KU LEUVEN

Morph-KGC^{star}: Declarative Generation of RDF-star Datasets from Heterogeneous Data

David Chaves-Fraga (w/ Julián Arenas, Ana Iglesias, Daniel Garijo, Oscar Corcho, Anastasia Dimou)

david.chaves@kuleuven.be

@dchavesf

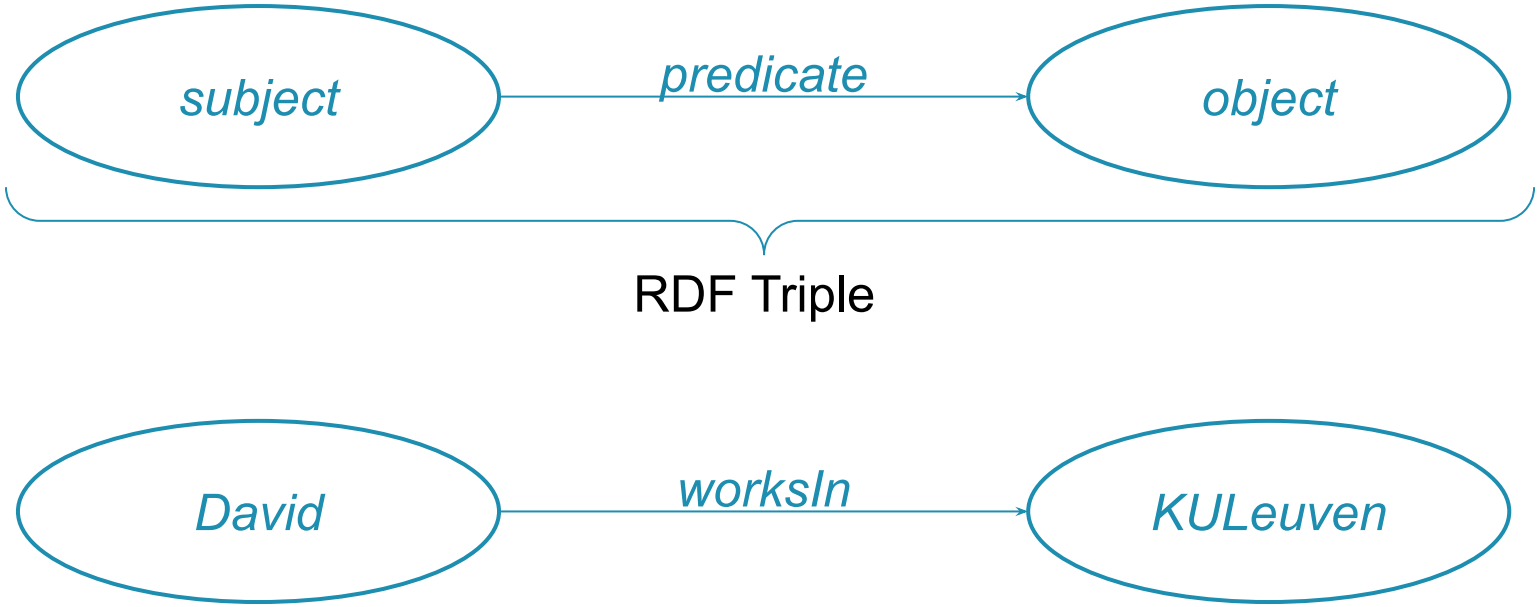
Reject at ISWC 2022
* This work is under review at Semantic Web Journal

What is RDF?

RDF is the **Resource Description Framework** for exchanging data (on the web)

