KU LEUVEN LEUVEN.AI

Aligning heterogeneous semi-structured data and knowledge graphs

Anastasia Dimou
📨 anastasia.dimou@kuleuven.be
🐦 @natadimou

my story so far…

Tenure-track assistant professor
Data & Cost-efficient ML & AI

Declarative Languages & Artificial Intelligence (**DTAI**) dtai.cs.kuleuven.be/
Computer Science Department
wms.cs.kuleuven.be
KULeuven kuleuven.be/

Leuven.AI ai.kuleuven.be/
KULeuven Institute for Artificial Intelligence

Flanders Make
flandersmake.be/

KU LEUVEN

Knowledge Graphs construction

Knowledge Graphs are

crowdsourced, e.g., Wikidata.org

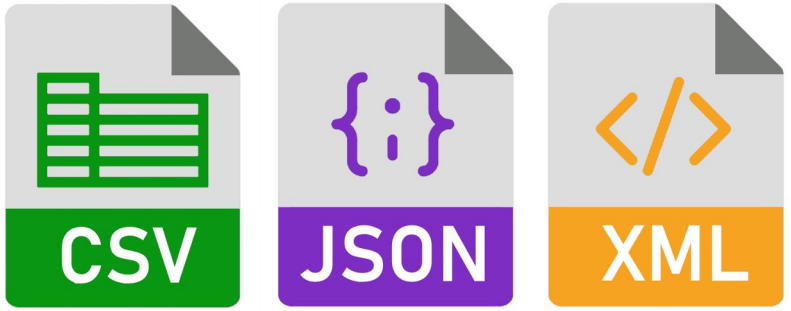
extracted from

unstructured data, e.g., plain text

(semi-)structured data

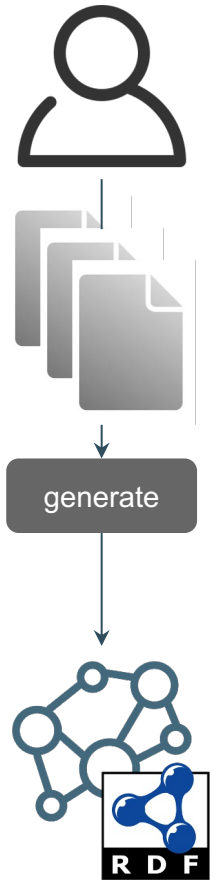
tabular data, e.g., DBs, CSV

hierarchical data, e.g., XML, JSON



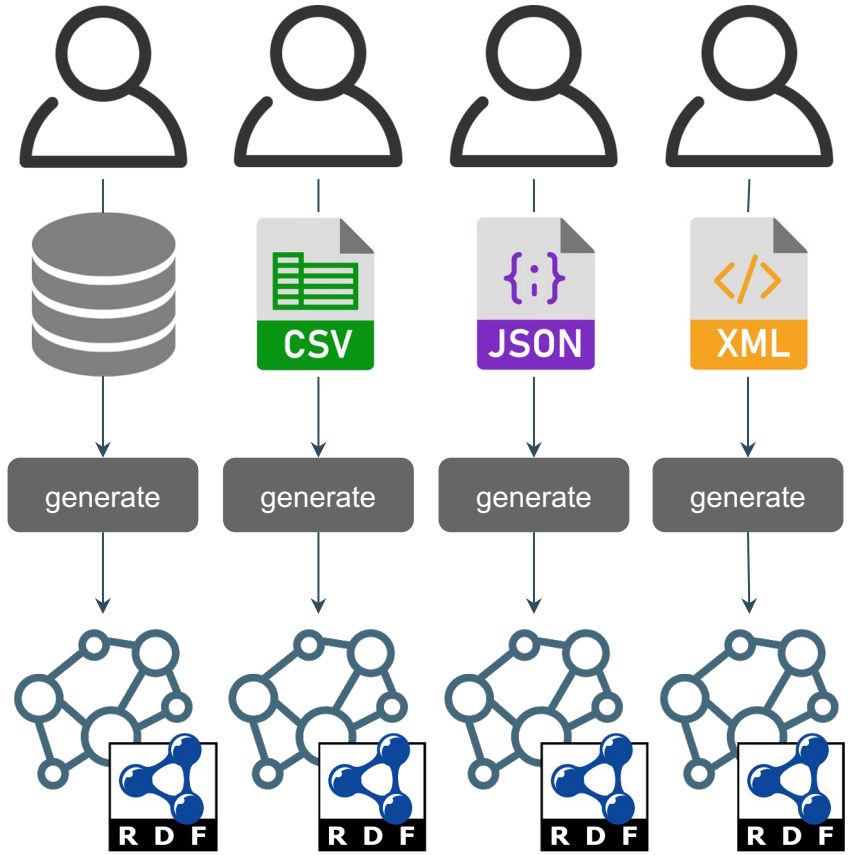
Declarative RDF graph generation from heterogeneous (semi-)structured data: a Systematic Literature Review.
D. Van Assche, T. Delva, G. Haesendonck, P. Heyvaert, B. De Meester, A. Dimou. 2022. JWS
https://doi.org/10.1016/j.websem.2022.100753

hard-coded



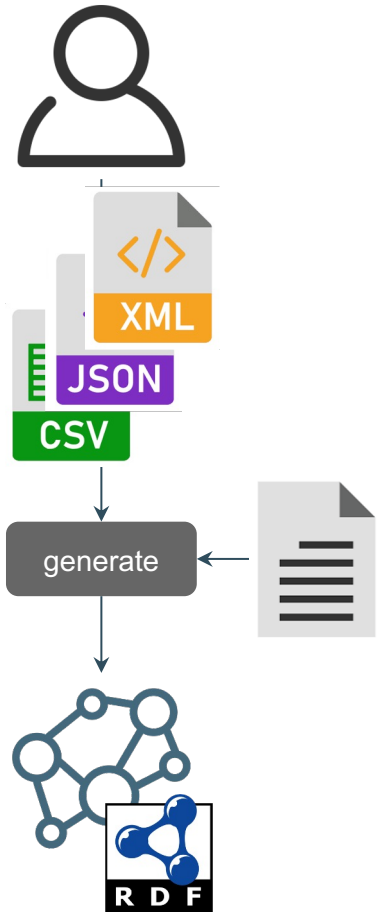
(–) new development cycle for every modification

format-specific



(–) learn & maintain multiple solutions if data in different formats

declarative

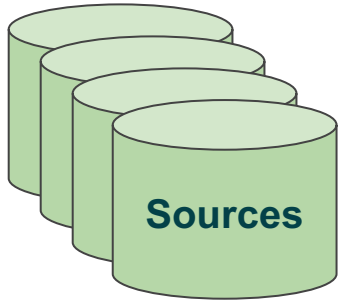


(–) learn & maintain single solutions
(+) declarative describe rules

Declarative RDF graph generation from heterogeneous (semi-)structured data: a Systematic Literature Review.
D. Van Assche, T. Delva, G. Haesendonck, P. Heyvaert, B. De Meester, A. Dimou. 2022. JWS
https://doi.org/10.1016/j.websem.2022.100753

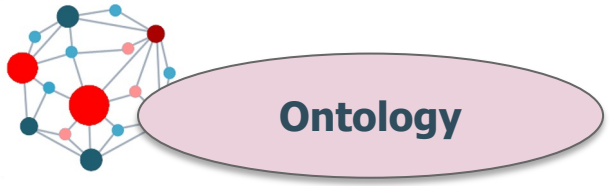
Declarative Knowledge Graph Construction

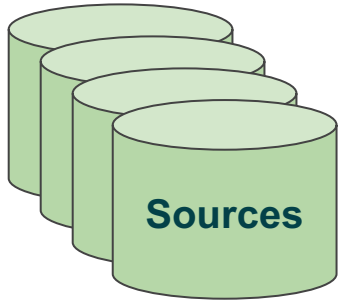
rank	name	mark
1	Anzhelika Sidorova	4.95
2	Sandi Morris	4.90
3	Katerina Stefanidi	4.85
4	Holly Bradshaw	4.80
5	Alysha Newman	4.80
6	Angelica Bengtsson	4.80



Declarative Knowledge Graph Construction

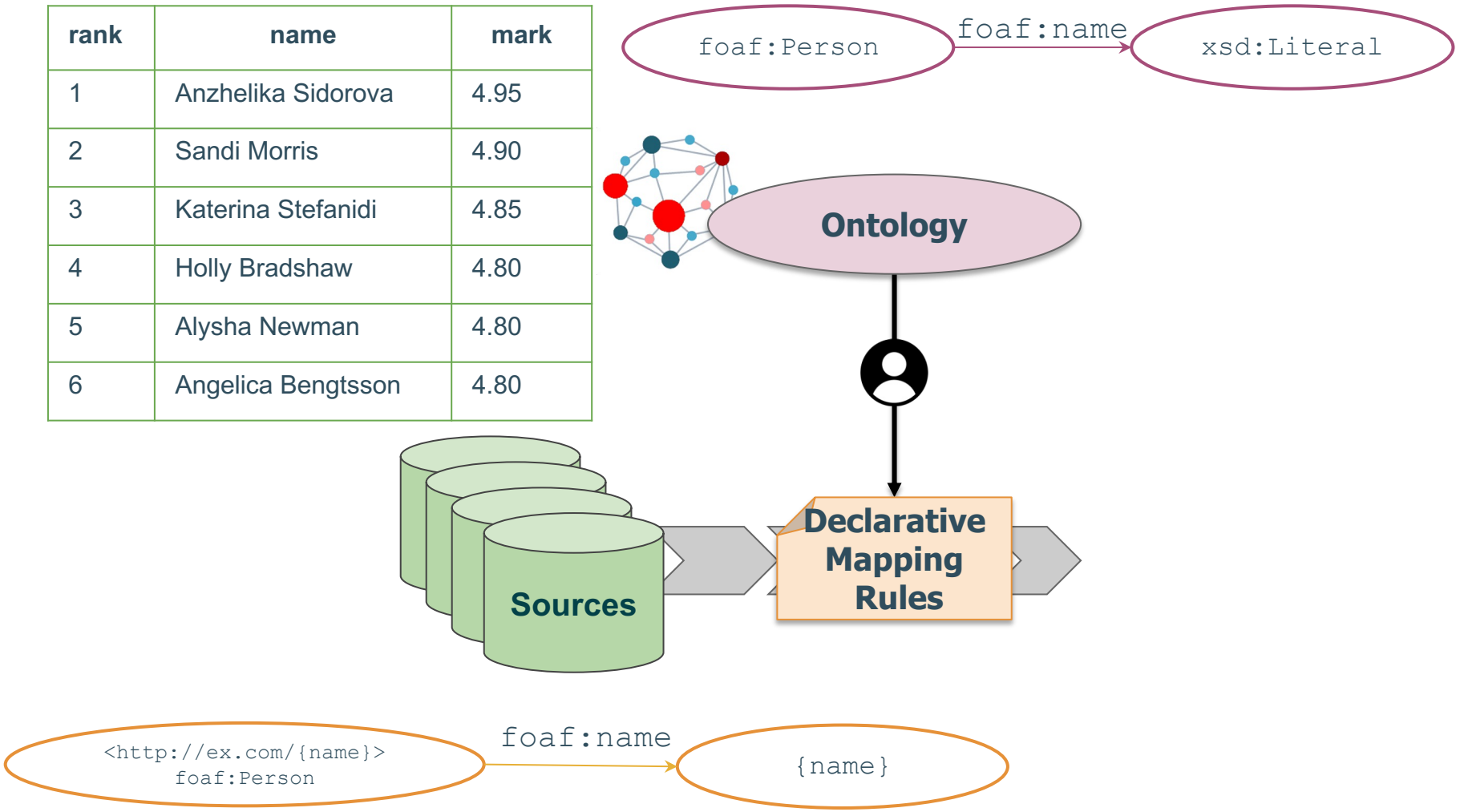
rank	name	mark
1	Anzhelika Sidorova	4.95
2	Sandi Morris	4.90
3	Katerina Stefanidi	4.85
4	Holly Bradshaw	4.80
5	Alysha Newman	4.80
6	Angelica Bengtsson	4.80





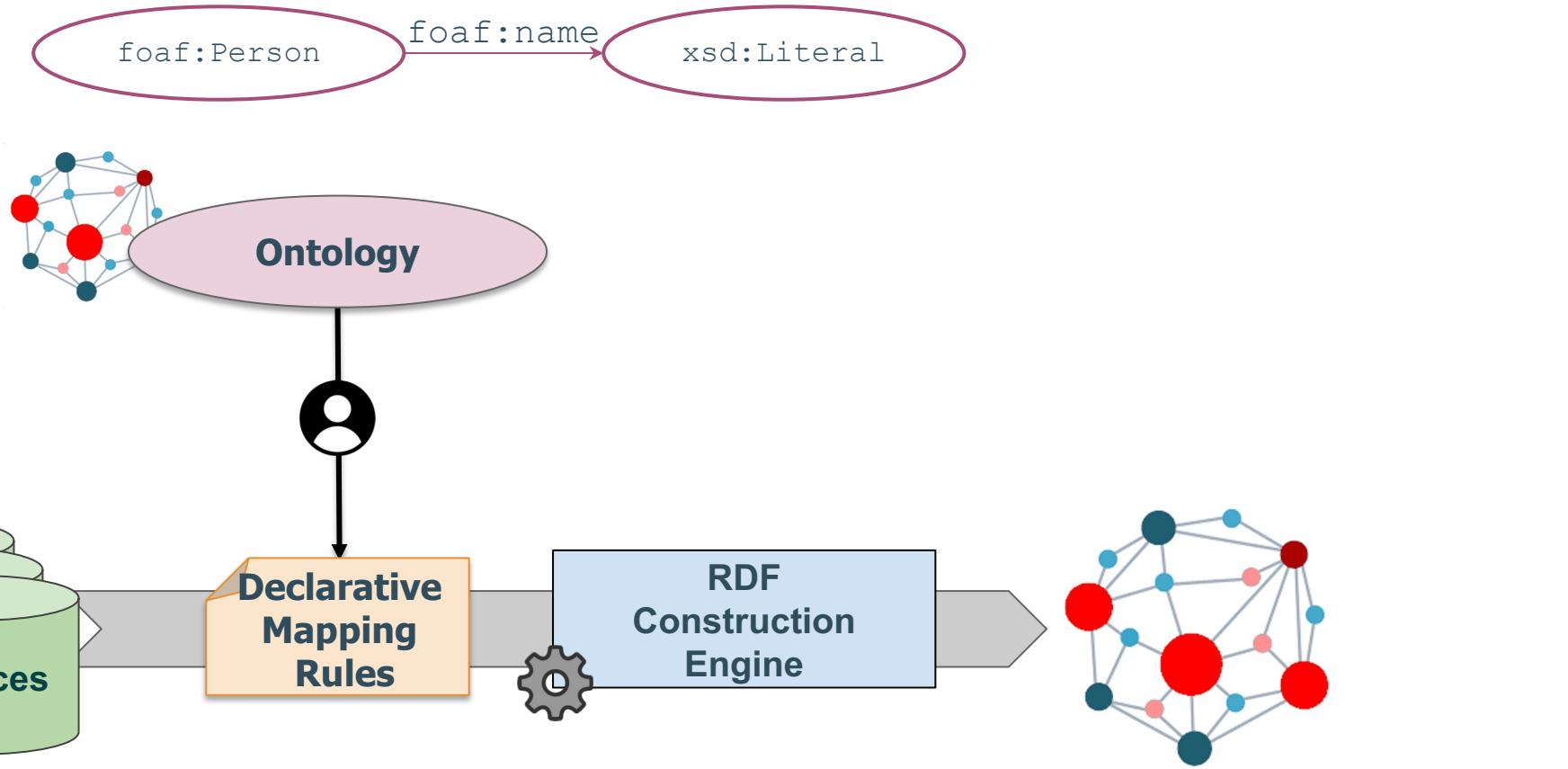
Declarative Knowledge Graph Construction

rank	name	mark
1	Anzhelika Sidorova	4.95
2	Sandi Morris	4.90
3	Katerina Stefanidi	4.85
4	Holly Bradshaw	4.80
5	Alysha Newman	4.80
6	Angelica Bengtsson	4.80



Declarative Knowledge Graph Construction

rank	name	mark
1	Anzhelika Sidorova	4.95
2	Sandi Morris	4.90
3	Katerina Stefanidi	4.85
4	Holly Bradshaw	4.80
5	Alysha Newman	4.80
6	Angelica Bengtsson	4.80



<http://ex.com/Anzhelika%20Sidorova> rdf:type foaf:Person.
<http://ex.com/Anzhelika%20Sidorova> foaf:name "Anzhelika Sidorova".
<http://ex.com/Sandi%20Morris> rdf:type foaf:Person.
<http://ex.com/Sandi%20Morris> foaf:name "Sandi Morris".

KU LEUVEN

What did I do?

