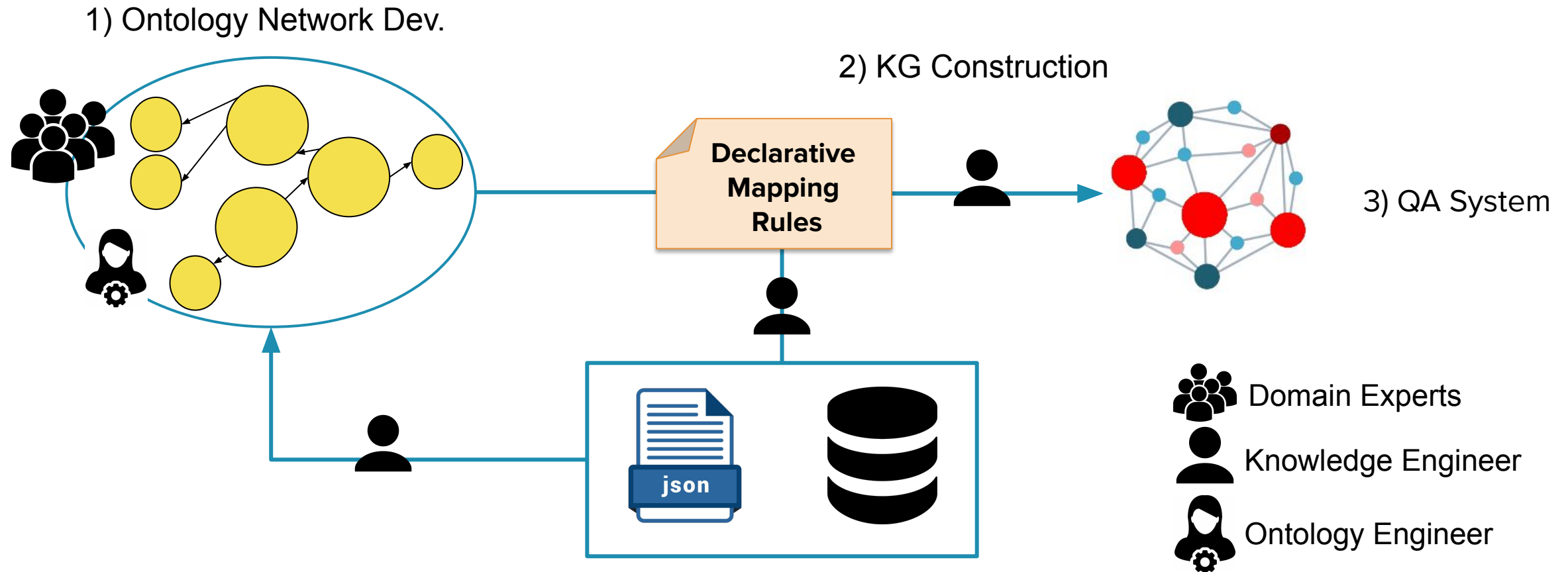
Scenario 2

Vocab/ontology defined (e.g., standard)
Domain expert with technical skills

Scenario 3

Target KG (e.g., DBpedia)
Reduce domain experts effort

S1: Everything under your control (more or less)



Corcho, O., Chaves-Fraga, D., et al. (2021). A High-Level Ontology Network for ICT Infrastructures. In *International Semantic Web Conference* (pp. 446-462)

S1: Everything under your control (more or less)

Outcomes:

- Development of the ontology network (~ 6 months)
- Mapping templates with OWL2YARRRML (automatic)
- Mapping rule creation in YARRRML (~1 month)
- Complex environment for testing/development of a KGC engine
 - Morph-KGC* (<https://github.com/oeg-upm/morph-kgc>)