The main data model for **Knowledge Graphs**

A **simple but very powerful model,** for describing **facts**

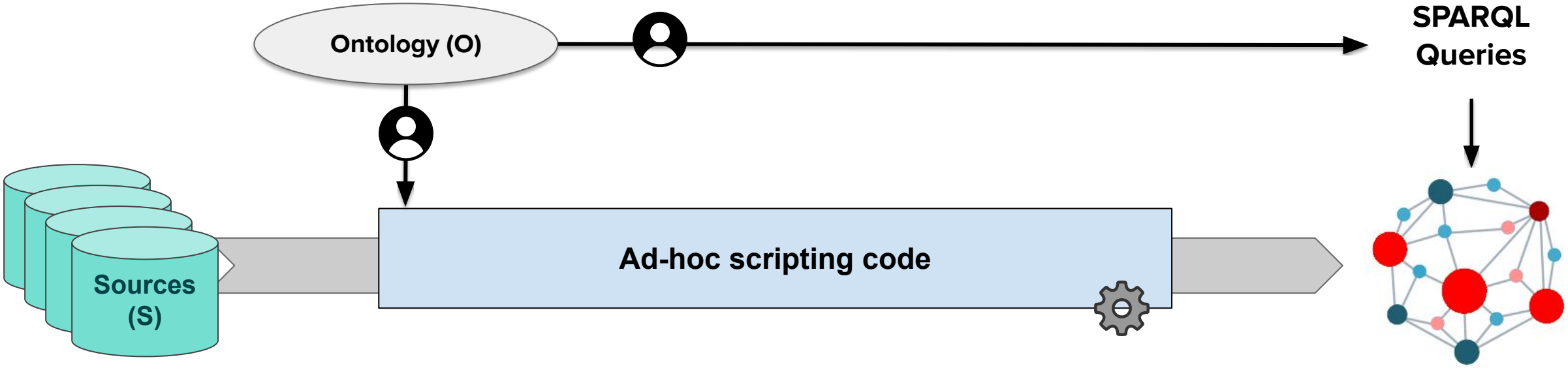


RDF example

<http://ex.com/Anzhelika%20Sidorova> rdf:type foaf:Person.
<http://ex.com/Sandi%20Morris> rdf:type foaf:Person.
<http://ex.com/Katerina%20Stefanidi> rdf:type foaf:Person.
<http://ex.com/Holly%20Bradshaw> rdf:type foaf:Person.
<http://ex.com/Alysha%20Newman> rdf:type foaf:Person.
<http://ex.com/Angelica%20Bengtsson> rdf:type foaf:Person.

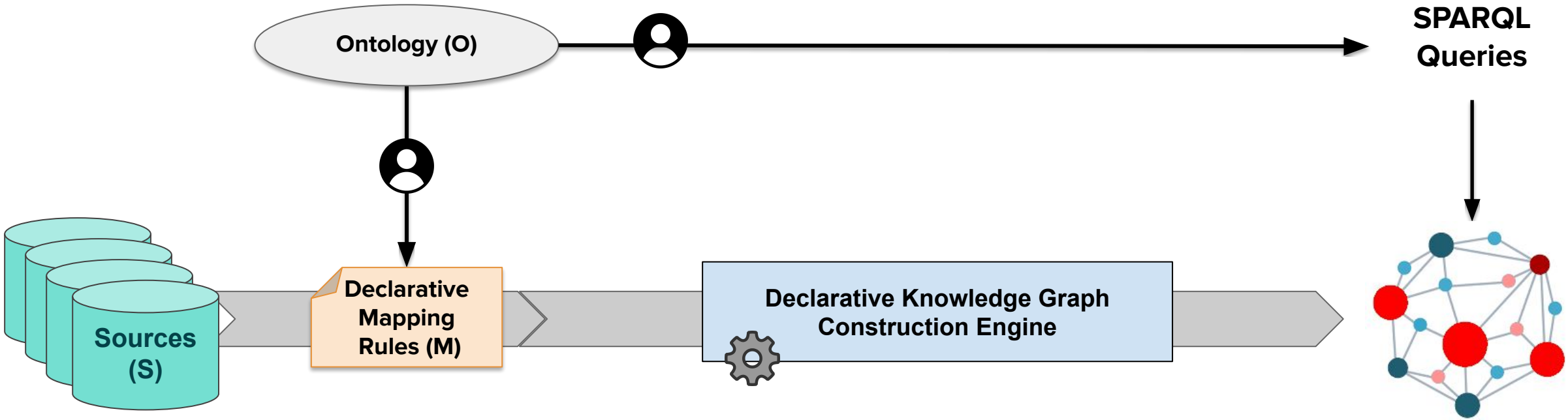
<http://ex.com/Anzhelika%20Sidorova> foaf:name “Anzhelika Sidorova”@en.
<http://ex.com/Sandi%20Morris> foaf:name "Sandi Morris"@en.
<http://ex.com/Katerina%20Stefanidi> foaf:name "Katerina Stefanidi"@en.
<http://ex.com/Holly%20Bradshaw> foaf:name "Holly Bradshaw"@en.
<http://ex.com/Alysha%20Newman> foaf:name "Alysha Newman"@en .
<http://ex.com/Angelica%20Bengtsson> foaf:name "Angelica Bengtsson"@en .

