

Declarative Knowledge Graph Construction: A Practical Introduction

David Chaves-Fraga

✉ david.chaves@kuleuven.be

🐦 @dchavesf



ONTO
COMMONS

ONTOLOGY-DRIVEN
DATA DOCUMENTATION
FOR INDUSTRY COMMONS

Why am I here?

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)



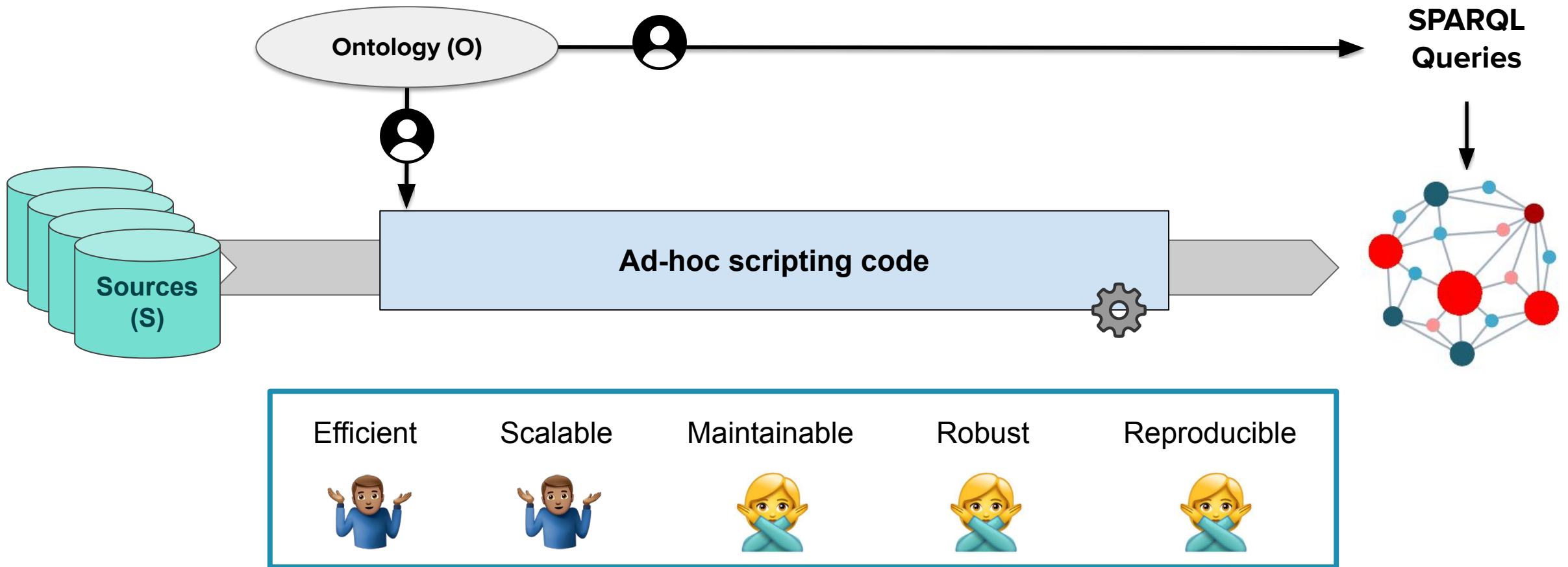
Joint postdoctoral researcher (2022- now)