Mapping Languages & RML

Mapping rules & User support

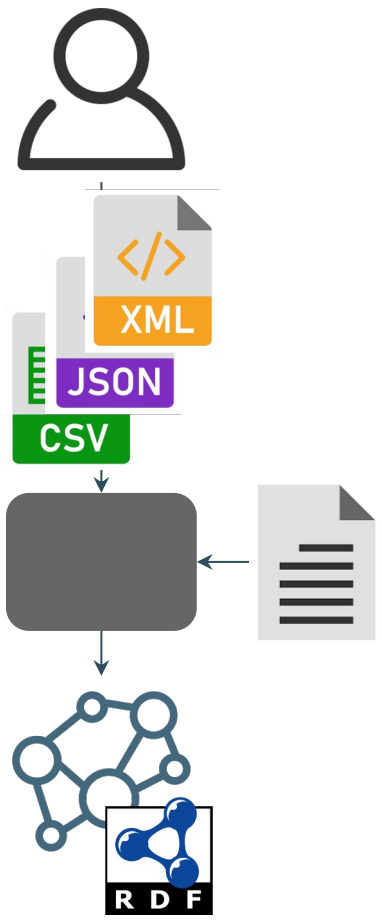
Mapping rules & RDF graphs validation

Mapping rules & SHACL shapes

Implementations & benchmarks

Declarative Mapping Languages

declarative



query languages
SPARQL syntax

XSPARQL (2008)
RML (2013)
xR2RML (2015)
SPARQL-Generate (2017)
D2RML (2018)
YARRRML (2018)
D-REPR (2019)
ShExML (2020)
(Facade-X) SPARQL-Anything (2021)

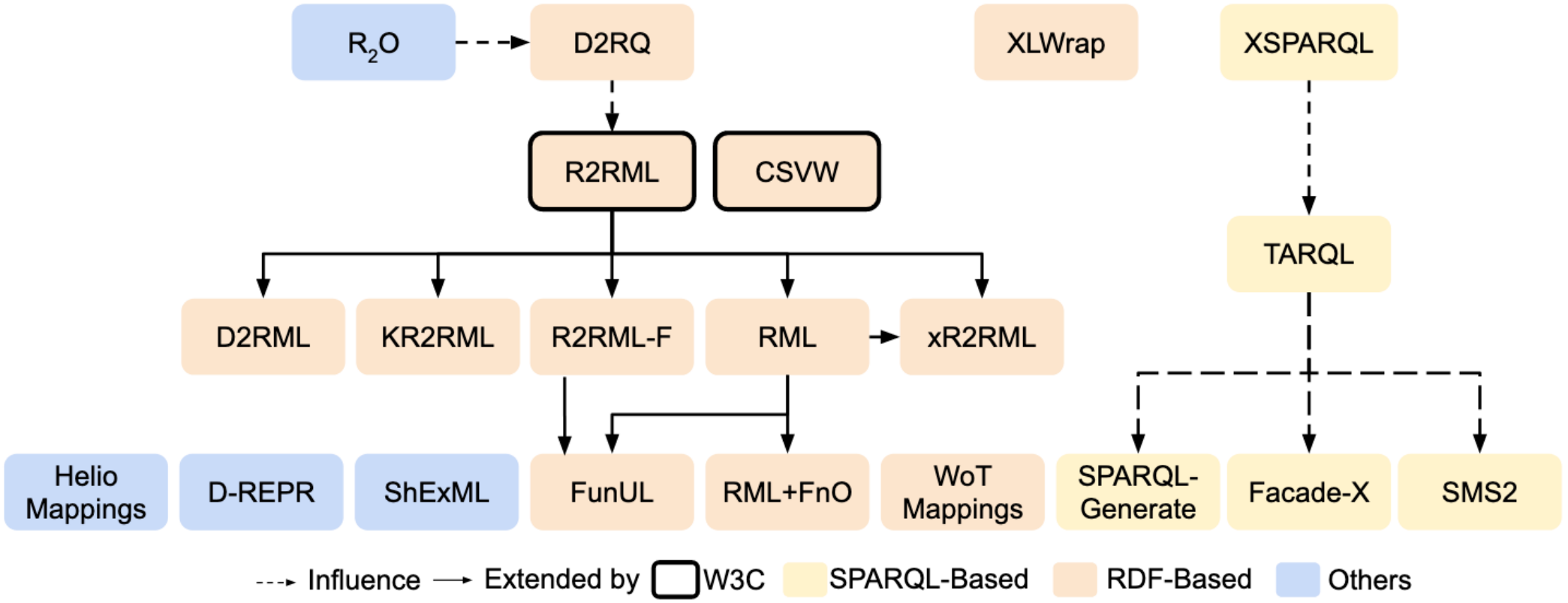
dedicated languages
R2RML syntax

data-serialisation languages
YAML syntax

constraint languages
ShEx syntax

Declarative RDF graph generation from heterogeneous (semi-)structured data: a Systematic Literature Review.
D. Van Assche, T. Delva, G. Haesendonck, P. Heyvaert, B. De Meester, A. Dimou. 2022. JWS
doi.org/10.1016/j.websem.2022.100753

Relationships of Declarative Mapping Languages



cross-mapping-language alignments

YARRRML-to-RML

https://github.com/RMLio/yarrrml-parser

https://github.com/oeg-upm/yarrrml-translator

RML-to-SPARQL-Generate

https://github.com/sparql-generate/rml-to-sparql-generate

ShExML-to-RML

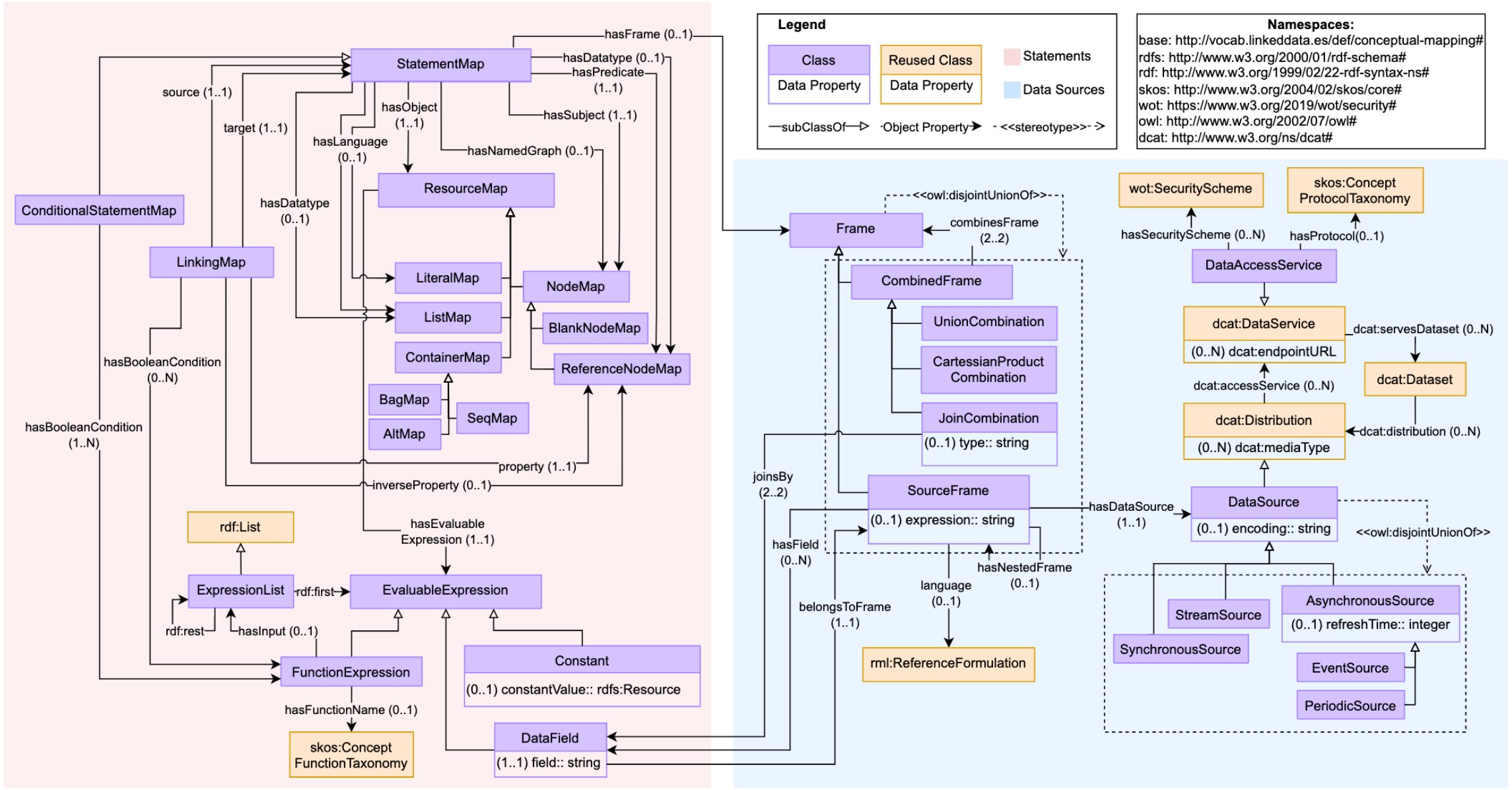
https://github.com/herminiogg/ShExML

RML-to-SPARQL-Anything

https://github.com/DylanVanAssche/rml2sparqlanything

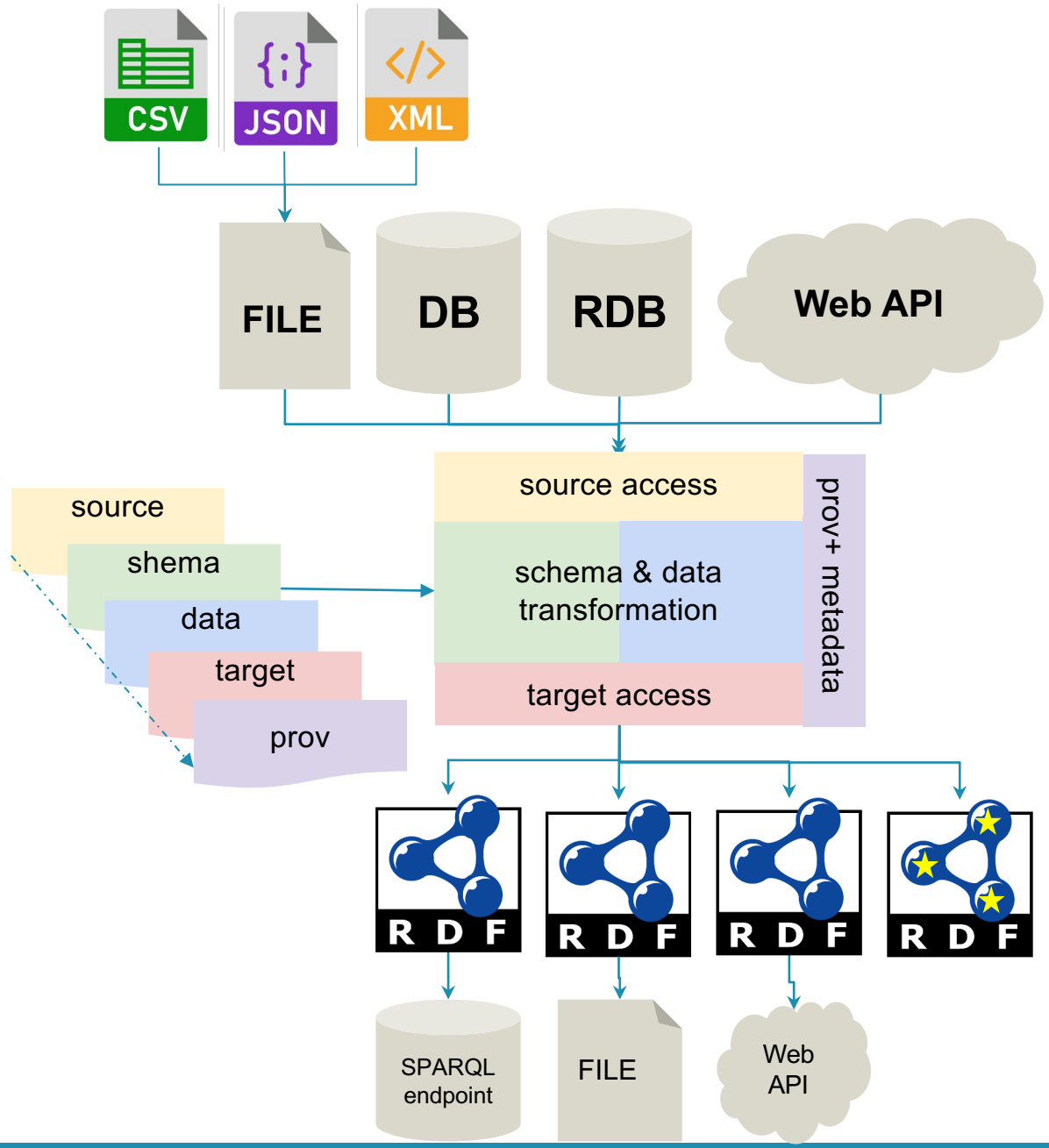
D2RQL-to-RML

https://github.com/RMLio/D2RQ_to_R2RML



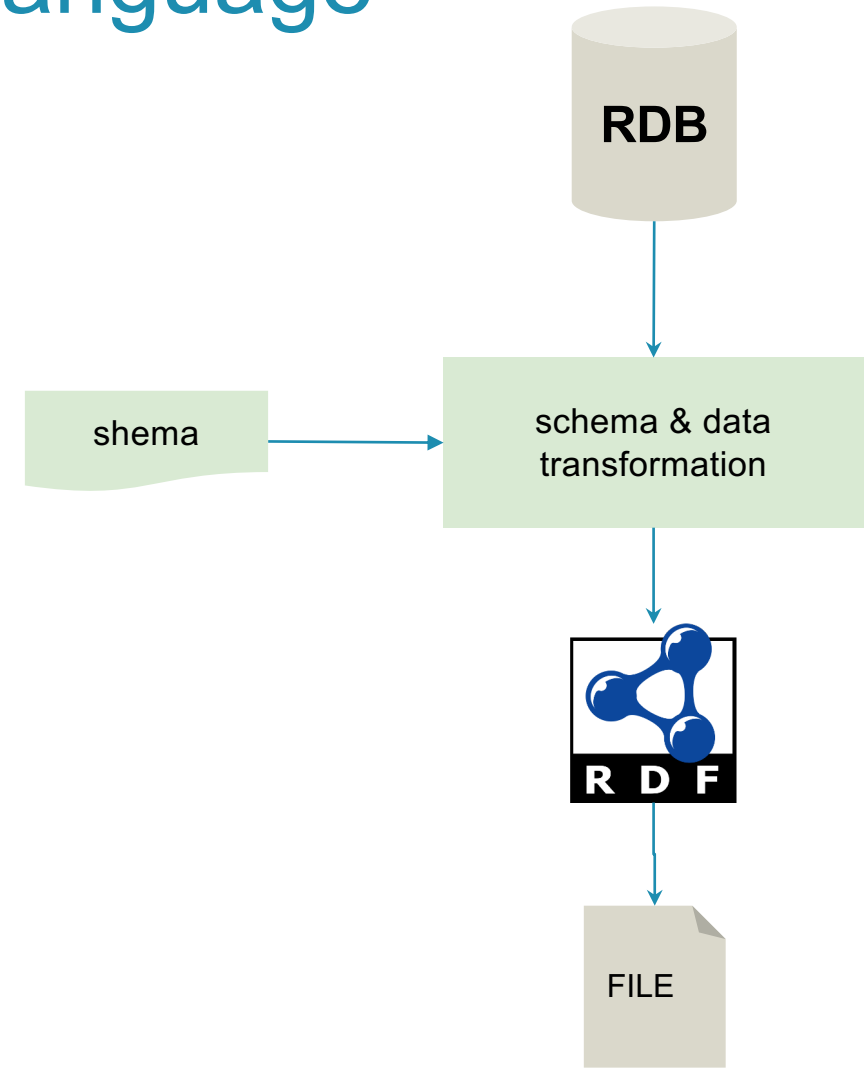
An Ontological Approach for Representing Declarative Mapping Languages. Ana Iglesias-Molina et al. (2022) SWJ /www.semantic-web-journal.net/content/ontological-approach-representing-declarative-mapping-languages

RDF Mapping Language (RML)



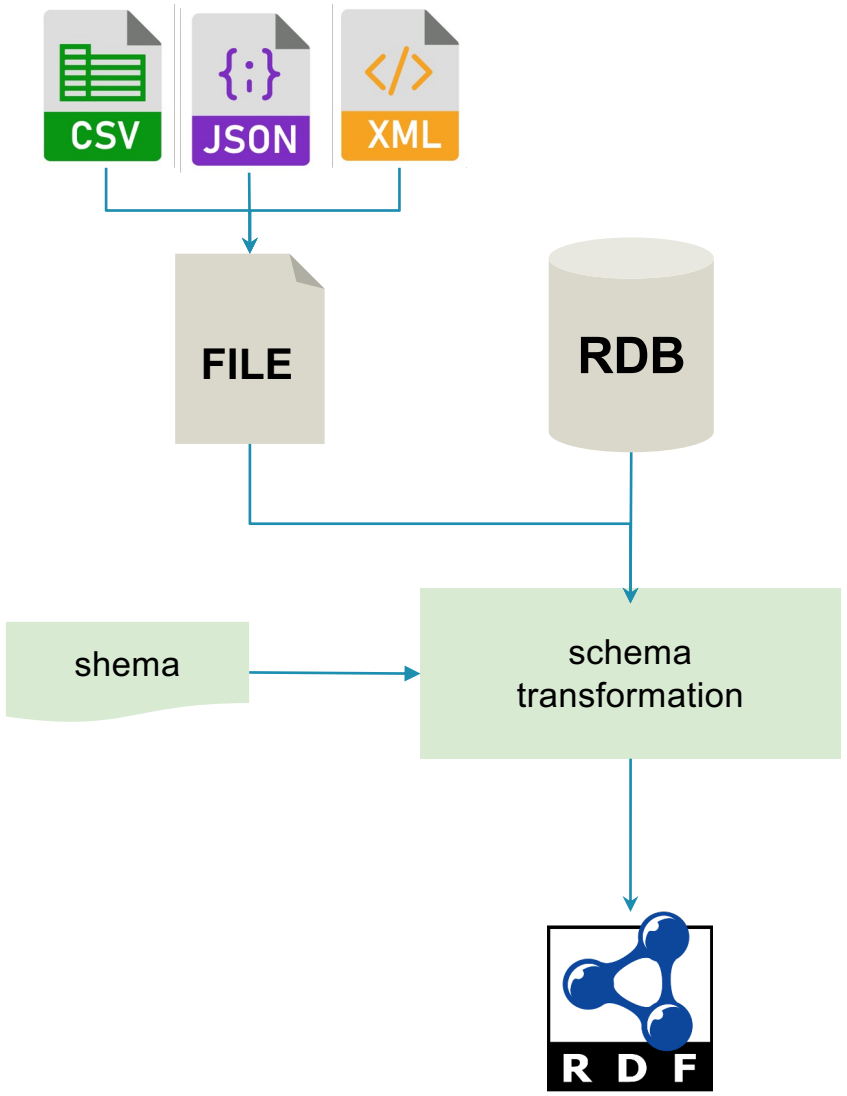
RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. A. Dimou et al. LDOW 2014
ceur-ws.org/Vol-1184/ldow2014_paper_01.pdf

Relational to RDF Mapping Language (R2RML)



https://www.w3.org/TR/r2rml/

RDF Mapping Language (RML)