Knowledge Graph Construction: Scripting-based



Efficient	Scalable	Maintainable	Robust	Reproducible
🤷🏽	🤷🏽	🙆	🙆	🙆

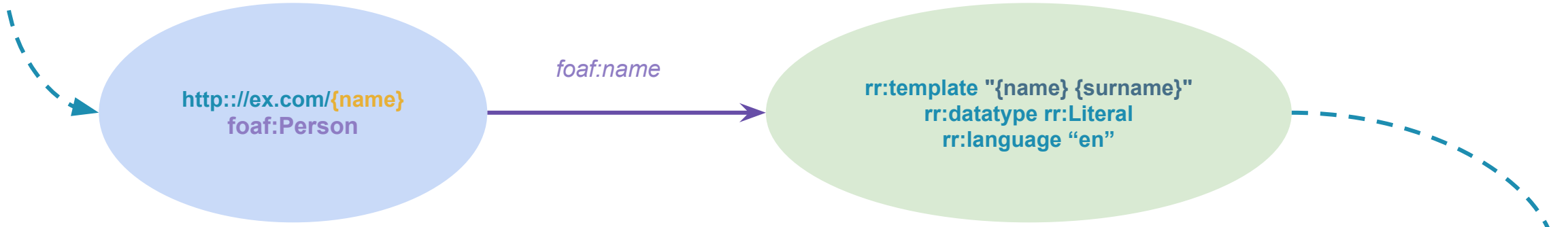
KG Construction with Mapping Rules



Efficient Scalable Maintainable Robust Reproducible

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"] ] ].
```



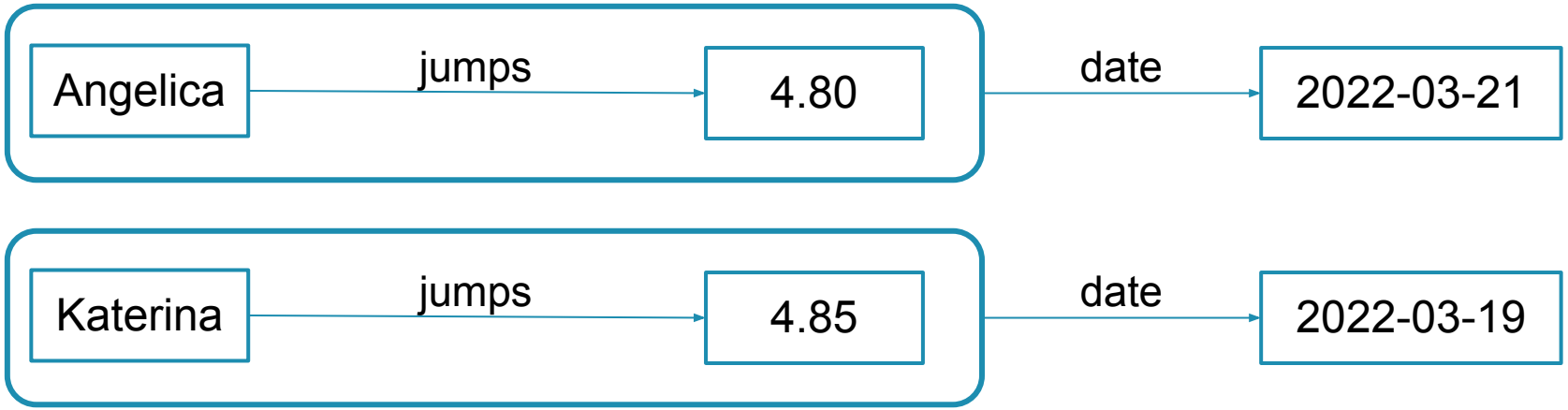
<http://ex.com/Anzhelika%20Sidorova> rdf:type foaf:Person.
<http://ex.com/Sandi%20Morris> rdf:type foaf:Person.
<http://ex.com/Katerina%20Stefanidi> rdf:type foaf:Person.
<http://ex.com/Holly%20Bradshaw> rdf:type foaf:Person.
<http://ex.com/Alysha%20Newman> rdf:type foaf:Person.
<http://ex.com/Angelica%20Bengtsson> rdf:type foaf:Person.

<http://ex.com/Anzhelika%20Sidorova> foaf:name “Anzhelika Sidorova”@en.
<http://ex.com/Sandi%20Morris> foaf:name "Sandi Morris"@en.
<http://ex.com/Katerina%20Stefanidi> foaf:name "Katerina Stefanidi"@en.
<http://ex.com/Holly%20Bradshaw> foaf:name "Holly Bradshaw"@en.
<http://ex.com/Alysha%20Newman> foaf:name "Alysha Newman"@en .
<http://ex.com/Angelica%20Bengtsson> foaf:name "Angelica Bengtsson"@en .

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

Statements about Statements

ID	DATE	MARK	PERSON
1	2022-03-21	4.80	Angelica
2	2022-03-19	4.85	Katerina



Option I: Standard Reification