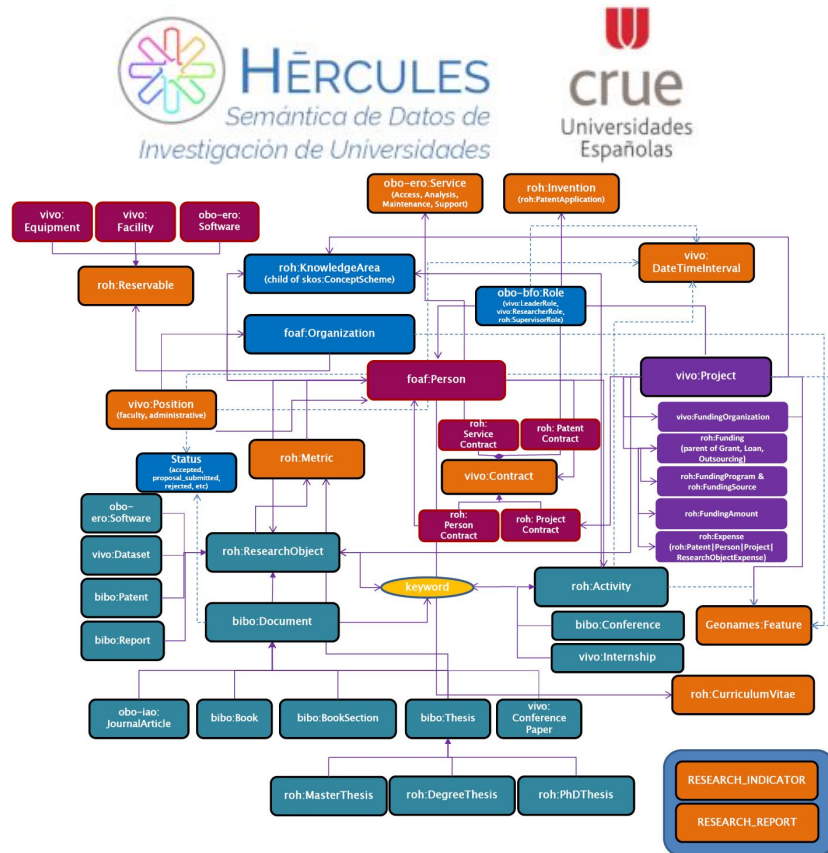
Lessons Learned:

- Simple but useful support tools (OWL2YARRRML)
- Domain experts w/o technical skills → ontology conceptualization
- Independent maintainability difficult to guarantee
- Ontology is stable → mapping rules key resource for the KGC



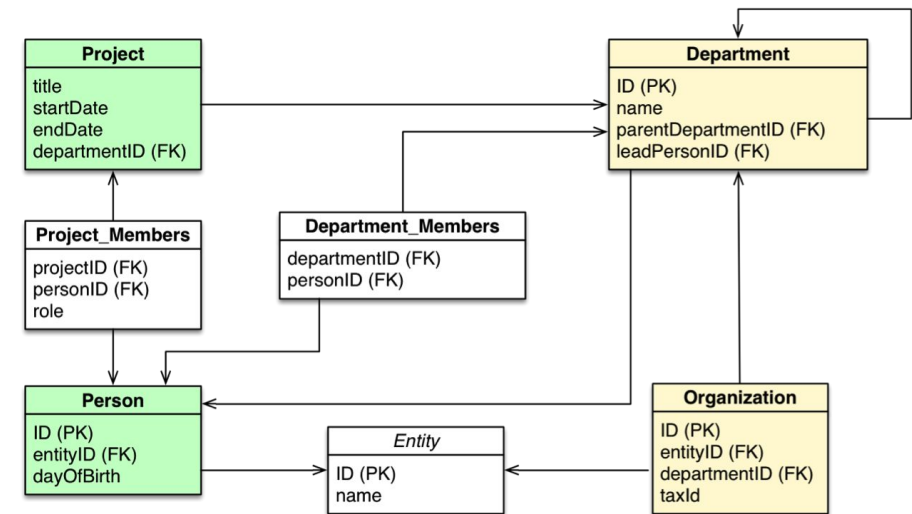
* Arenas-Guerrero, J., Chaves-Fraga, D., Toledo, J., Pérez, M. S., & Corcho, O. (2022). Morph-KGC: Scalable knowledge graph materialization with mapping partitions. *Semantic Web Journal* (accepted).

S2: Nothing under your control




- Ontology v0.3 (will change)
- Not standard documentation (PDF file)