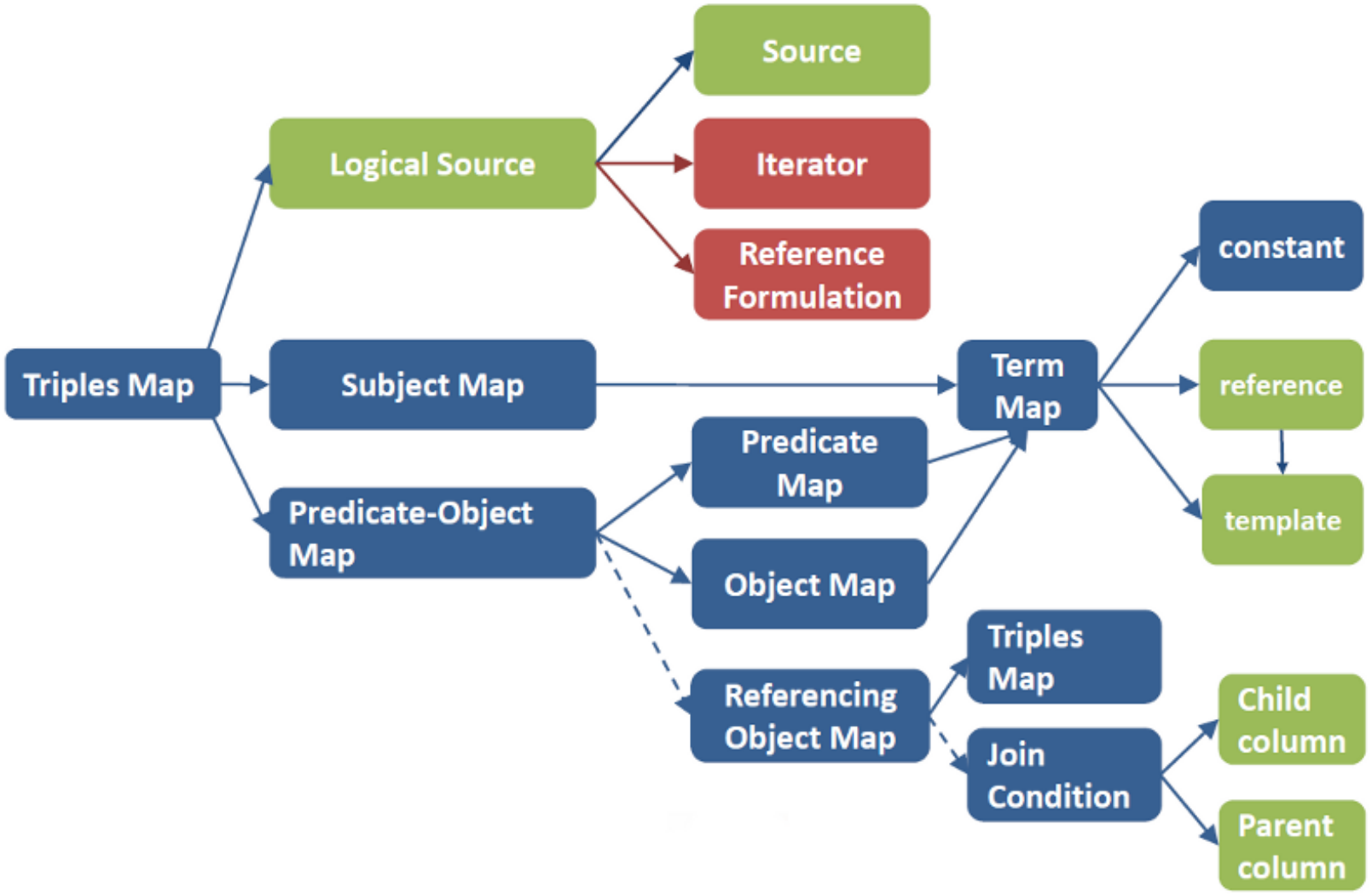
rank	name	mark
1	Anzhelika Sidorova	4.95
2	Sandi Morris	4.90
3	Katerina Stefanidi	4.85
4	Holly Bradshaw	4.80
5	Alysha Newman	4.80
6	Angelica Bengtsson	4.80


```
ex:Anzhelika%20Sidorova rdf:type foaf:Person.
ex:Anzhelika%20Sidorova foaf:name "Anzhelika Sidorova".
ex:Sandi%20Morris rdf:type foaf:Person.
ex:Sandi%20Morris foaf:name "Sandi Morris".
```

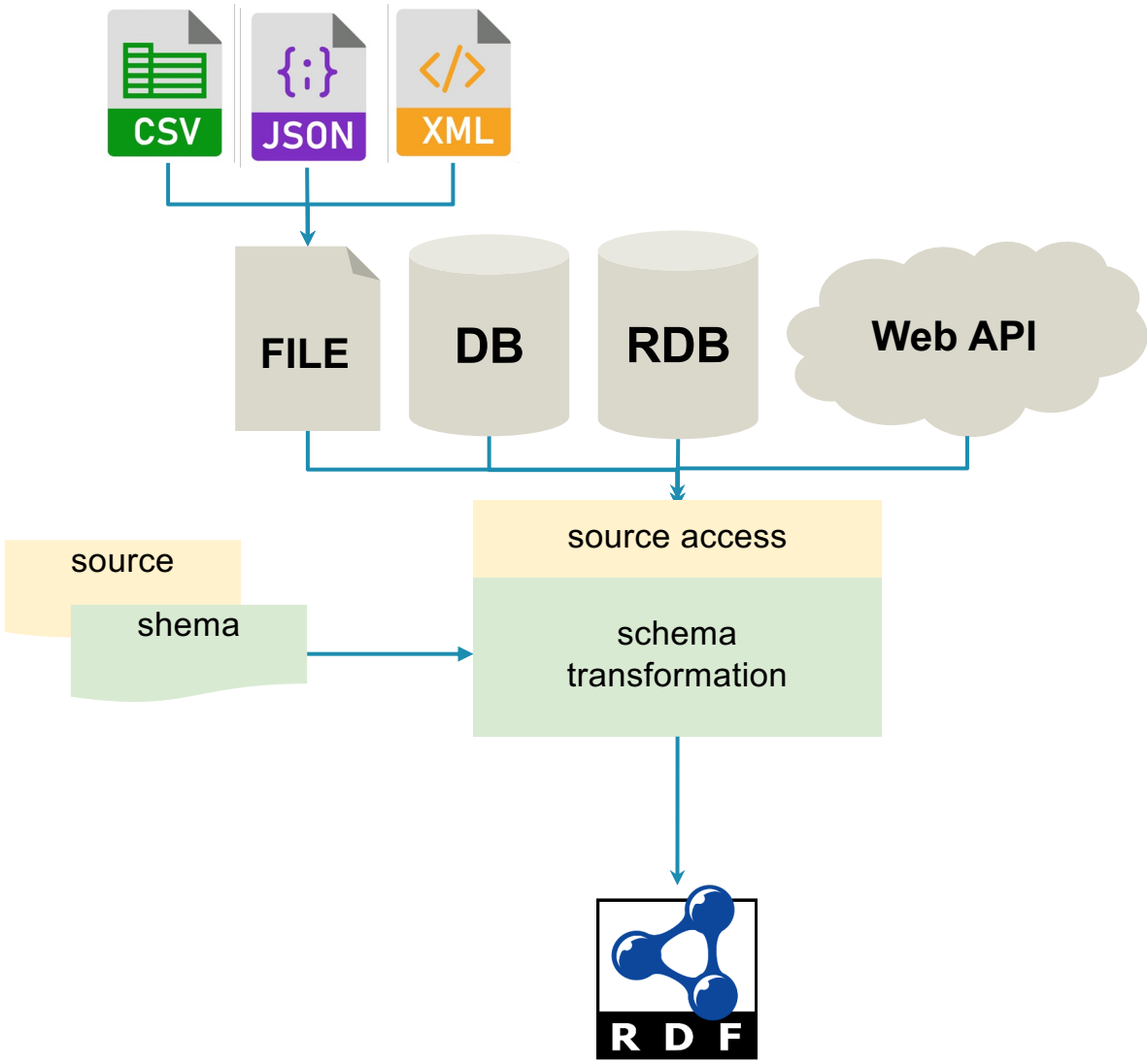
RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. A. Dimou et al. LDOW 2014
ceur-ws.org/Vol-1184/ldow2014_paper_01.pdf

KU LEUVEN

R2RML Vs RML


Source
Logical Source
Iterator
Reference Formulation
constant
Triples Map
Subject Map
Term Map
reference
Predicate Map
template
Predicate-Object Map
Object Map
Triples Map
Referencing Object Map
Child column
Join Condition
Parent column

RML: source description



Local File(s)
Database connectivity
- **D2RQ** www.d2rq.org/d2rq-language

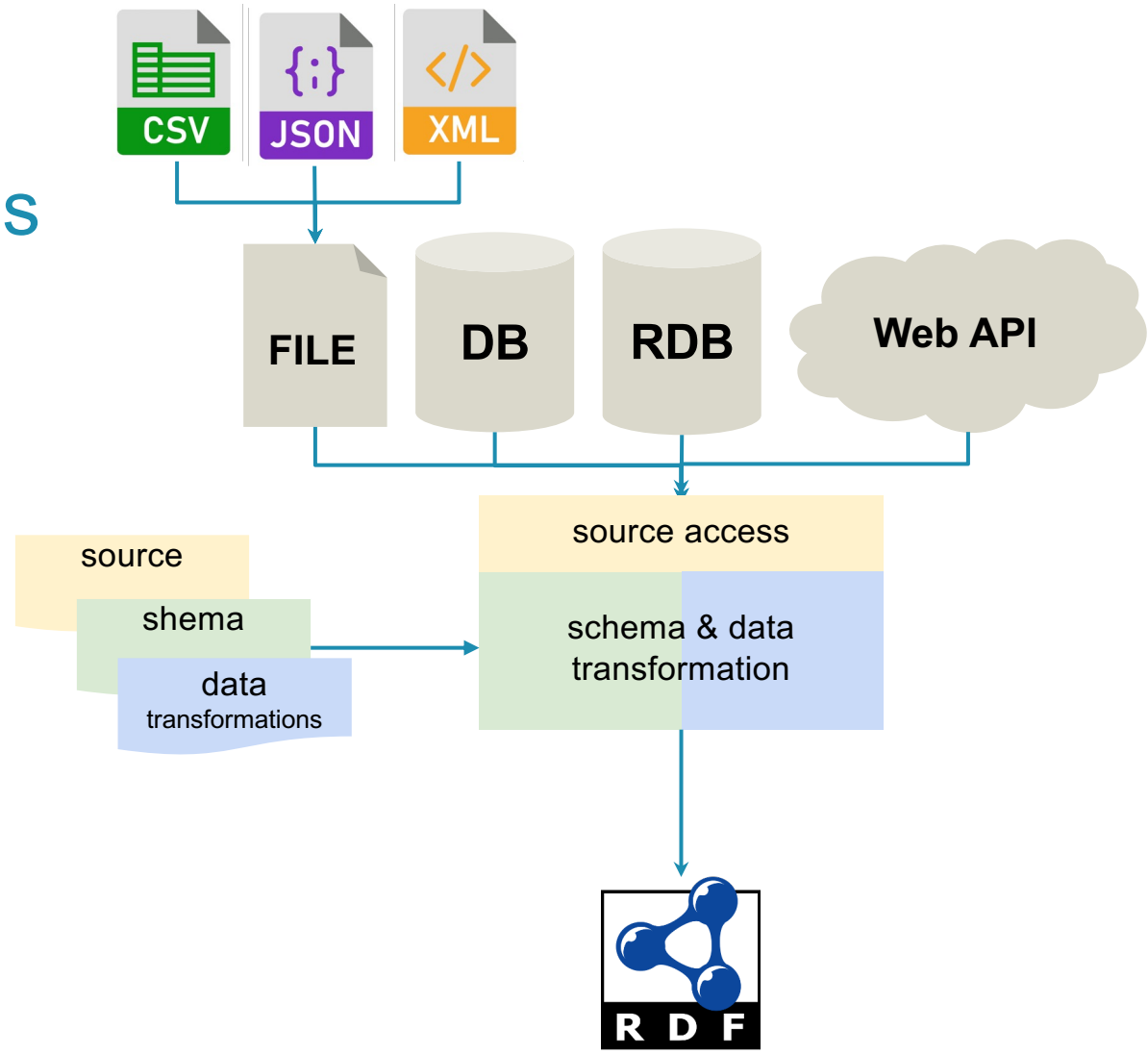
Web source(s) (Web API/service)
- **DCAT** www.w3.org/TR/vocab-dcat-2/
- **CSVW** www.w3.org/ns/csvw
- **Hydra** www.hydra-cg.com/spec/latest/core/
- **VOiD** www.w3.org/TR/void/

RDF source(s)
- **VOiD** (endpoint) www.w3.org/TR/void/
- **SPARQL-SD** www.w3.org/TR/sparql11-service-description/

RML+FnO: data transformations

https://fno.io/spec/
alternatives:
R2RML-F
FunUL

rank	name	mark
1	Anzhelika Sidorova	4.95
2	Sandi Morris	4,90
3	Katerina Stefanidi	4.85
4	Holly Bradshaw	4.80
5	Alysha Newman	4.80
6	Angelica Bengtsson	4.80

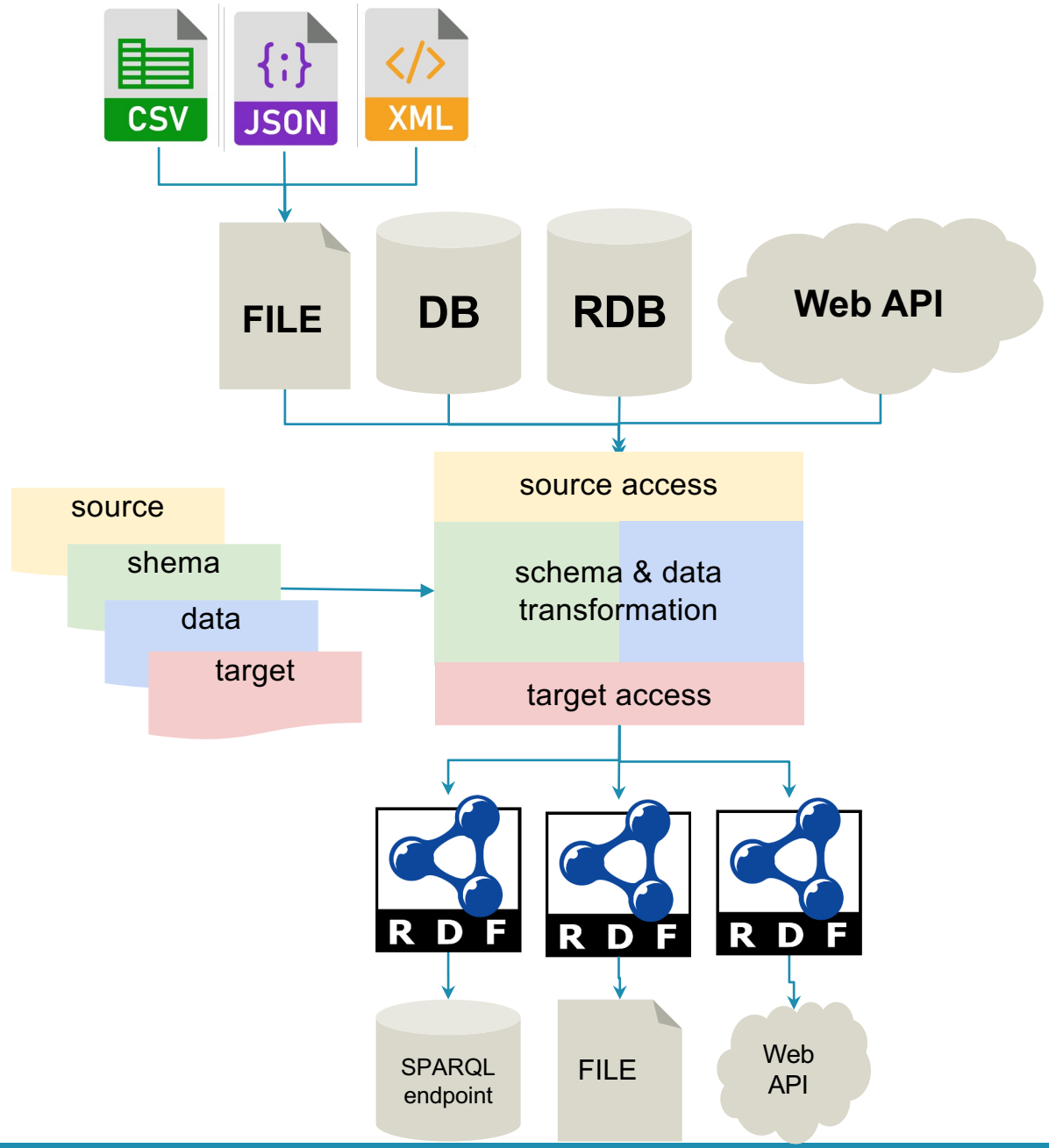



```
ex:Anzhelika%20Sidorova rdf:type foaf:Person.
ex:Anzhelika%20Sidorova foaf:name ”ANZHELIKA SIDOROVA".
ex:Sandi%20Morris rdf:type foaf:Person.
ex:Sandi%20Morris foaf:name ”SANDI MORRIS".
```


Implementation-independent function reuse. B.De Meester et al. (2020) FGCS. doi.org/10.1016/j.future.2019.10.006
Declarative data transformations for Linked Data generation: the case of DBpedia. B.De Meester et al. ESWC2017 doi.org/10.1007/978-3-319-58451-5_3
An Ontology to Semantically Declare and Describe Functions. B.De Meester et al. ESWC2016 P&D doi.org/10.1007/978-3-319-47602-5_10