```
#row 1
_:1 rdf:type rdf:Statement .
_:1 rdf:subject :Angelica .
_:1 rdf:predicate :jumps .
_:1 rdf:object "4.80" .
_:1 :date "2022-03-21" .

#row 2
_:2 rdf:type rdf:Statement .
_:2 rdf:subject :Katerina .
_:2 rdf:predicate :jumps .
_:2 rdf:object "4.85" .
_:2 :date "2022-03-19" .
```

```
<#TM> a rr:TriplesMap ;
rml:logicalSource :marks ;
rr:subjectMap [
    rml:reference "ID" ;
    rr:termType rr:BlankNode ;
    rr:class rdf:Statement ] ;
rr:predicateObjectMap [
    rr:predicate rdf:subject ;
    rr:objectMap [
        rr:template ":{PERSON}" ] ] ;
rr:predicateObjectMap [
    rr:predicate rdf:predicate ;
    rr:object :jumps ] ;
rr:predicateObjectMap [
    rr:predicate rdf:object ;
    rr:objectMap [
        rml:reference "MARK" ] ] ;
rr:predicateObjectMap [
    rr:predicate :date ;
    rr:objectMap [
        rml:reference "DATE" ] ] .
```

Option II: Singleton Property


```
#row 1
:Angelica :jumps#1 "4.80" .
:jumps#1 :date "2022-03-21" .
:jumps#1 rdf:singletonPropertyOf :jumps .

#row 2
:Katerina :jumps#2 "4.85" .
:jumps#2 :date "2022-03-19" .
:jumps#2 rdf:singletonPropertyOf :jumps .
```

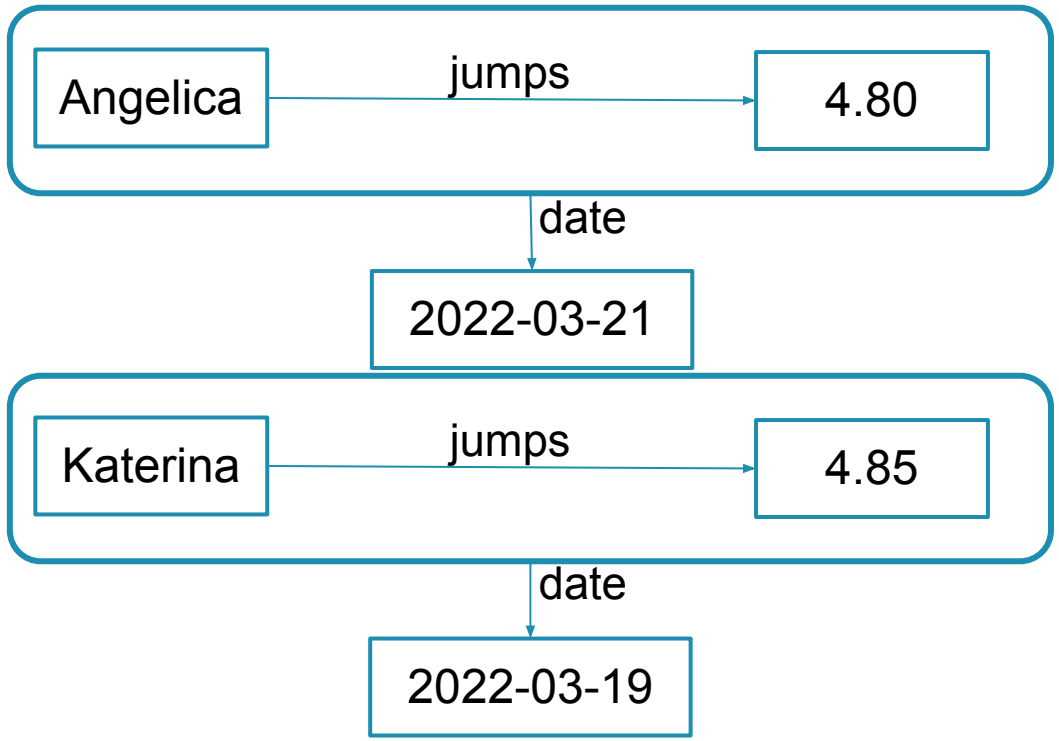
```
<#TM> a rr:TriplesMap ;
rml:logicalSource :marks ;
rr:subjectMap [
    rr:template ":{PERSON}" ] ;
rr:predicateObjectMap [
    rr:predicateMap [
        rr:template ":jumps#{ID}" ] ;
    rr:objectMap [
        rml:reference "MARK" ] ] .