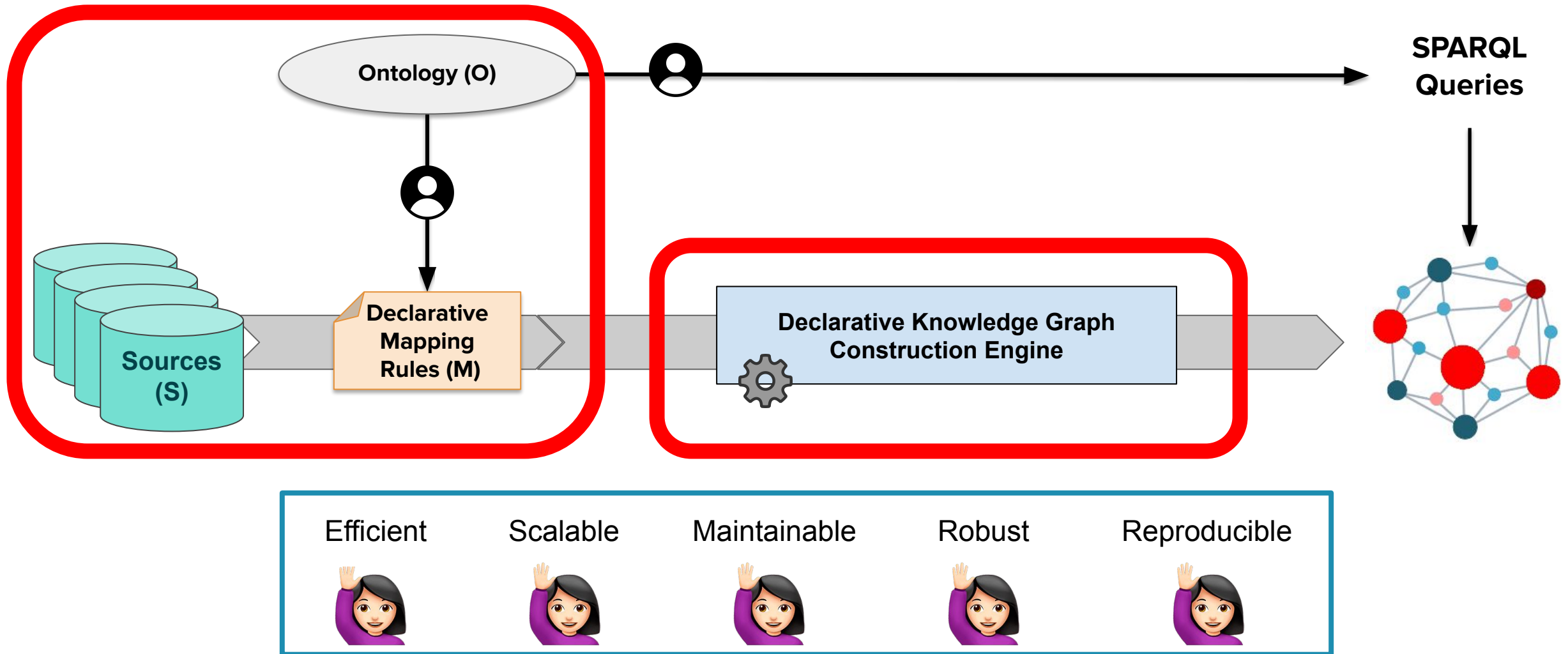
Knowledge Graph Construction Workshop Organizer (2019-now)



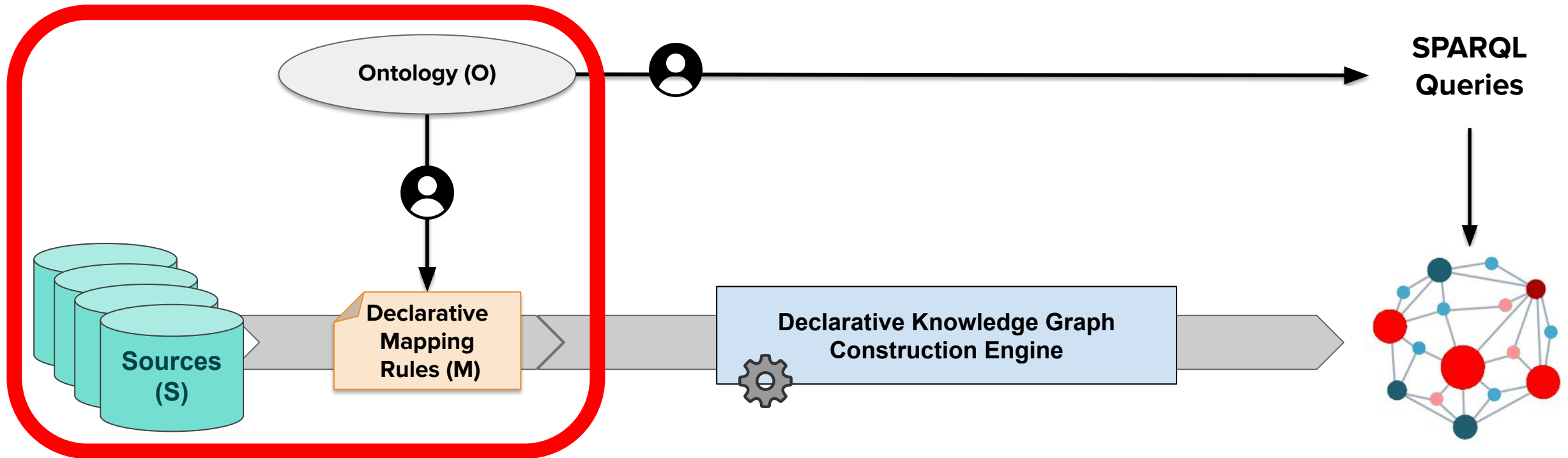
Knowledge Graph Construction: Scripting-based