KU LEUVEN

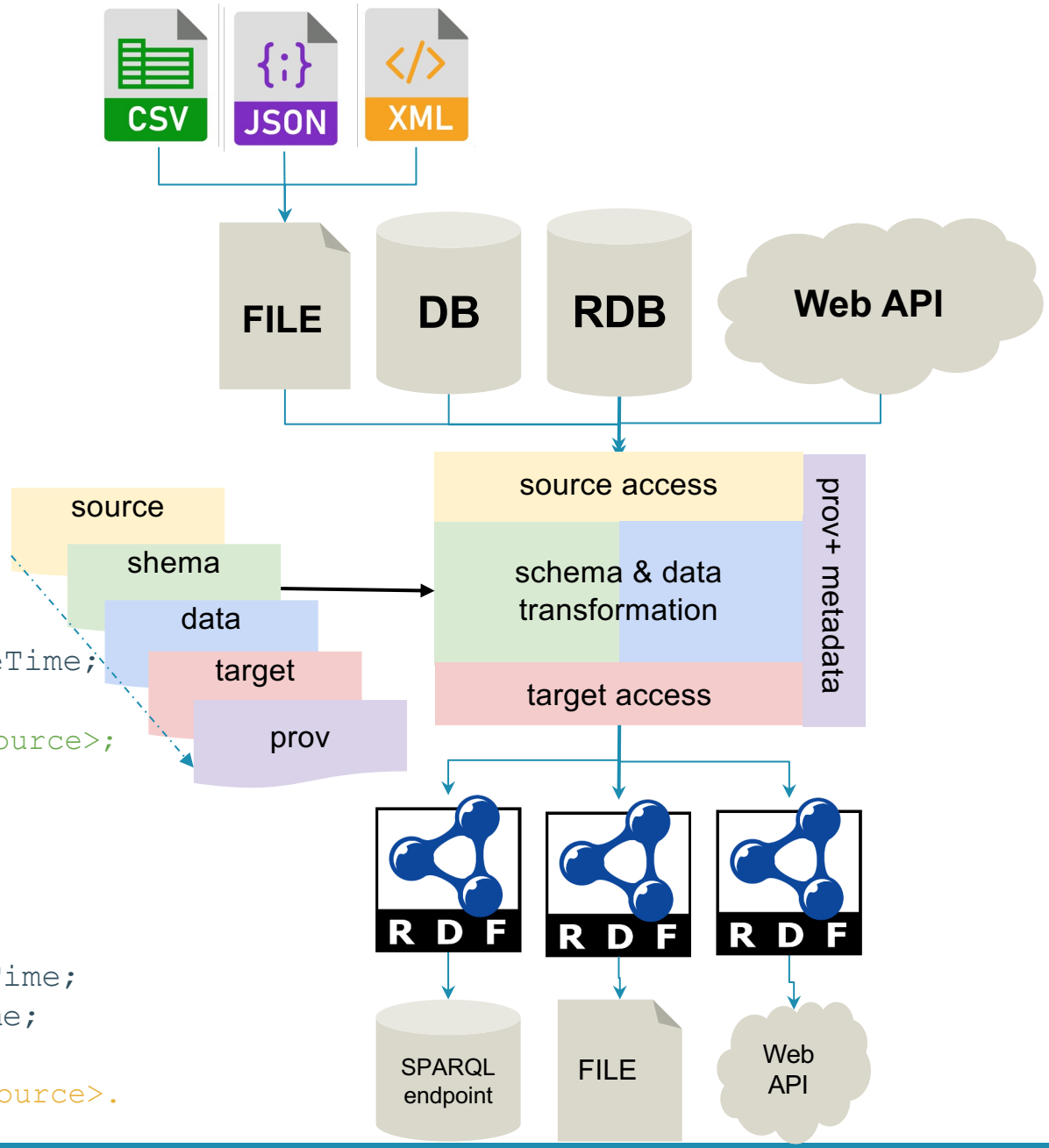
RML: target description



Leveraging Web of Things W3C recommendations for knowledge graphs generation. D.Van Assche et al. ICWE2021
doi.org/10.1007/978-3-030-74296-6_26

RML: provenance & metadata

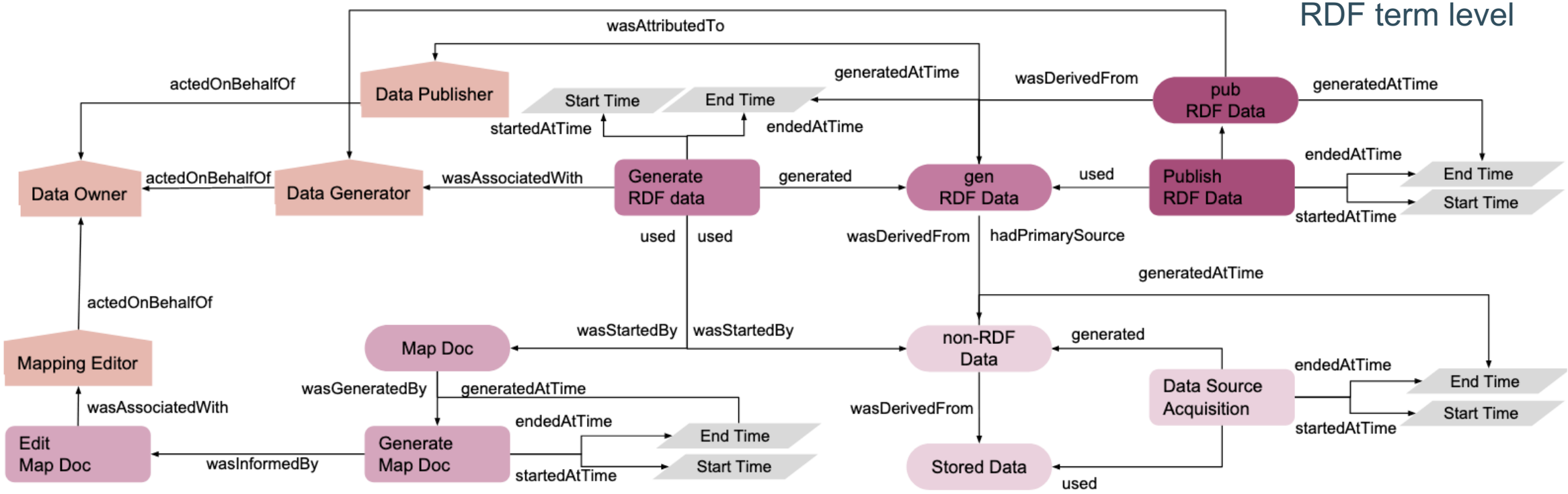
```
ex:Anzhelika%20Sidorova rdf:type foaf:Person.
ex:Anzhelika%20Sidorova foaf:name "Anzhelika Sidorova".
ex:Sandi%20Morris rdf:type foaf:Person.
ex:Sandi%20Morris foaf:name "Sandi Morris".

<#RDF_Dataset> a prov:Entity, void:Dataset;
prov:generatedAtTime "2022-01-05T17:10:00Z"^^xsd:dateTime;
prov:wasGeneratedBy <#RDFdataset_Generation>;
prov:wasDerivedFrom <#DB_LogicalSource>,<#DCAT_LogicalSource>;
prov:wasAssociatedWith <#RMLProcessor>;
prov:wasAttributedTo <http://natadimou.com>.

<#RDFdataset_Generation> a prov:Activity;
 prov:generated <#RDF_Dataset>;
 prov:startedAtTime "2022-01-05T17:00:00Z"^^xsd:dateTime;
 prov:endedAtTime "2022-01-05T17:10:00Z"^^xsd:dateTime;
 prov:wasInformedBy <#MapDoc_Generation>;
 prov:used <#MapDoc>,<#DB_LogicalSource>,<#DCAT_LogicalSource>.
```



Automated Metadata Generation for Linked Data Generation and Publishing Workflows. A. Dimou et al. LDOW2016 ceur-ws.org/Vol-1593/article-04.pdf
Detailed Provenance Capture of Data Processing. B. De Meester, A. Dimou, R. Verborgh, E. Mannens. SemSci2017 ceur-ws.org/Vol-1931/paper-05.pdf

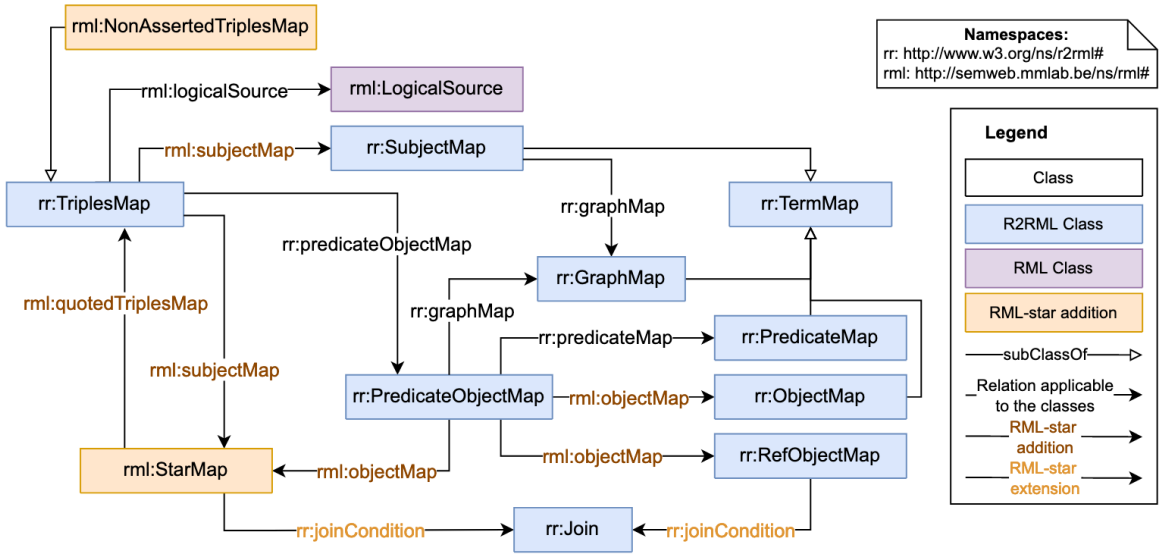
RML: provenance

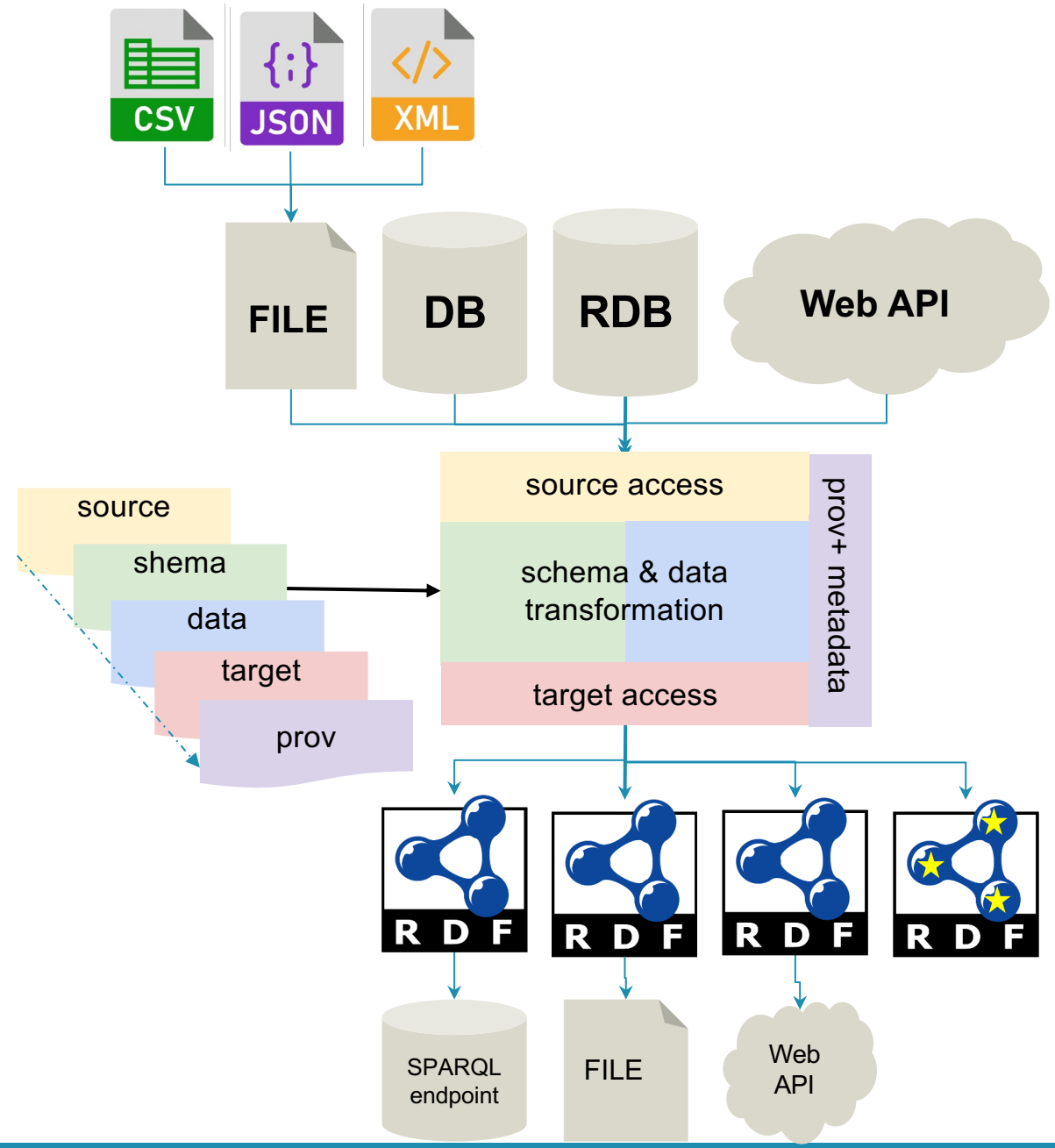
Metadata details levels:
- dataset level
- named-graph level
- subgraph level
- RDF triple level
- RDF term level



Automated Metadata Generation for Linked Data Generation and Publishing Workflows. A. Dimou et al. LDOW2016 ceur-ws.org/Vol-1593/article-04.pdf
Detailed Provenance Capture of Data Processing. B. De Meester, A. Dimou, R. Verborgh, E. Mannens. SemSci2017 ceur-ws.org/Vol-1931/paper-05.pdf

RML-star

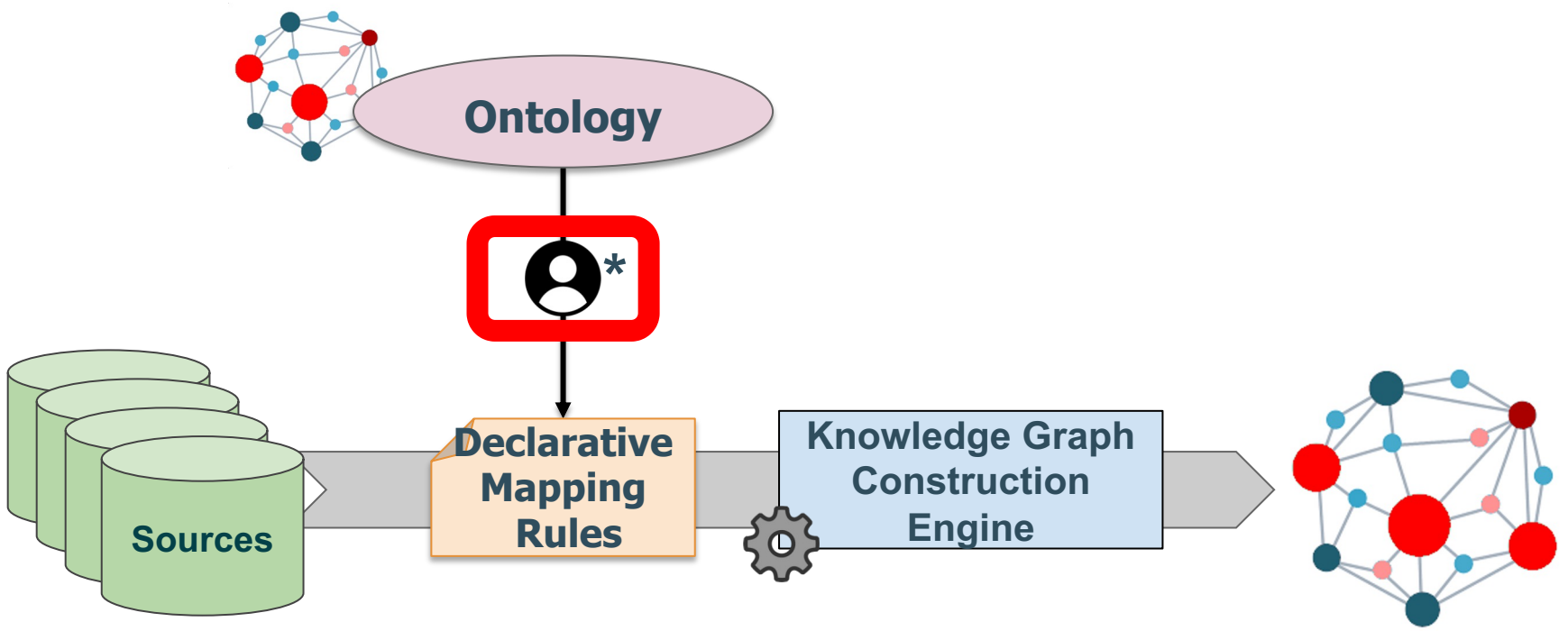




RML-star: A Declarative Mapping Language for RDF-star Generation. Delva et al. ceur-ws.org/Vol-2980/paper374.pdf
Morph-KGC^star: Declarative Generation of RDF-star Datasets from Heterogeneous Data. J. Arenas-Guerrero et al. www.semantic-web-journal.net/content/morph-kgcstar-declarative-generation-rdf-star-datasets-heterogeneous-data

KU LEUVEN

Declarative Knowledge Graph Construction



*** On average it takes 6 Person-Month to create a knowledge graph**

Declarative Description of Knowledge Graphs Construction Automation: Status & Challenges. Chaves-Fraga & Dimou. KGCW2022
http://ceur-ws.org/Vol-3141/paper5.pdf

What did I do?