<#TM-SP> a rr:TriplesMap ;
rr:logicalSource :marks ;
rr:subjectMap [
    rr:template ":jumps#{ID}" ] ;
rr:predicateObjectMap [
    rr:predicate :date ;
    rr:objectMap [
        rml:reference "DATE" ];
rr:predicateObjectMap [
    rr:predicate rdf:singletonPropertyOf;
    rr:object :jumps ] ] .
```

The RDF-star solution

Triples that include a **triple as a subject or an object** are known as RDF-star triples

An RDF-star graph is a **set of RDF-star triples**.

SPARQL-star extends SPARQL to query RDF-star graphs



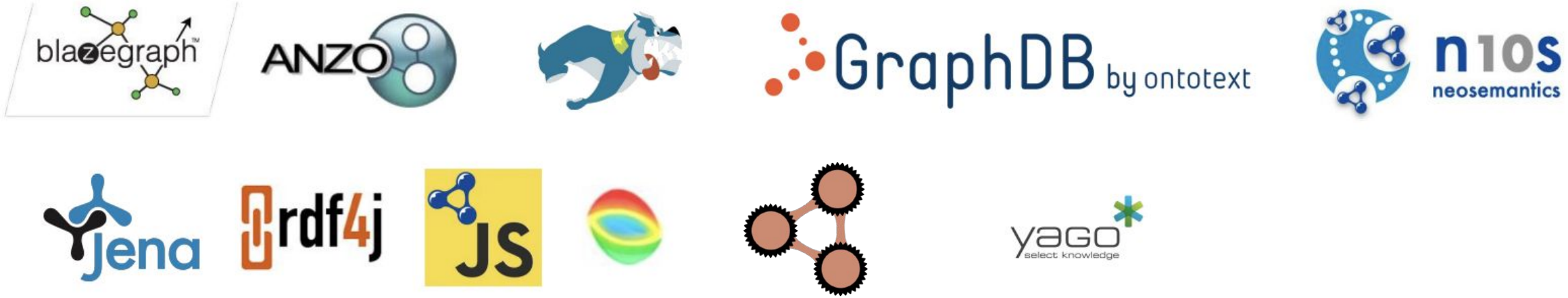

```
<< :Angelica :jumps "4.80" >> :date "2022-03-21" .
<< :Katerina :jumps "4.85" >> :date "2022-03-19" .