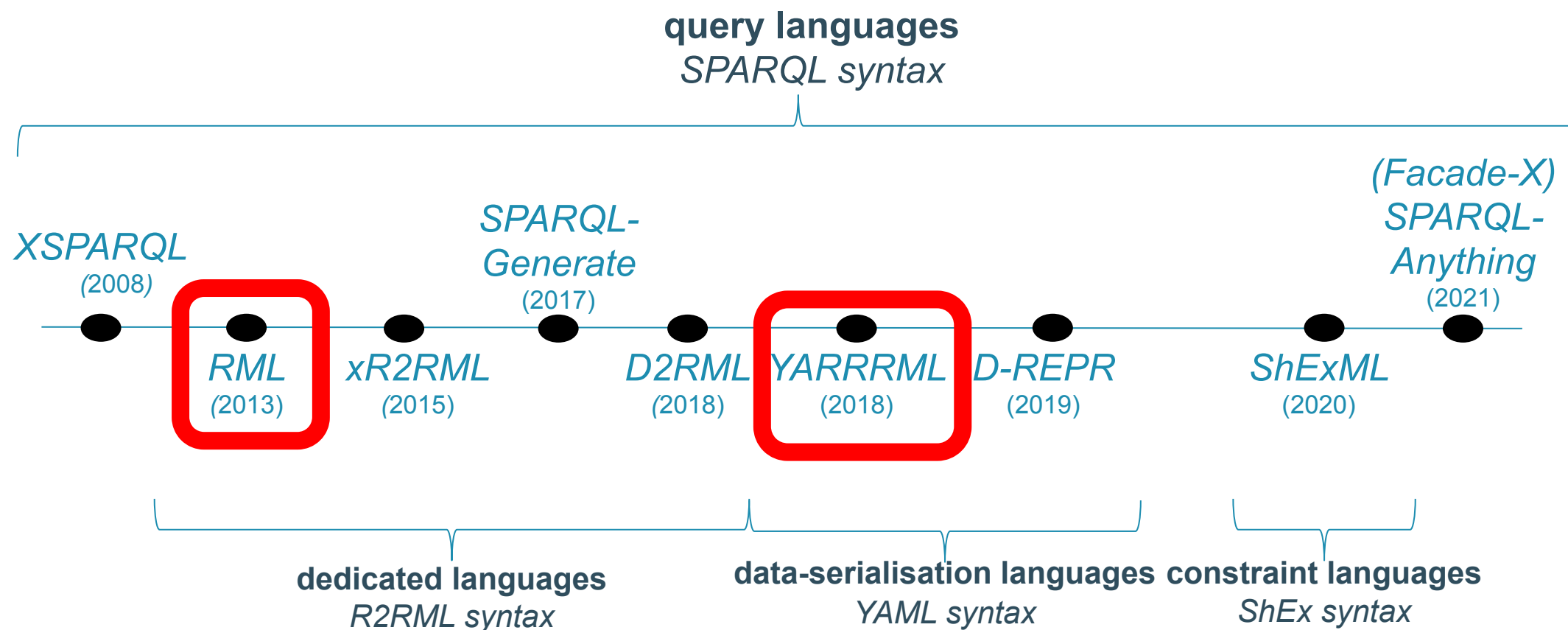
KG Construction with Mapping Rules



Mapping Rules Creation: (YARR)RML



Declarative mapping languages - *schema* transformations