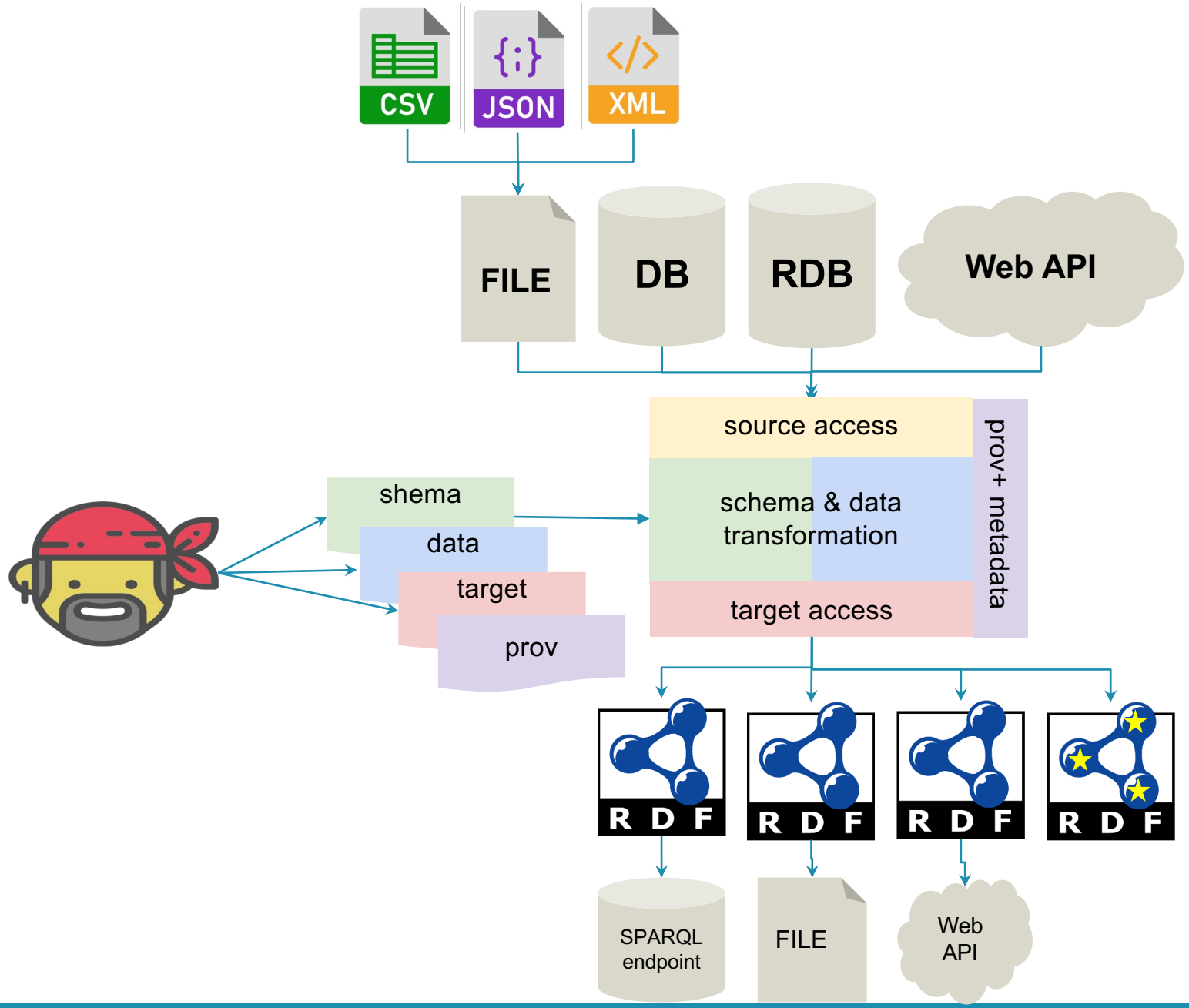
Mapping Languages & RML

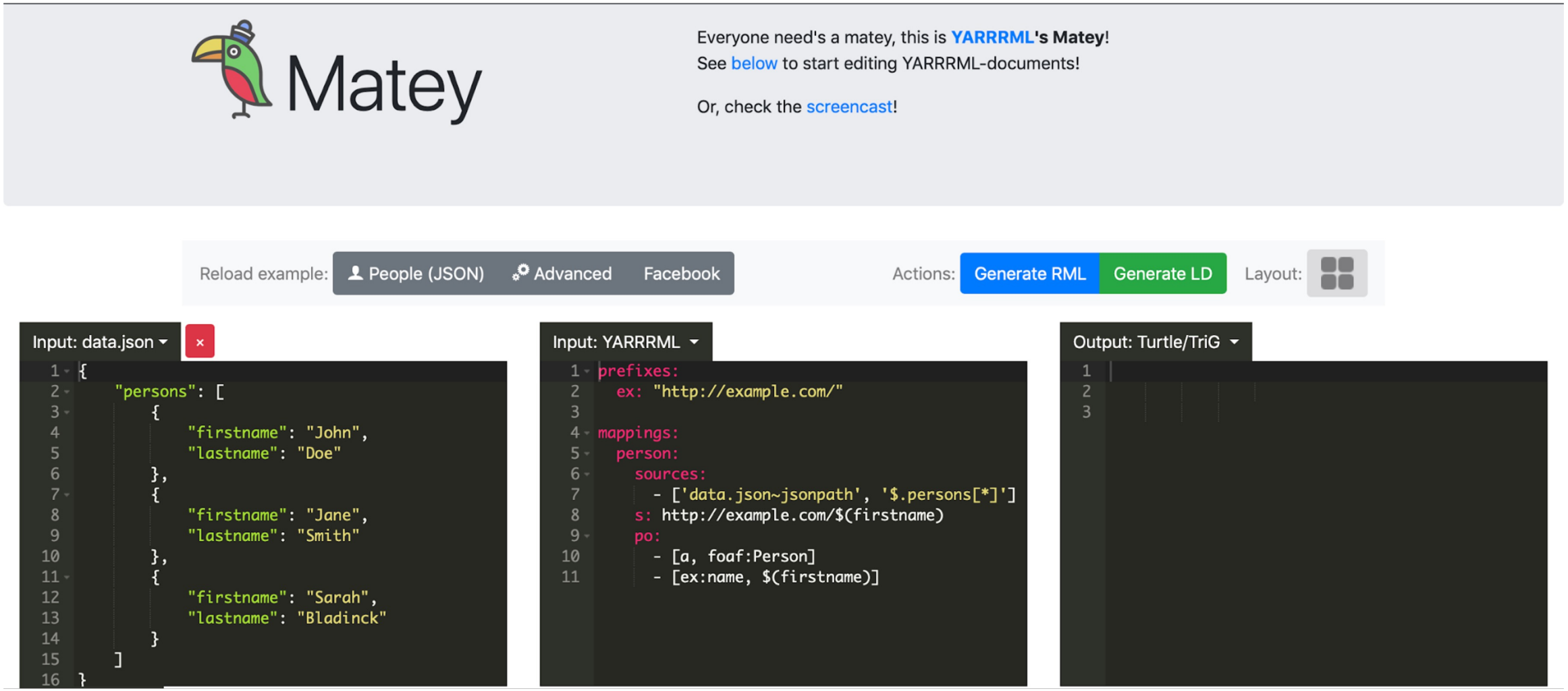
Mapping rules & User support

Mapping rules & RDF graphs validation

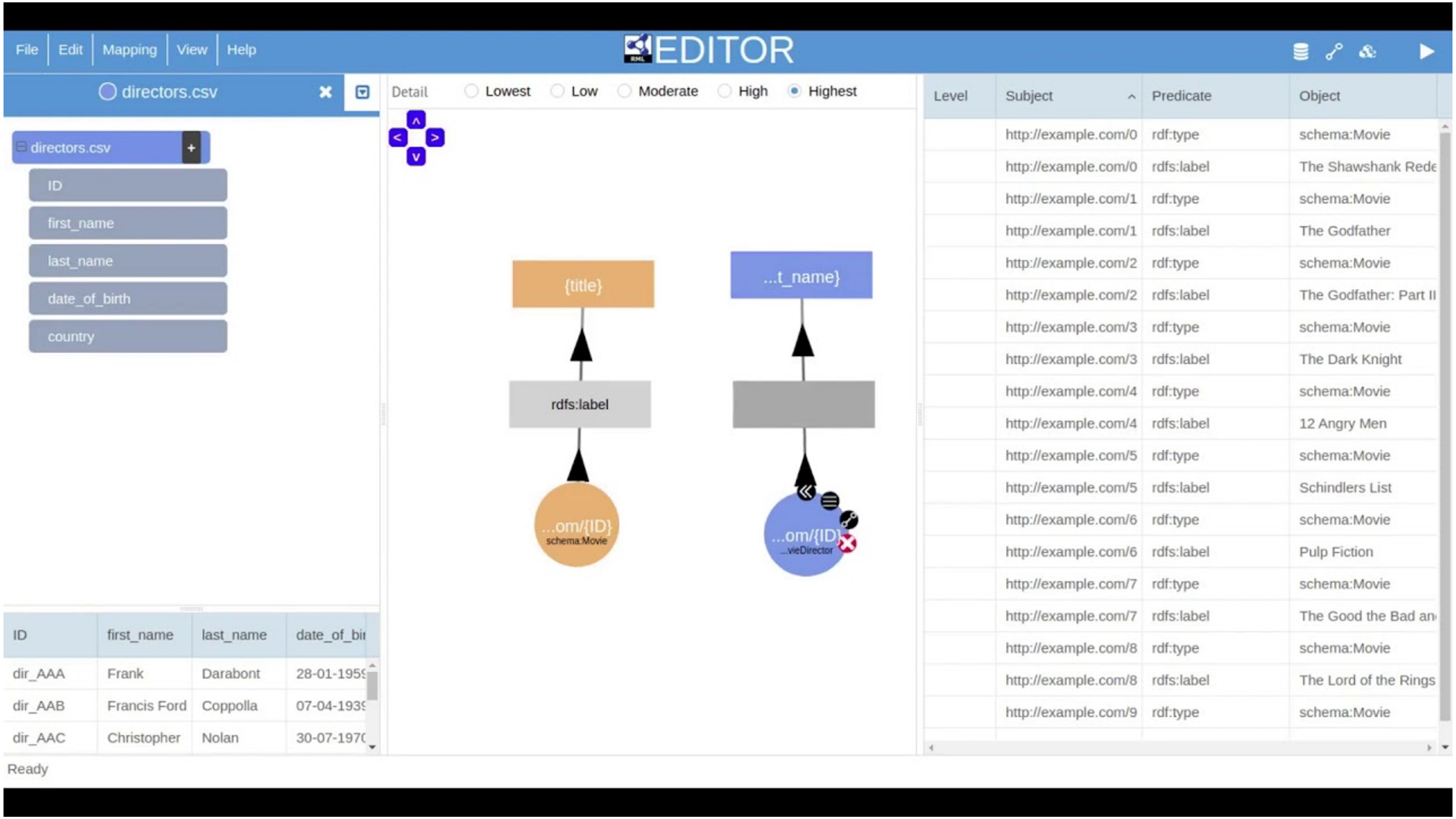
Mapping rules & SHACL shapes

Implementations & benchmarks


CSV
JSON
XML
FILE
DB
RDB
Web API
source access
schema & data transformation
target access
prov+ metadata
shema
data
target
prov
R D F
R D F
R D F
R D F
SPARQL endpoint
FILE
Web API



Declarative Rules for Linked Data Generation at your Fingertips! P. Heyvaert, B. De Meester, A. Dimou, R. Verborgh ESWC 2018
doi.org/10.1007/978-3-319-98192-5_40



Specification and implementation of mapping rule visualization and editing: MapVOWL and the RMLEditor.
P. Heyvaert & A. Dimou, et al. JWS 2018 doi.org/10.1016/j.websem.2017.12.003

KU LEUVEN

Editors

Matey
https://github.com/rmlio/matey

RMLEditor
https://app.rml.io/rmleditor/, https://rml.io/tools/rmleditor/, https://github.com/RMLio/rmleditor-ce

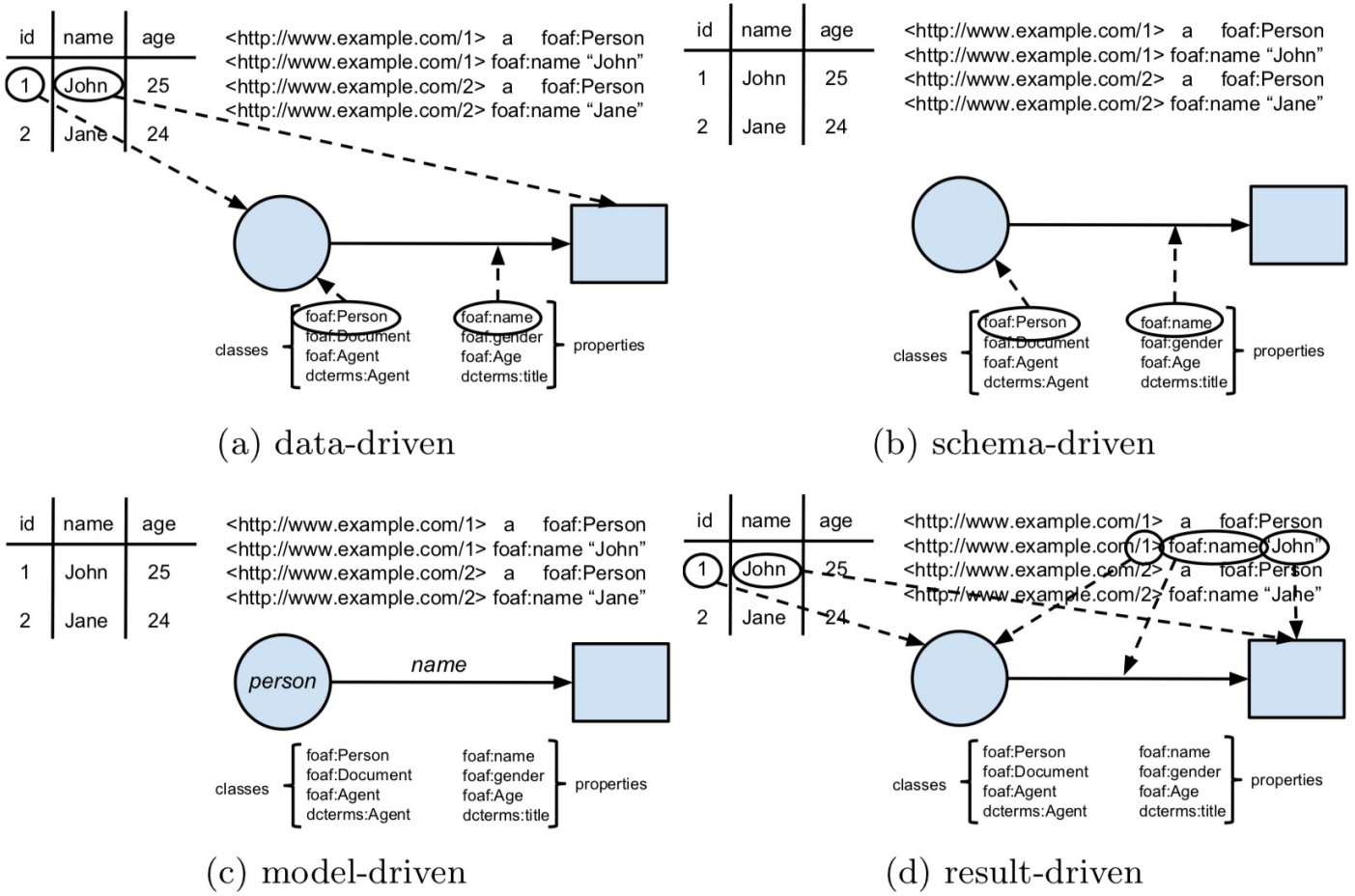
Mapeathor
https://morph.oeg.fi.upm.es/tool/mapeathor, https://github.com/oeg-upm/morph-website

Map-On
http://semanco-tools.eu/map-on, https://github.com/arc-lasalle/Map-On

RMLx Visual Editor
http://pebbie.org/mashup/rml

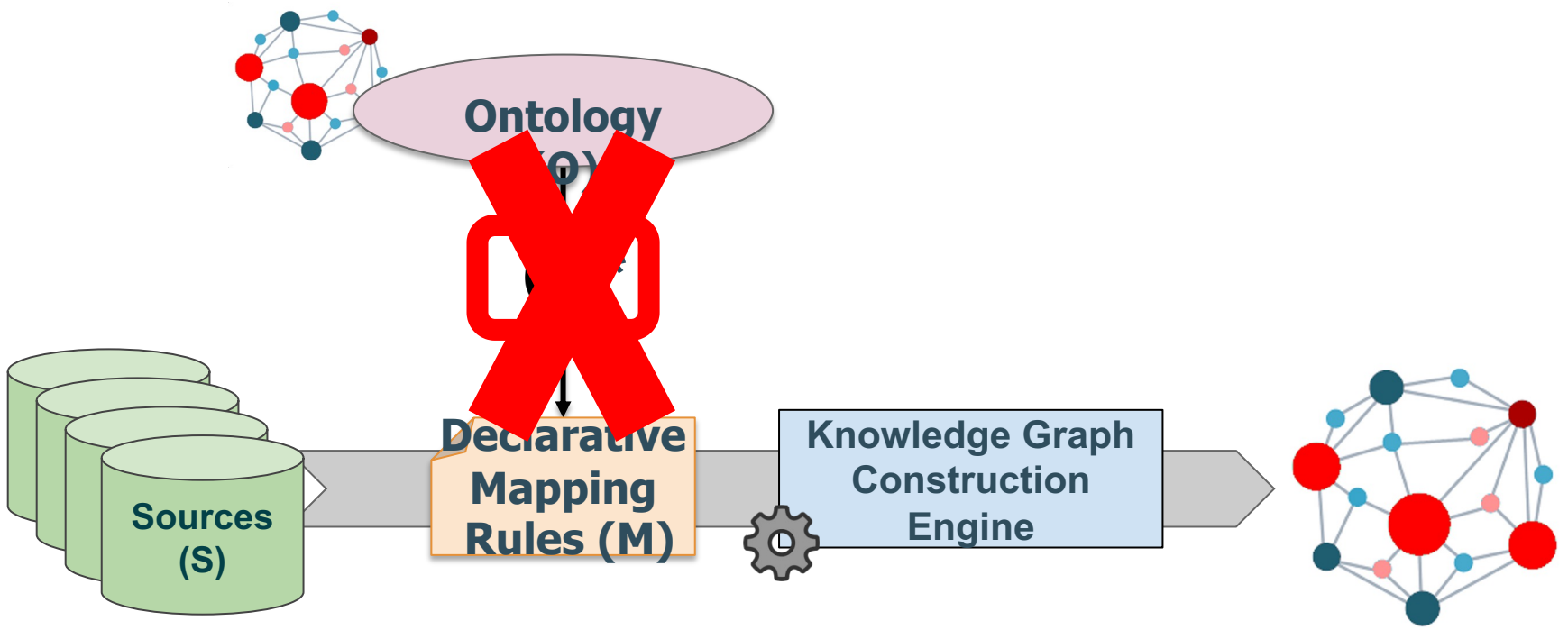
Specification and implementation of mapping rule visualization and editing: MapVOWL and the RMLEditor.
P. Heyvaert & A. Dimou, et al. (2018) JWS doi.org/10.1016/j.websem.2017.12.003

Approaches for generating mapping rules



Towards Approaches for Generating RDF Mapping Definitions P. Heyvaert et al. ISWC2015 P&D
ceur-ws.org/Vol-1486/paper_70.pdf

Declarative Knowledge Graph Construction



*** On average it takes 6 Person-Month to create a knowledge graph**

Declarative Description of Knowledge Graphs Construction Automation: Status & Challenges. Chaves-Fraga & Dimou. KGCW2022
http://ceur-ws.org/Vol-3141/paper5.pdf

The million dollar question!