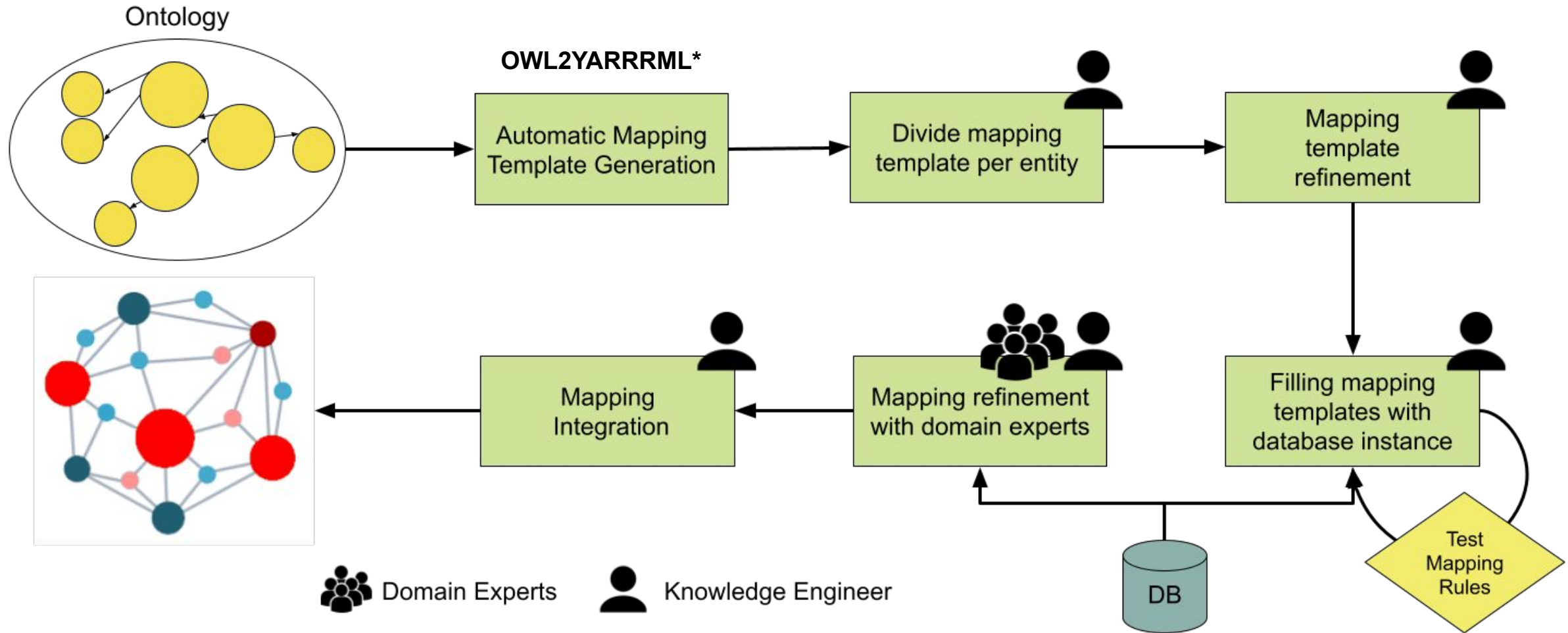
ORACLE



- More than 1800 tables
- Database very well documented
- Oracle supports R2RML mappings

 Chaves-Fraga, D., Corcho, O., et al. (2022). Systematic Construction of Knowledge Graphs for Research Performing Organizations in Spain (Under Review)

S2: Nothing under your control



* <https://github.com/oeg-upm/owl2yarrml>

S2: Nothing under your control

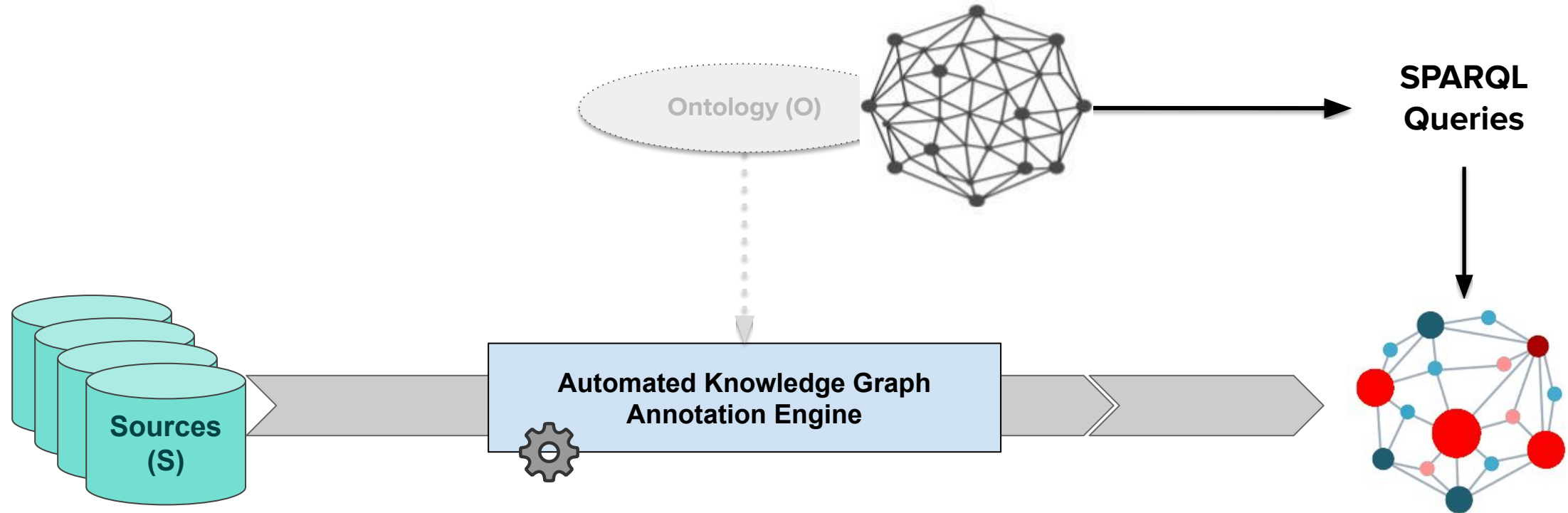
Outcomes:

- Total time: 7 months for mapping creation
- More than 5K rules in R2RML (N-Triples syntax)
- Virtual KG over each entity
- Materialized KG to feed a central repository

Lessons Learned:

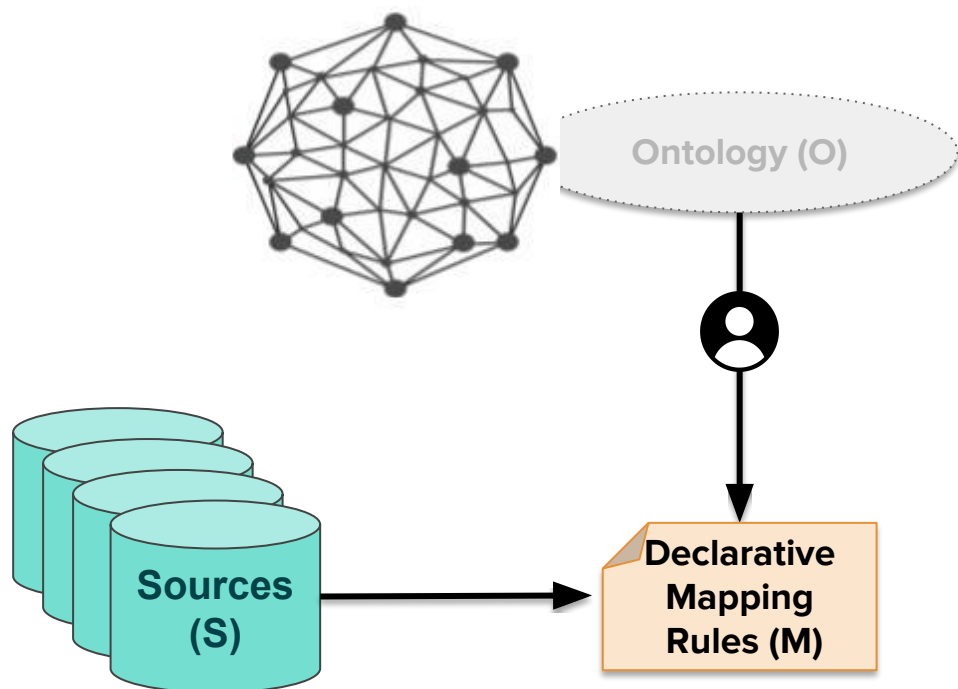
- Simple but useful support tools (OWL2YARRML)
- Domain experts with technical knowledge in the loop
- Divide and Conquer in complex scenarios
- Delegate complex tasks to the DBMS

S3: Automation KG Construction



Chaves-Fraga, D. & Dimou, A. (2022). Declarative Description of Knowledge Graphs Construction Automation: Status & Challenges. In *Third International Workshop on Knowledge Graph Construction co-located with ESWC2022*.

S3: Automation KG Construction



Rules

- Declarative approach
- Understanding the domain
- Target KG
- Linear iteration
- Time consuming
- High quality KG
- Non-reproducible task
- Explainable

Automation

- No manual work
- No knowledge about the domain
- Target Annotation
- Multiple iterations
- Faster
- Quality can be compromised
- Reproducible tasks?
- Non-explainable

RQ1) Are hybrid approaches feasible to explain and optimize a knowledge graph construction process?

RQ2) Can we describe a knowledge graph construction automation process using declarative rules?

Conclusions and Final Remarks

- Mapping rules (in any form) are the central resource of KG generation
- Background of domain experts / users have to be considered
- Adaptability means successful
- Trade-offs: Automation VS Data quality
- Governance of Data Integration Systems (Sources, Mappings, Ontology)

Do you want to know more/get involved? <http://w3id.org/kg-construct>

Awesome KGC tools: <https://github.com/kg-construct/awesome-kgc-tools>



Why Mapping Rules are Important for Constructing your Knowledge Graph?



David Chaves-Fraga

✉ david.chaves@kuleuven.be

🐦 @dchavesf