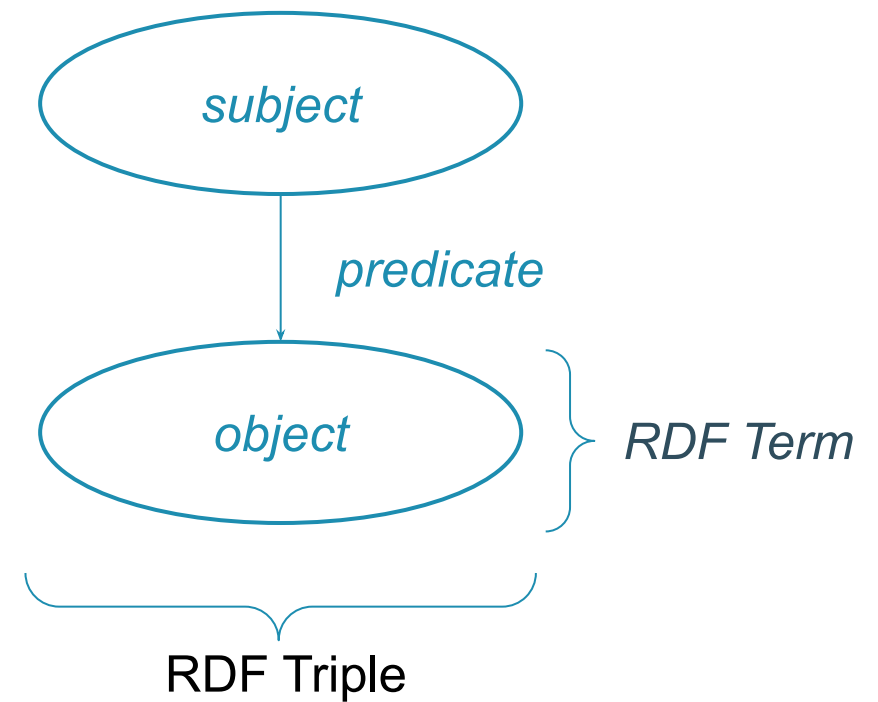


RML spec: <https://rml.io/specs/rml/>

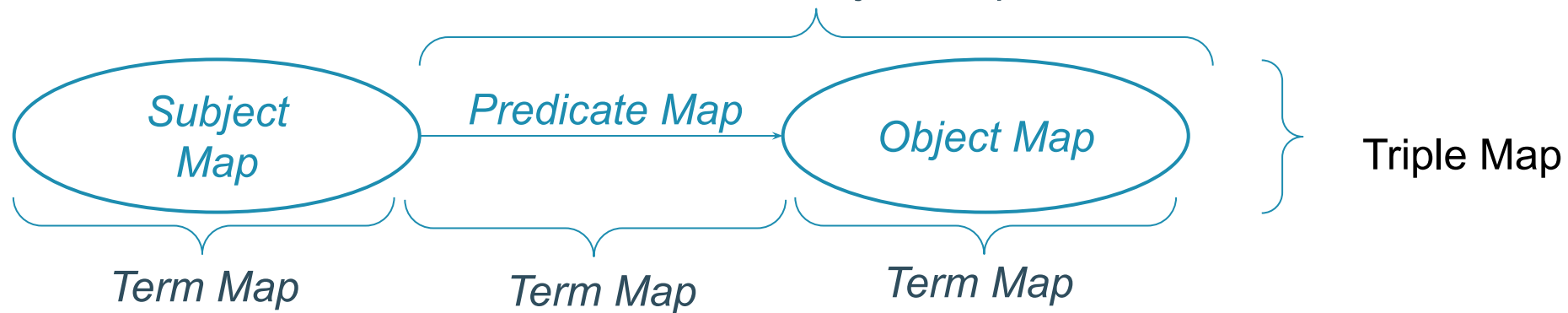
YARRRML spec: <https://rml.io/yarrrrml/spec/>