How can we automate the knowledge graph construction and reduce its time and effort?

KU LEUVEN

Direct Knowledge Graph Construction

wrapper systems

FDR2 (2004)

D2RQ (2004)

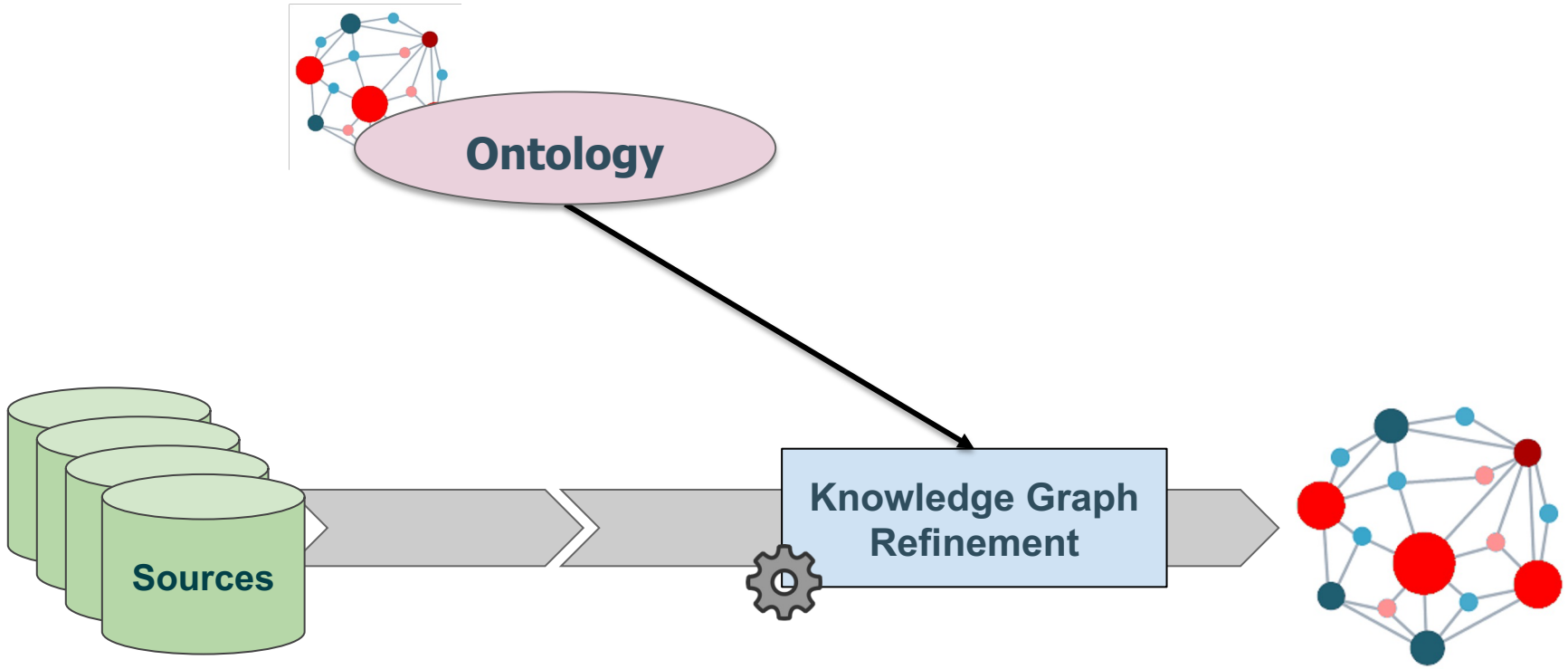
METAmorphoses (2004)

Relational.OWL (2005)

R_2O (2006)

RDB2Onto (2006)

D2OMapper (2006)



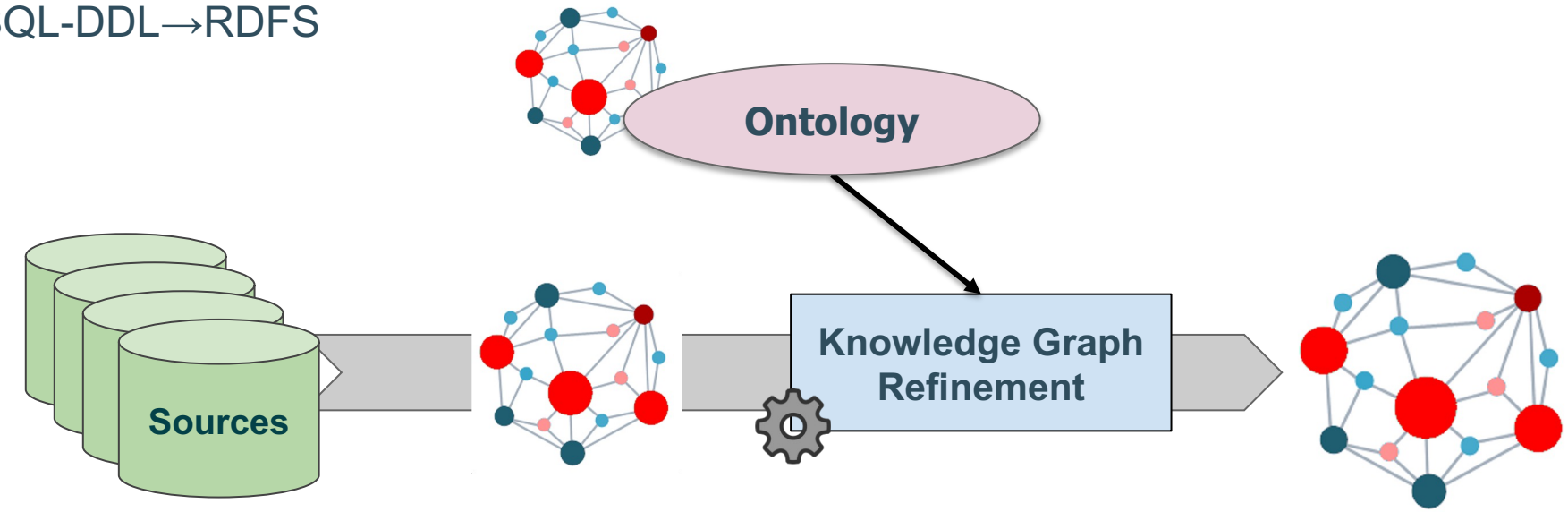
Direct Mapping SQL Databases to the Semantic Web: a Survey. Sequeda et al. The Knowledge Engineering Review (2011)

Direct Knowledge Graph Construction

direct mapping

Buccella et al. (2004)
Li et al. (2005)
Shen et al. (2006)
} SQL-DDL→OWL

Stojanovic et al. (2002)
Astrova et al. (2004)
} SQL-DDL→RDFS



Direct Mapping SQL Databases to the Semantic Web: a Survey. Sequeda et al. The Knowledge Engineering Review (2011)

Direct Knowledge Graph Construction

declarative automation

Mirror (2015)

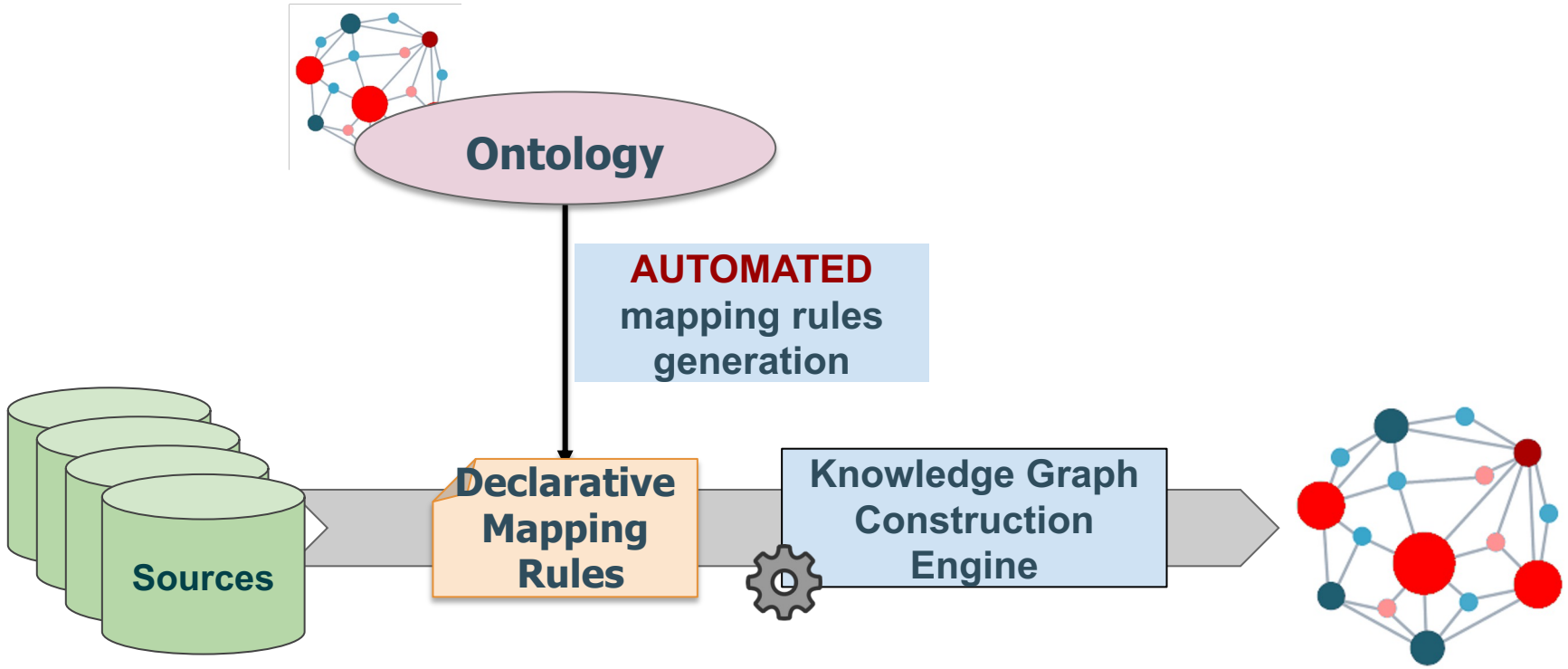
AutoMap4OBDA (2016)

BootOX (2015)

IncMap (2013)

COMA++ (2005)

Ontop (... - nowadays)



RODI benchmark → disappointing results

RODI: Benchmarking Relational-to-Ontology Mapping Generation Quality. Pinkel et al. SWJ 2016

Direct Knowledge Graph Construction

SemTab challenge

MTab (2019,2020,2021)

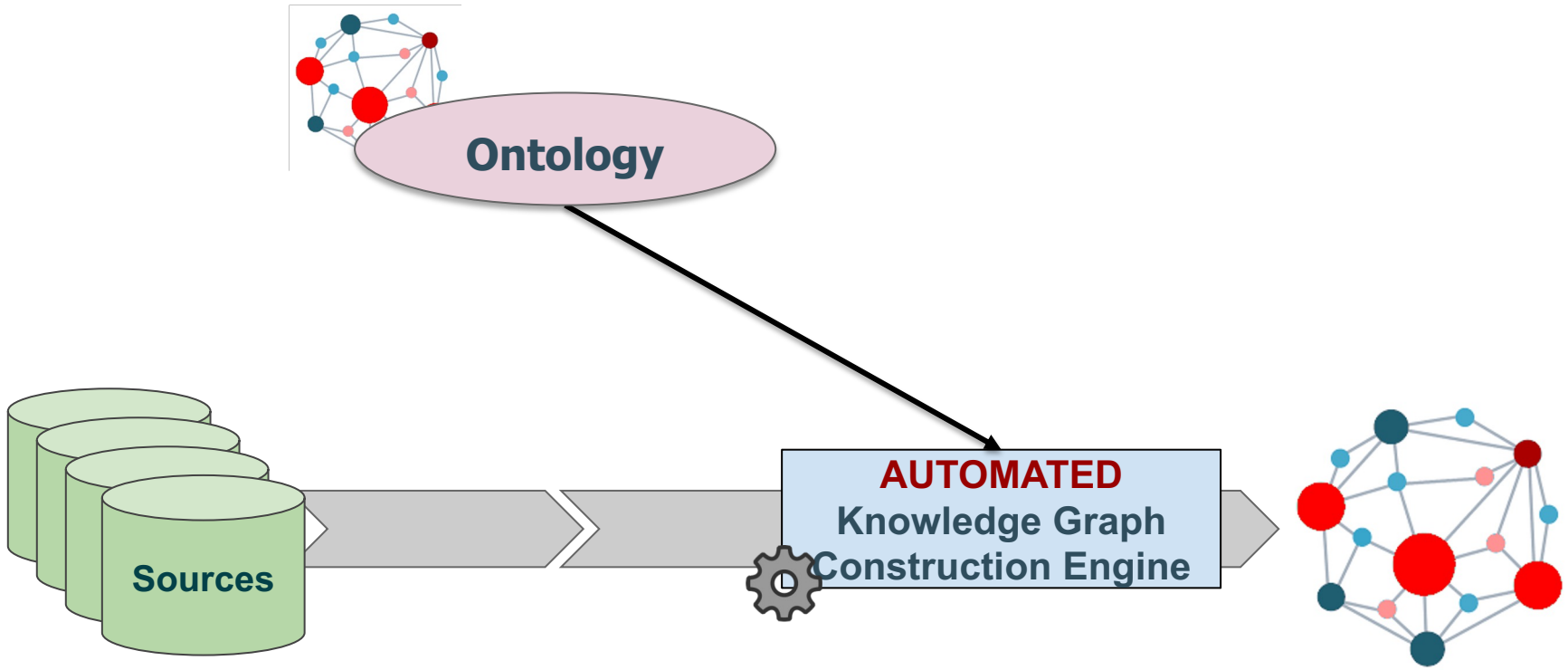
MAGIC (2020)

MantisTable (2021)

JenTab (2021)

DAGOBAH (2019,2020,2021)

....

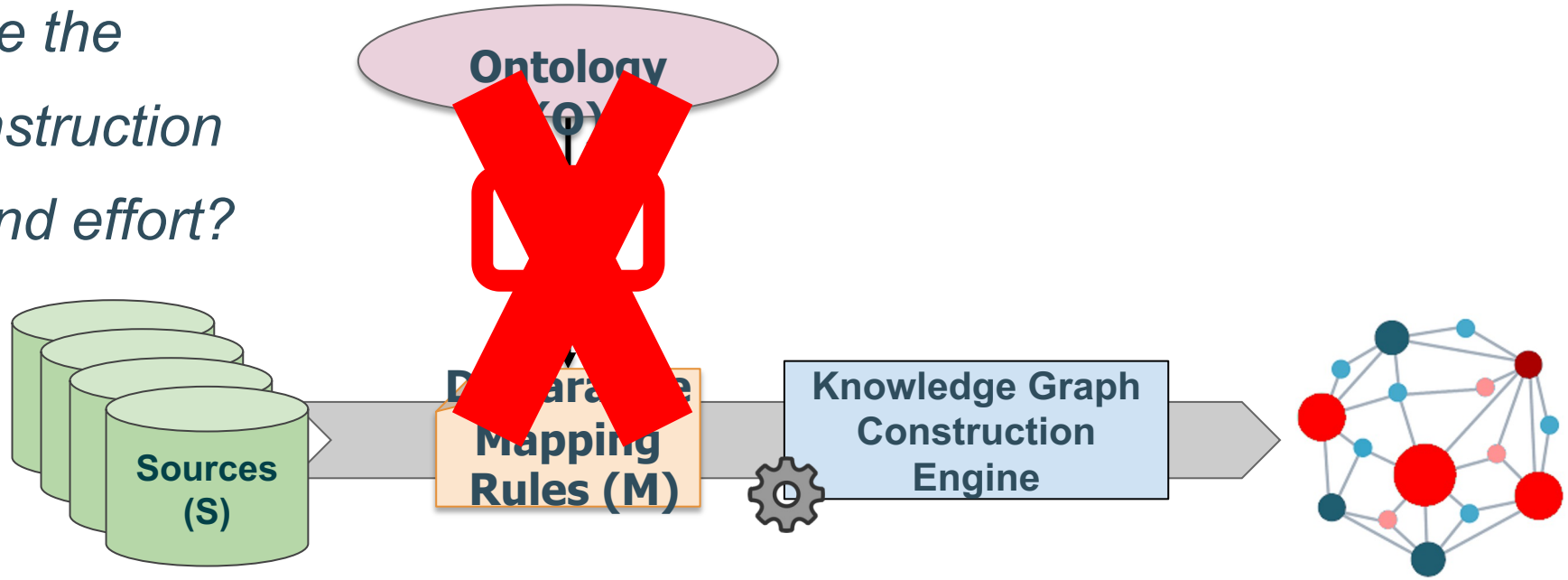


SemTab Challenge → overfit ⇒ lack of generalisation

https://www.cs.ox.ac.uk/isg/challenges/sem-tab/

The million dollar question!

How can we automate the knowledge graph construction and reduce its time and effort?



What did I do?