SELECT ?jumper ?mark ?date WHERE {
      << ?jumper :jumps ?mark >> :date ?date
}
```

RDF-star features

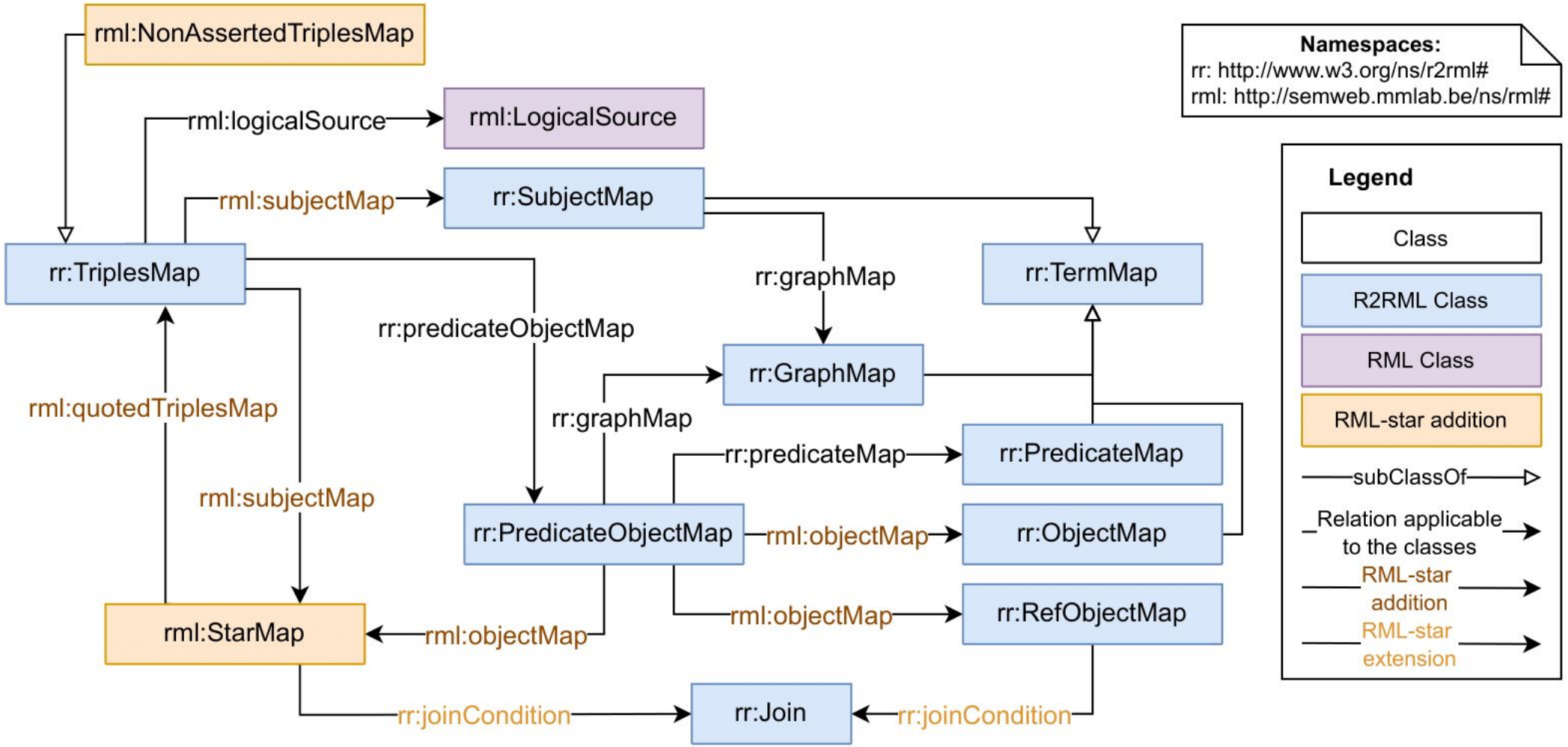
Wide adoption of the approach from industry and vendors

Standardization process through the World Wide Web Consortium (W3C)

No sustainable procedure to generate RDF-star graphs



RML-star: A declarative generator for RDF-star



KU LEUVEN

Option III: RDF-star

```
#row 1
<< :Angelica :jumps "4.80" >>
        :date "2022-03-21" .