* Adapted from Knowledge Graph Construction Tutorial @ ESWC 2022 (Anastasia Dimou)

rank	name	nationality	mark	notes
1	Anzhelika Sidorova	Russia	4.95	WL,PB
2	Sandi Morris	USA	4.90	SB
3	Katerina Stefanidi	Greece	4.85	SB
4	Holly Bradshaw	UK	4.80	-
5	Alysha Newman	Canada	4.80	-
6	Angelica Bengtsson	Sweden	4.80	NR



Predicate Object Map



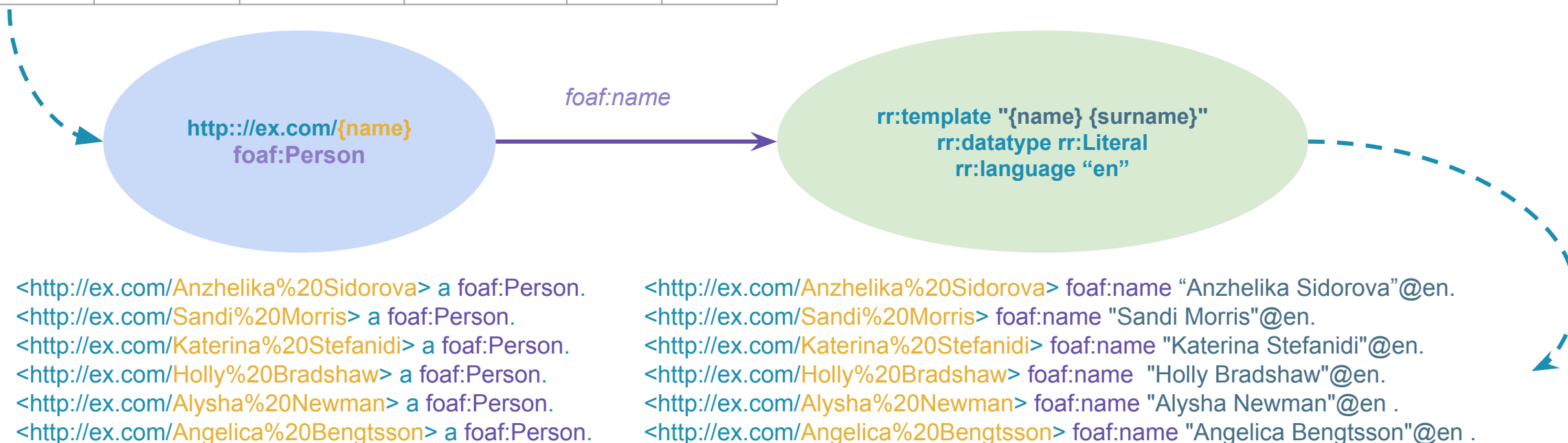
* Adapted from Knowledge Graph Construction Tutorial @ ESWC 2022 (Anastasia Dimou)

rank	name	surname	nationality	mark	notes
1	Anzhelika	Sidorova	Russia	4.95	WL,PB
2	Sandi	Morris	USA	4.90	SB
3	Katerina	Stefanidi	Greece	4.85	SB
4	Holly	Bradshaw	UK	4.80	-
5	Alysha	Newman	Canada	4.80	-
6	Angelica	Bengtsson	Sweden	4.80	NR

```

<#TriplesMap_1> [
  rr:subjectMap [
    rr:template "http://ex.com/{name}";
    rr:class foaf:Person; ]
  rr:predicateObjectMap [
    rr:predicateMap [rr:constant foaf:name];
    rr:objectMap [ rr:template "{name} {surname}";
                  rr:termType rr:Literal;
                  rr:language "en"] ] ].

```



* Adapted from Knowledge Graph Construction Tutorial @ ESWC 2022 (Anastasia Dimou)


```

<#TriplesMap_1> [
  rml:logicalSource [
    rml:source "poleVaulters.csv";
    rml:referenceFormulation ql:CSV; ]; ]
rr:subjectMap [
  rr:template "http://ex.com/{name}"; ];
rr:predicateObjectMap [
  rr:predicateMap [rr:constant ex:score];
  rr:objectMap [ rml:reference "mark";
    rr:datatype xsd:decimal]; ];
rr:predicateObjectMap [
  rr:predicateMap [rr:constant foaf:name];
  rr:objectMap [ rml:reference "name"; rr:language "en"]; ];
rr:predicateObjectMap [
  rr:predicateMap [rr:constant ex:country];
  rr:objectMap [ rr:parentTriplesMap <#TriplesMap_2>;
    rr:joinCondition [
      rr:parent "country_name";
      rr:child "nationality"] ]; ] ].

<#TriplesMap_2> [
  rml:logicalSource [
    rml:source "countries.xml";
    rml:referenceFormulation ql:XPath;
    rml:iterator "countries/country" ];
rr:subjectMap [
  rr:template "http://ex.com/{country_abb}";
  rr:graphMap [ rr:constant ex:CountryGraph ]; ].