Mapping Languages & RML

Mapping rules & User support

Mapping rules validation

Mapping rules & SHACL shapes

Implementations & benchmarks

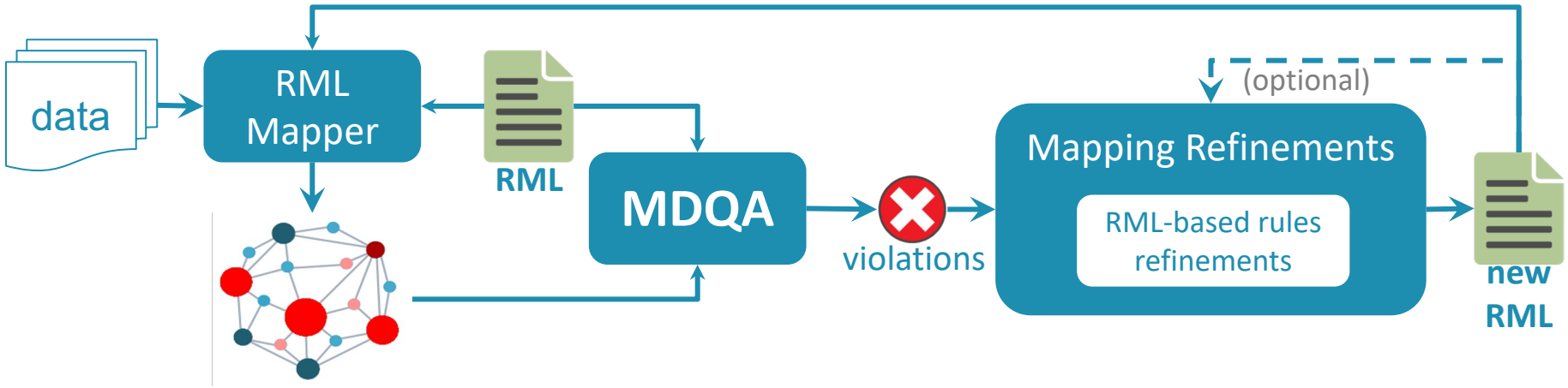
rank	name	mark
1	Anzhelika Sidorova	4.95
2	Sandi Morris	4.90
3	Katerina Stefanidi	4.85
4	Holly Bradshaw	4.80
5	Alysha Newman	4.80
6	Angelica Bengtsson	4.80

ex:Anzhelika%20Sidorova rdf:type ~~foaf:Person~~ bibo: Document .
ex:Anzhelika%20Sidorova foaf:name "Anzhelika Sidorova"^^xsd:integer.
ex:Sandi%20Morris rdf:type ~~foaf:Person~~ bibo: Document.
ex:Sandi%20Morris foaf:name "Sandi Morris"^^xsd:integer.

rank	name	mark
1	Anzhelika Sidorova	4.95
2	Sandi Morris	4.90
3	Katerina Stefanidi	4.85
4	Holly Bradshaw	4.80
5	Alysha Newman	4.80
6	Angelica Bengtsson	4.80

ex:Anzhelika%20Sidorova rdf:type ~~foaf:Person~~ bibo: Document .
ex:Anzhelika%20Sidorova foaf:name "Anzhelika Sidorova"^^xsd:integer.
ex:Sandi%20Morris rdf:type ~~foaf:Person~~ bibo: Document.
ex:Sandi%20Morris foaf:name "Sandi Morris"^^xsd:integer.

Mapping rules quality assessment



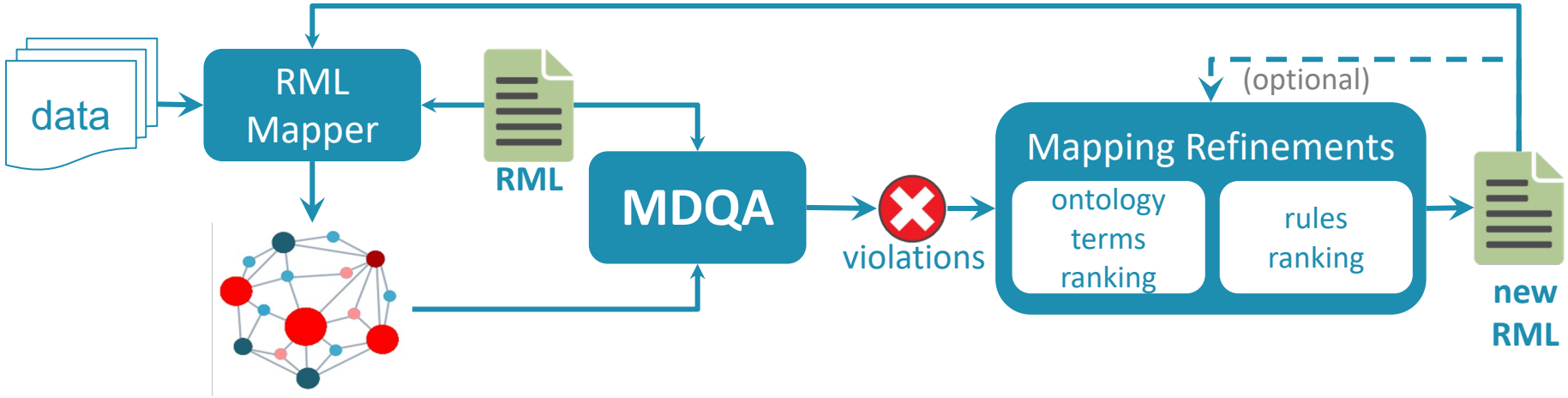
DBpedia Quality Assessment
knowledge graph: 16h
RML rules: 32s

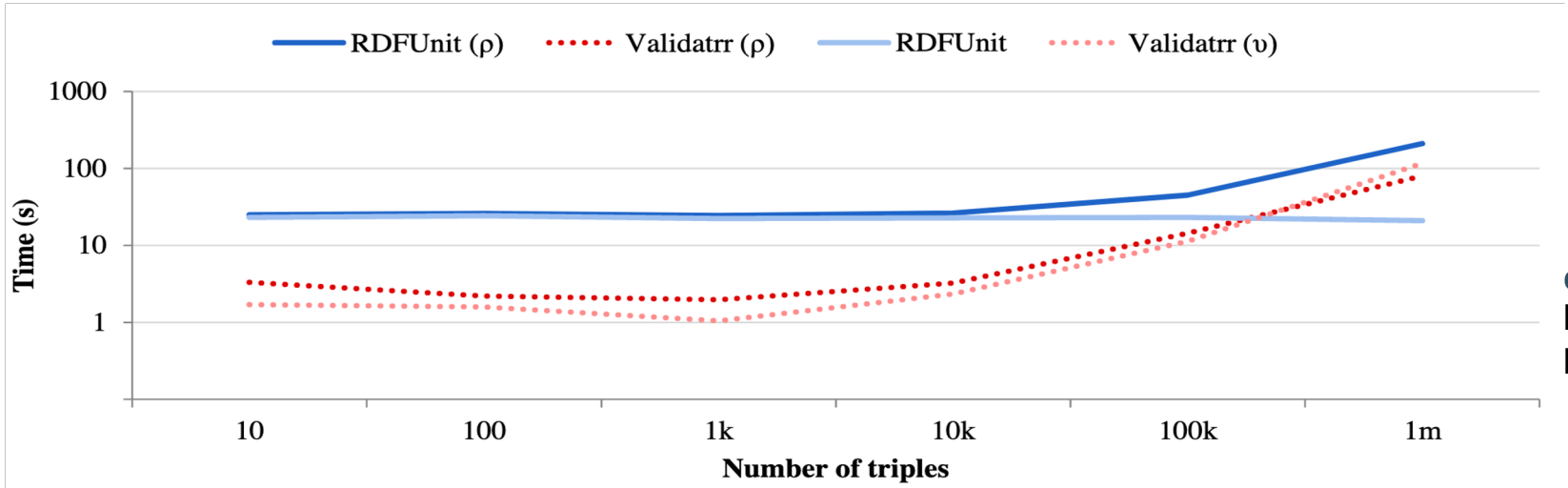
Certain test cases require a complete Linked Data set
e.g., (qualified) cardinality, (inverse) functionality, (a)symmetricity, irreflexivity

Assumption
ontology terms are perfect!

Assessing and Refining Mappings to RDF to Improve Dataset Quality. A. Dimou et al. ISWC 2015
doi.org/10.1007/978-3-319-25010-6_8

Mapping rules quality assessment





datasets have more than 100k triples
lodstats: 94%
lodlaundromat: 60%

Rule-driven inconsistency resolution for knowledge graph generation rules P. Heyvaert, A. Dimou, B. De Meester, R. Verborgh. 2019 (SWJ)
doi.org/10.3233/SW-190358

What did I do?

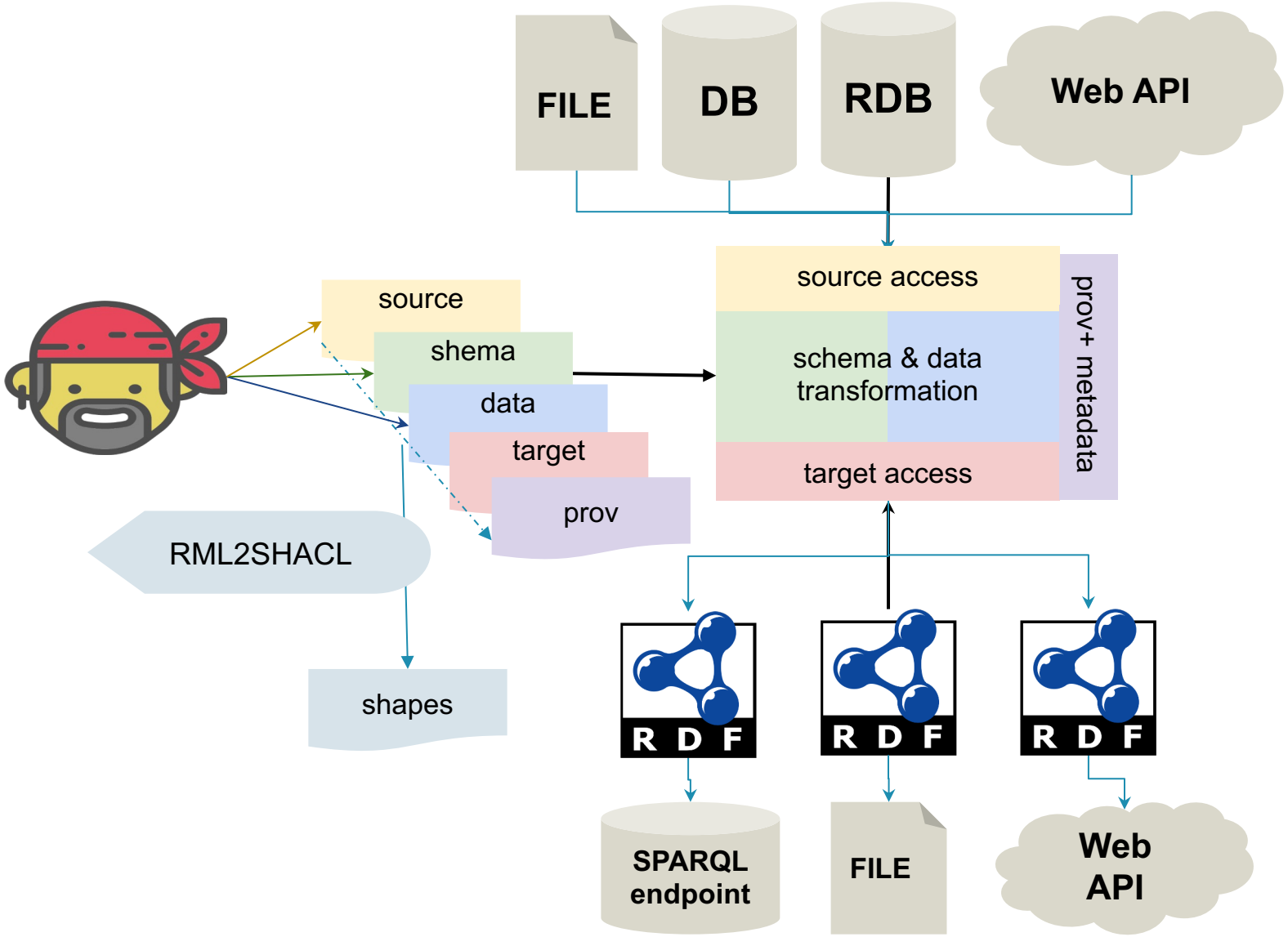
Mapping Languages & RML

Mapping rules & User support

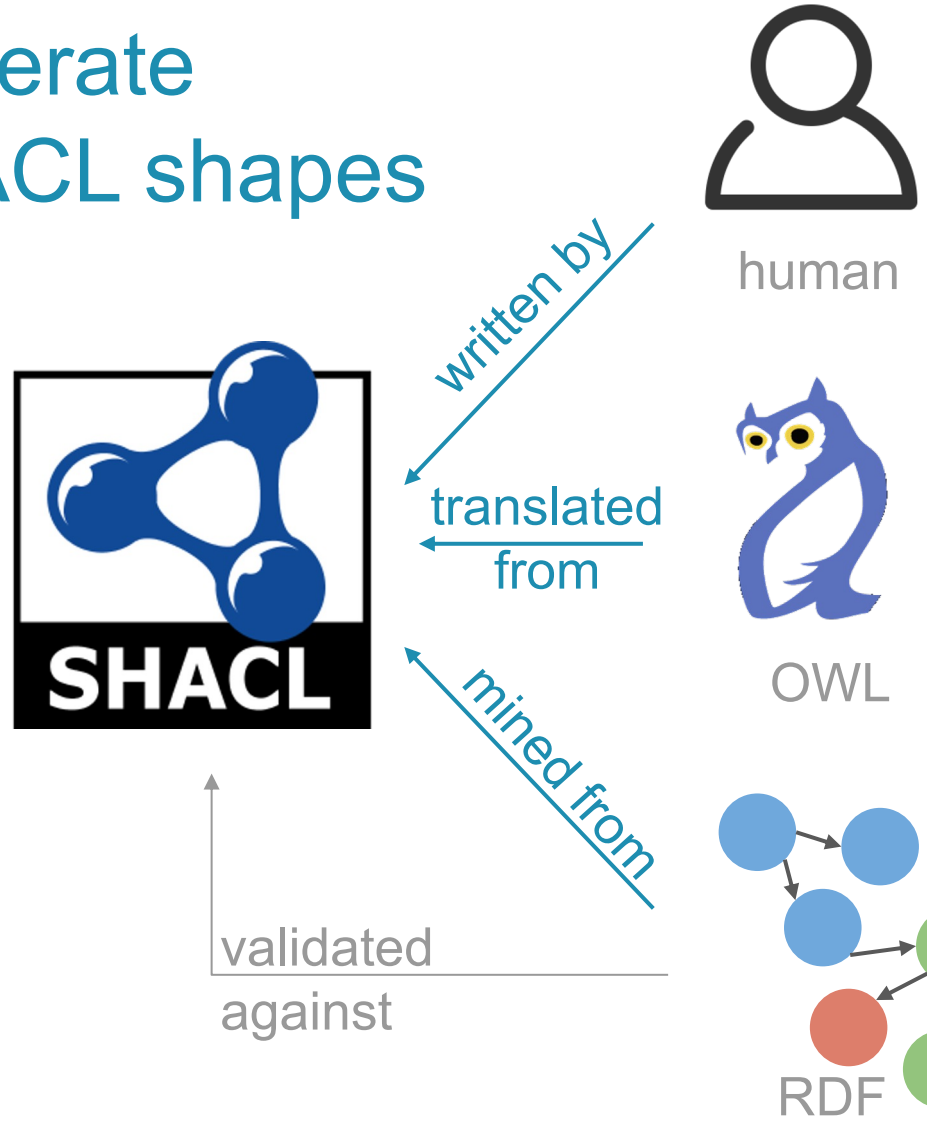
Mapping rules & RDF graphs validation

Mapping rules & SHACL shapes

Implementations & benchmarks


FILE
DB
RDB
Web API
source access
prov+ metadata
source
shema
data
target
prov
schema & data transformation
target access
RML2SHACL
shapes
R D F
R D F
R D F
SPARQL endpoint
FILE
Web API

Generate SHACL shapes



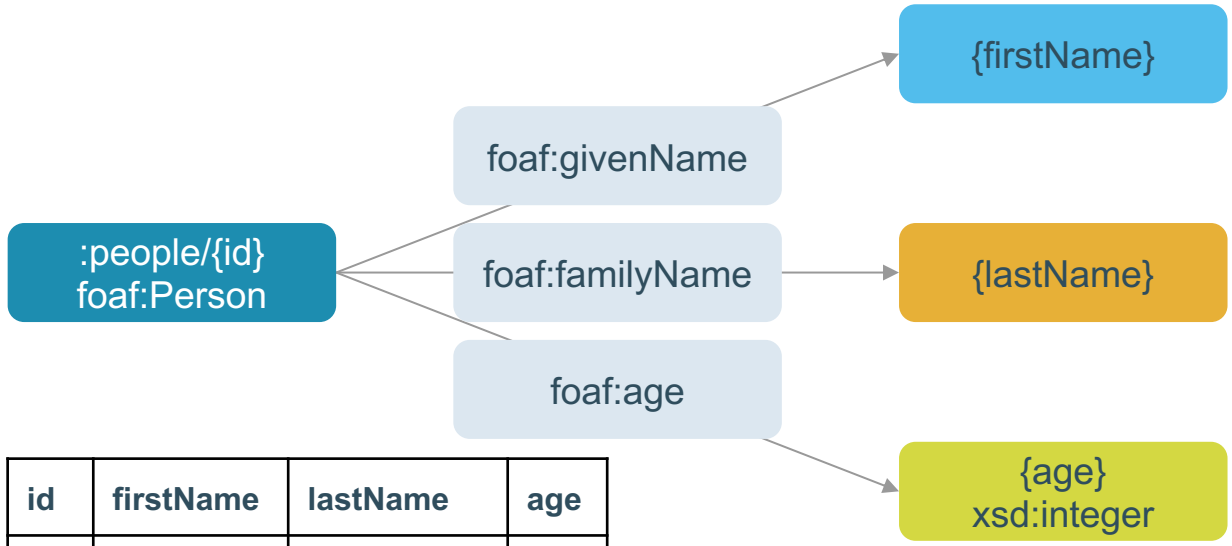
+ most control
− human effort needed

+ constant in data-size
− limited to ontological constraints, ontology required

+ always possible
− expensive for large data, uncertain in nature

Shape Designer for ShEx and SHACL Constraints. Boneva et al. ISWC 2019
Visual Notations for Viewing RDF Constraints with UnSHACLed. Lieber, De Meester, Heyvaert, Brückmann, Wambacq, Mannens, Verborgh, Dimou. SWJ 2021
Astrea: Automatic Generation of SHACL Shapes from Ontologies. A. Cimmino, A. Fernández-Izquierdo, and R. García-Castro. ESWC2020