#row 2
<< :Katerina :jumps "4.85" >>
        :date "2022-03-19" .
```

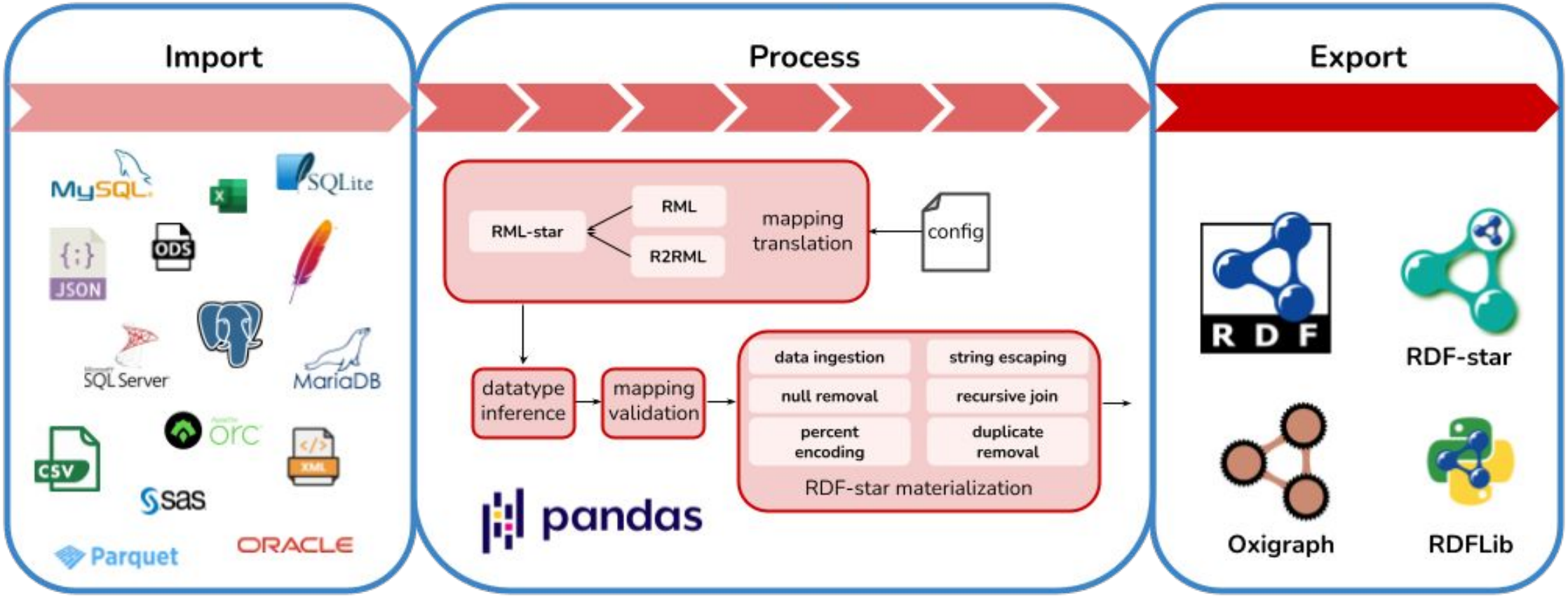
```
<#innerTM> a rml:NonAssertedTriplesMap ;
rml:logicalSource :marks ;
rml:subjectMap [
    rr:template ":{PERSON}" ] ;
rr:predicateObjectMap [
    rr:predicate :jumps ;
    rml:objectMap [
        rml:reference "MARK" ] ] .
```

```
<#outerTM> a rr:TriplesMap ;
rml:logicalSource :marks ;
rml:subjectMap [
    rml:quotedTriplesMap <#innerTM> ] ;
rr:predicateObjectMap [
    rr:predicate :date ;
    rml:objectMap [
        rml:reference "DATE" ] ] .
```

Delva, T., Arenas-Guerrero, J., Iglesias-Molina, A., Corcho, O., **Chaves-Fraga, D., & Dimou, A.** (2021). RML-star: A declarative mapping language for RDF-star generation. In ISWC2021, the International Semantic Web Conference

Morph-KGC^{star}

morph^{star}



KU LEUVEN

Demo time!

morph

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

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

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

https://short.upm.es/umdvm

KU LEUVEN

Use-Cases and Results

morph star

	Biomedical Research Literature				Scientific Software Metadata Extraction			
	Mapping		Generation Time (s)	Number of Output Triples	Mapping		Generation Time (s)	Number of Output Triples
	TriplesMap	POM			TriplesMap	POM		
RML-star	10	10	1,796	36,067,636	78	122	1,085	15,102
Singleton Property	10	15	1,943	75,465,497	78	158	1,124	16,015
Std. Reification	9	20	4,876	27,697,142	39	199	1,201	21,268

Morph-KGC with RML-star is **faster** than Morph-KGC with S.Property/Std. Reification
RML-star **needs less rules** for generating an RDF-star graph equivalent to S. Property/Std. Reification

KU LEUVEN

Conclusions

Declarative solution for constructing RDF-star from heterogeneous data

First implementation with good results in terms of performance

Adoption from the community + well-known engine (68 stars, 8 forks in GitHub)

Future Work

Human-friendly serialization of RML-star (YARRRRML-star)

Implementation of the translation (YARRRML-translator)

Optimizations in the construction of RDF-star datasets

Virtual Knowledge Graph Generation (SPARQL-star to SQL using RML-star)

KU LEUVEN

Morph-KGC^{star}: Declarative Generation of RDF-star Datasets from Heterogeneous Data

David Chaves-Fraga (w/ Julián Arenas, Ana Iglesias, Daniel Garijo, Oscar Corcho, Anastasia Dimou)

david.chaves@kuleuven.be

@dchavesf