```

RML



YARRRML

<https://rml.io/yarrmrl/>

```

mapping:
  person:
    sources:
      - [poleVaulters.csv~csv]
    subject:
      - "http://ex.com/{name}"
    predicateobjects:
      - [ex:score, $(mark), xsd:decimal]
      - [foaf:name, $(name), en~lang]
      - [foaf:name, $(name) $(surname), en~lang]
      - predicates: ex:country
        objects:
          - mapping: country
            condition:
              function: equal
              parameters:
                - [str1, $(nationality), s]
                - [str2, $(country_name), o]
    country:
      sources:
        - [countries.xml~xPath, countries/country]
      subjects: http://ex.com/{country_abb}
      graph: ex:CountryGraph

```

YARRRML

* Adapted from Knowledge Graph Construction Tutorial @ ESWC 2022 (Dylan Van Assche)

Mapping Helpers

User interfaces:

- Matey: <https://github.com/rmlio/matey> (for YARRRML)
- Mapeathor (based on Excel): <https://morph.oeg.fi.upm.es/tool/mapeathor> (for [R2]RML)
- RMLEditor: <https://app.rml.io/rmleditor/> (for RML)

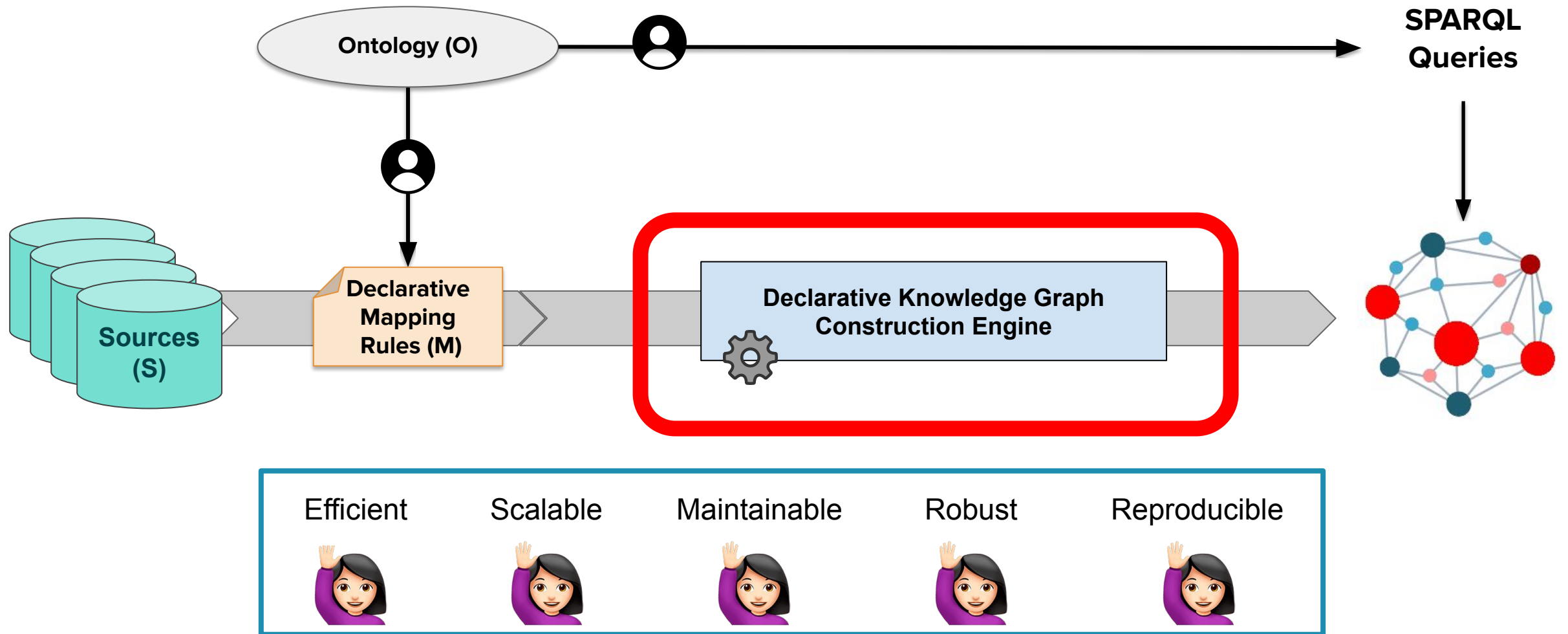
Mapping creators:

- OWL2YARRRML: <https://github.com/oeg-upm/owl2yarrml> (from OWL Ontology to YARRRML template)
- Spread2RML: <https://www.dfki.uni-kl.de/~mschroeder/demo/spread2rml/> (from SpreadSheet to RML)

Mapping translators:

- YARRRML-parser: <https://github.com/rmlio/yarrml-parser> (from YARRRML to [R2]RML)
- Pretty-YARRRML2RML: <https://github.com/oeg-upm/pretty-yarrml2rml> (from YARRRML to Pretty-RML)

Engine for KG Construction



DB2triples (<https://github.com/antidot/db2triples>)
R2RML Parser (<https://github.com/nkons/r2rml-parser>)
XSPARQL (<http://xsparql.sourceforge.net/>)
Morph-KGC (<https://github.com/oeg-upm/morph-kgc>)
R2RML-F (<https://github.com/chrdebru/r2rml>)

RMLMapper: Java (<https://github.com/RMLio/rmlmapper-java>)
CARML: Java (<https://github.com/carmil/carmil>)
RocketRML: JavaScript (<https://github.com/semantifyit/RocketRML>)

RMLStreamer: Flink (<https://github.com/RMLio/RMLStreamer>)
Chimera: Camel (<https://github.com/cefriel/chimera>)

SDM-RDFizer: heuristic-based planning
(<https://github.com/SDM-TIB/SDM-RDFizer>)

FunMap: function-free planning
(<https://github.com/SDM-TIB/FunMap>)

MapSDI: deduplication-based optimizations
(<https://github.com/SDM-TIB/MapSDI>)

Morph-KGC: mapping planning
(<https://github.com/oeg-upm/morph-kgc>)

homogeneous
data sources

R2RML
(RDBs)

Morph-RDB (<https://github.com/oeg-upm/morph-rdb>)
Ontop (<https://github.com/ontop/ontop>)
TripleWave (<https://github.com/streamreasoning/TripleWave>)
SparqlMap-M (<https://github.com/tomatophantastico/sparqlmap>)
Morph-streams++ (<https://github.com/jpcik/morph-streams>)

heterogeneous
data sources

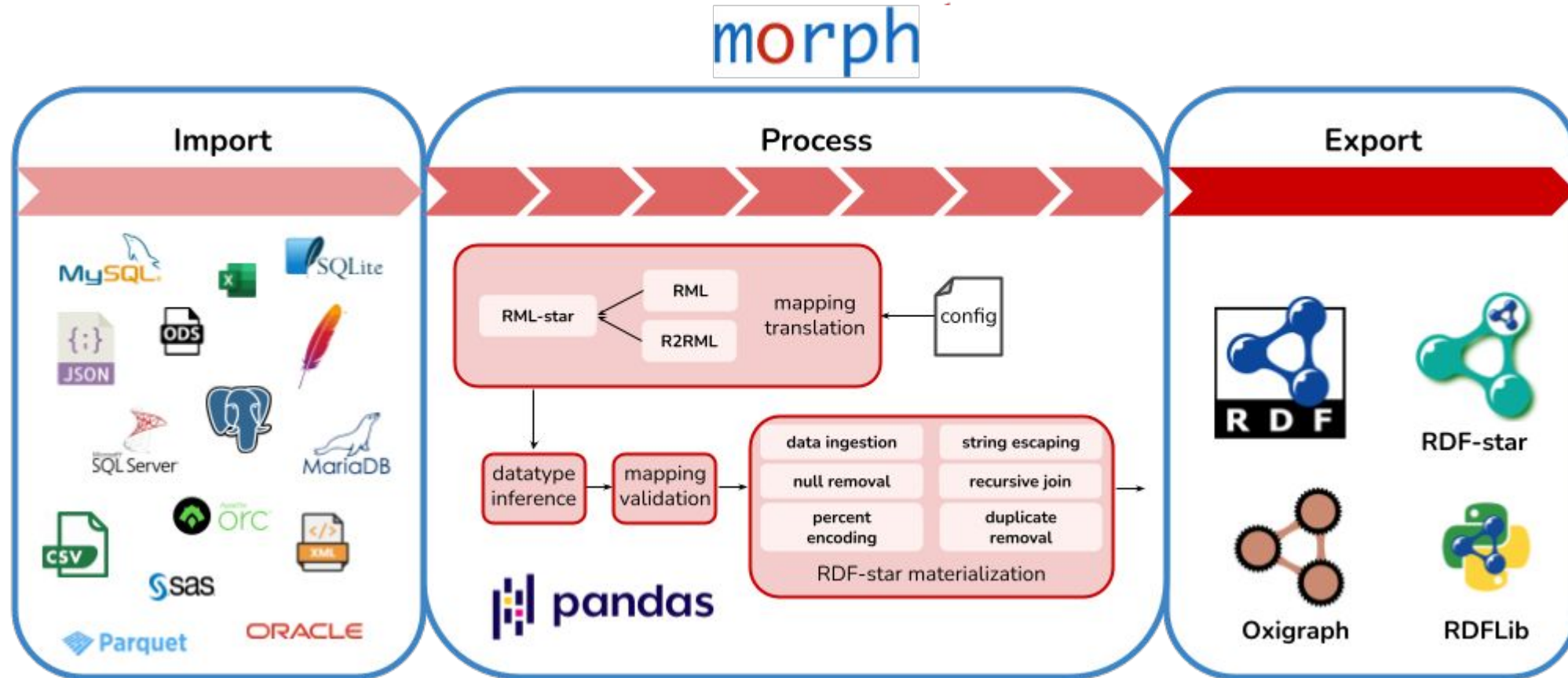
RML
*(RDBs, NoSQL, RDF,
 CSV, XML, JSON, HTML)*

Morph-xR2RML (<https://github.com/frmichel/morph-xr2rml>)
Squerall (<https://github.com/EIS-Bonn/Squerall>)
Ontario (<https://github.com/SDM-TIB/Ontario/>)

virtualisation

* Adapted from Knowledge Graph Construction Tutorial @ ESWC 2022 (Anastasia Dimou)

Morph-KGC: Scalable KG construction



Arenas-Guerrero, J., Chaves-Fraga, D., Toledo, J., S. Perez, M., & Corcho, O. (2022). Morph-KGC: Scalable knowledge graph materialization with mapping partitions. *Semantic Web Journal* (<http://dx.doi.org/10.3233/SW-223135>)



Arenas-Guerrero, J., Iglesias-Molina, A., Chaves-Fraga, D., Garijo, D., Corcho, O. & Dimou, A. (2022). Morph-KGC^{star}: Declarative generation of RDF-star graphs from heterogeneous data. *Under Review at Semantic Web Journal*

Hands-on time!

morph



<https://github.com/oeg-upm/morph-kgc/>



<https://pypi.org/project/morph-kgc/>



<https://morph-kgc.readthedocs.io/>



<https://short.upm.es/umdvms>

All info about Morph suite: <https://morph.oeg.fi.upm.es/>

W3C Community Group

Knowledge Graph Construction

Join us at <http://w3id.org/kg-construct>

Ask questions at <https://github.com/kg-construct/rml-questions/discussions>

Follow our work at <https://github.com/kg-construct/>



Declarative Knowledge Graph Construction: A Practical Introduction

David Chaves-Fraga

✉ david.chaves@kuleuven.be

🐦 @dchavesf



ONTO
COMMONS

ONTOLOGY-DRIVEN
DATA DOCUMENTATION
FOR INDUSTRY COMMONS