Generate RDF with RML

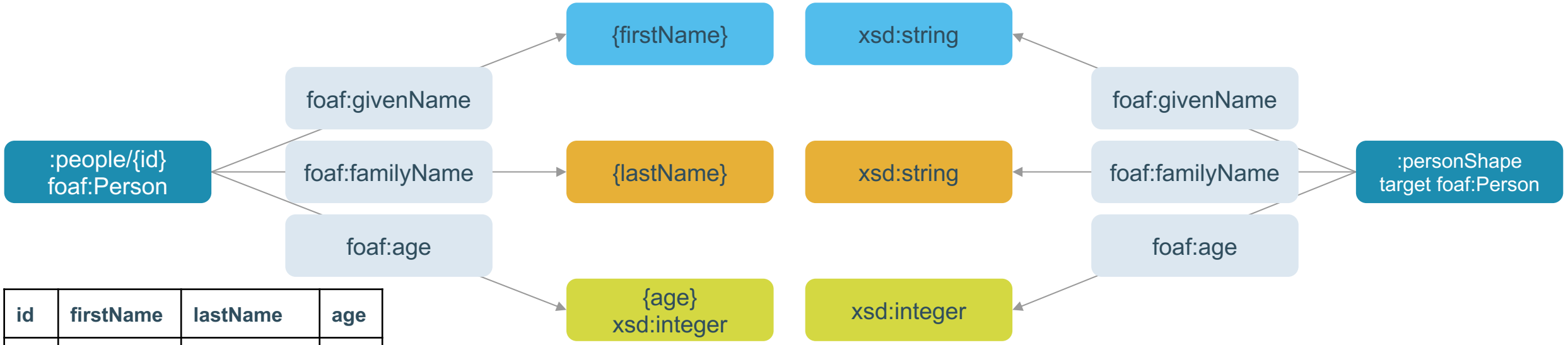


id	firstName	lastName	age
1	Thomas	Delva	26
2	Birte	De Smedt	22
3	Sitt	Min Oo	24
4	Dylan	Van Assche	26
5	Sven	Lieber	30
6	Anastasia	Dimou	20

```
:people/1 a foaf:Person;
          foaf:givenName "Thomas";
          foaf:familyName "Delva";
          foaf age "26"^xsd:integer.
```


RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data. A. Dimou et al. LDOW 2014
ceur-ws.org/Vol-1184/ldow2014_paper_01.pdf

RML mapping rules & SHACL shapes

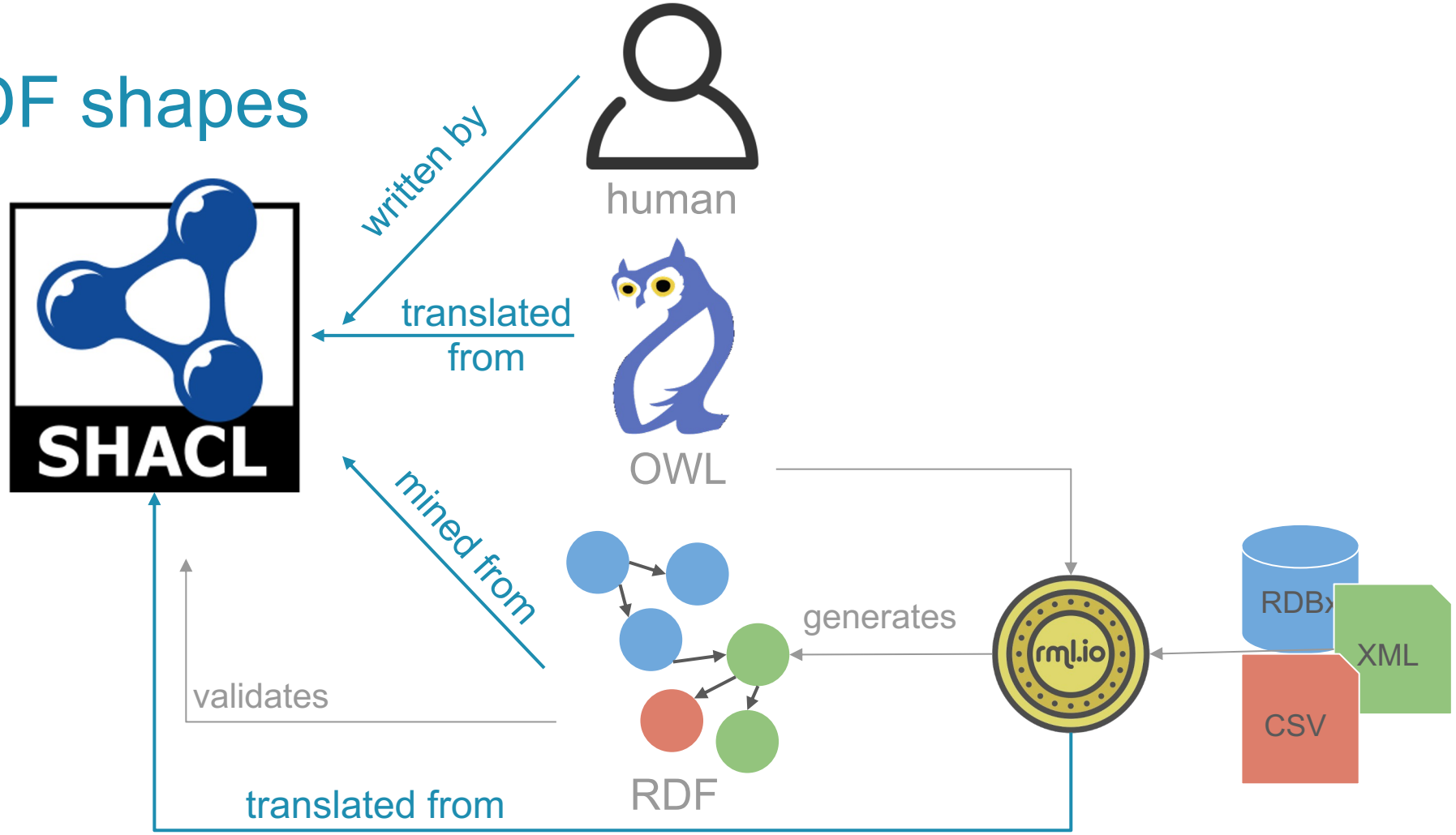


id	firstName	lastName	age
1	Thomas	Delva	26
2	Birte	De Smedt	22
3	Sitt	Min Oo	24
4	Dylan	Van Assche	26
5	Sven	Lieber	34
6	Anastasia	Dimou	20

```
:personShape a sh:NodeShape ;
    sh:targetClass foaf:Person ;
    sh:class foaf:Person ;
    sh:nodeKindsh:IRI ;
    sh:pattern ".*people/.*" ;
    sh:property [ sh:path foaf:givenName ;
                  sh:nodeKind sh:Literal ] ;
                [ sh:path foaf:familyName ;
                  sh:nodeKind sh:Literal ] ;
                [ sh:path foaf:age ;
                  sh:nodeKind xsd:integer ] ;
```

https://www.w3.org/TR/shacl/#

Generate RDF shapes from RML



RML2SHACL: RDF Generation Taking Shape T. Delva, B. De Smedt, S. Min Oo, D. Van Assche, S. Lieber, A. Dimou K-CAP 2021

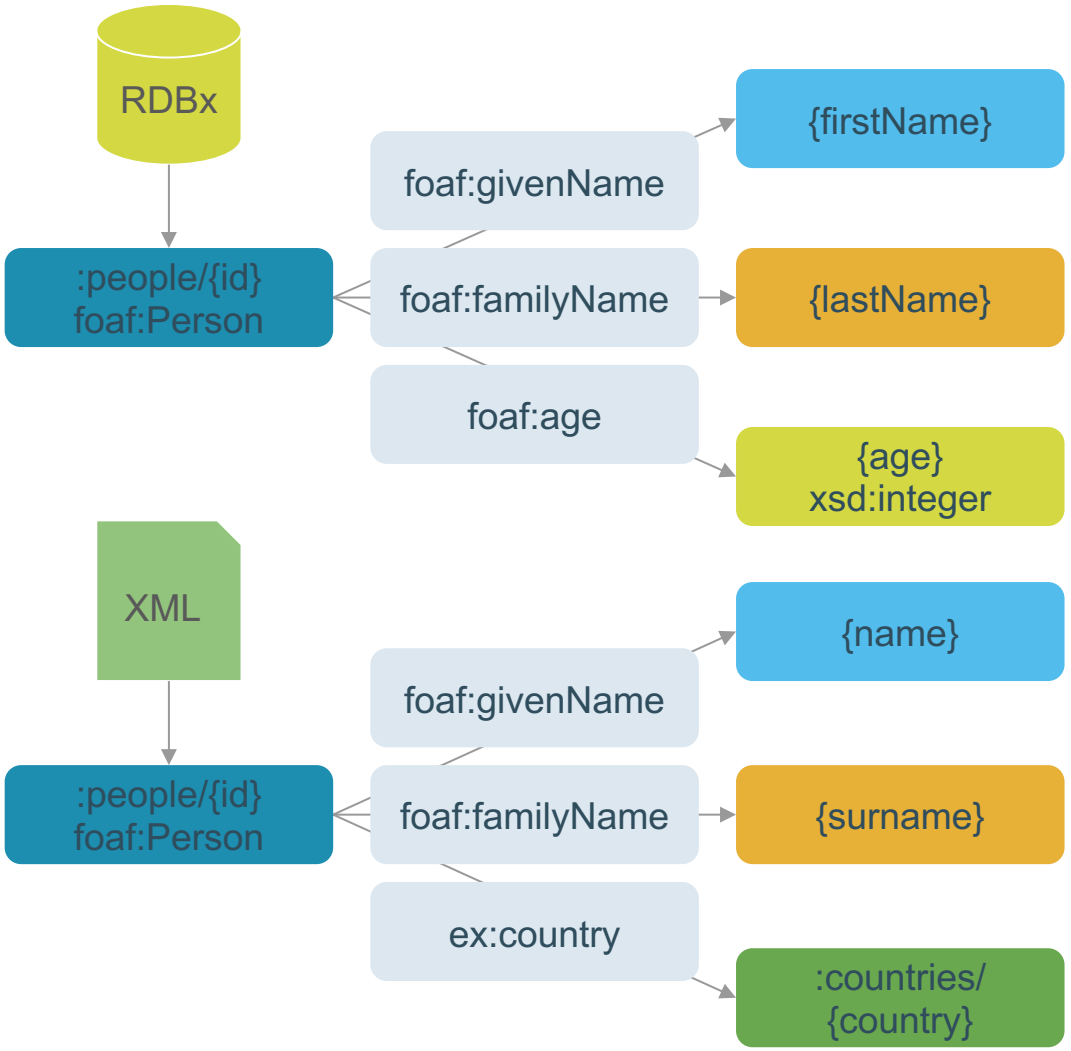
KU LEUVEN

RML & SHACL alignment

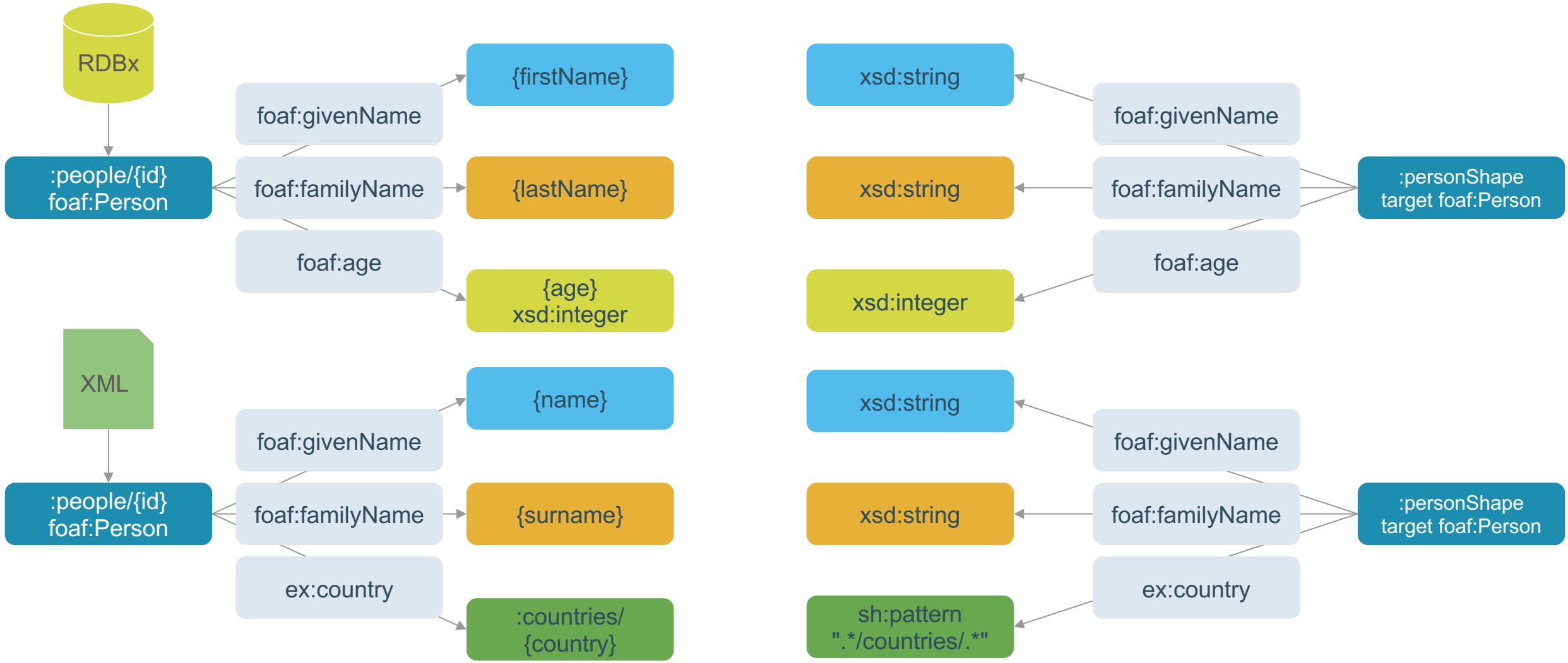
RML	SHACL
`rr:subjectMap & rr:SubjectMap`	`sh:NodeShape`
`rr:predicateObjectMap`	`sh:property`
`rr:PredicateObjectMap`	`sh:PropertyShape`
`rr:class`	`sh:class, sh:targetClass`
`rr:predicate`	`sh:path`
`rr:referencingObjectMap`	`sh:node`
`rr:termType`	`sh:nodeKind`
`rr:datatype`	`sh:datatype`
`rr:language`	`sh:languageIn`
`rr:constant`	`:in`
`rr:template`	`sh:pattern`

***note**: the alignment is not formally described yet*

RML mapping rules

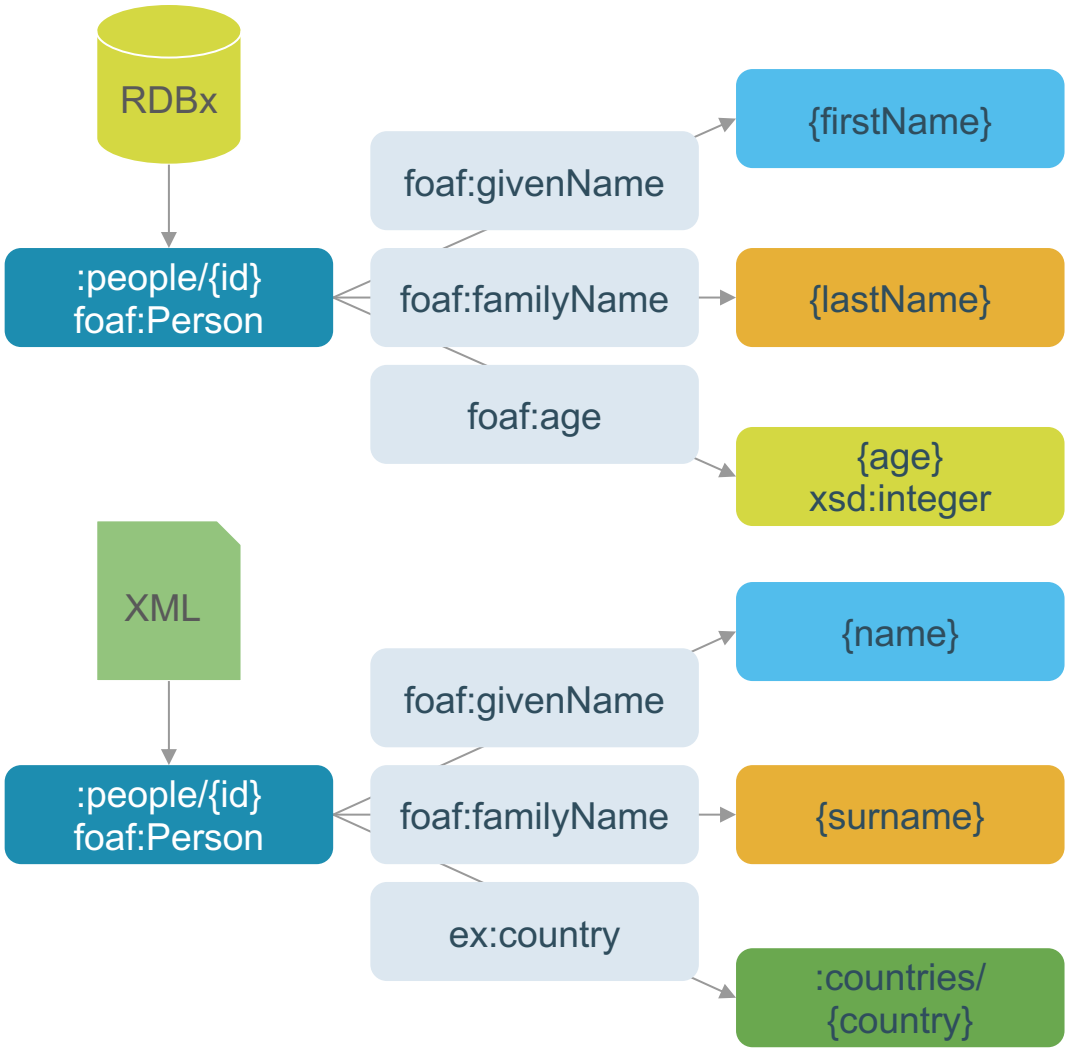


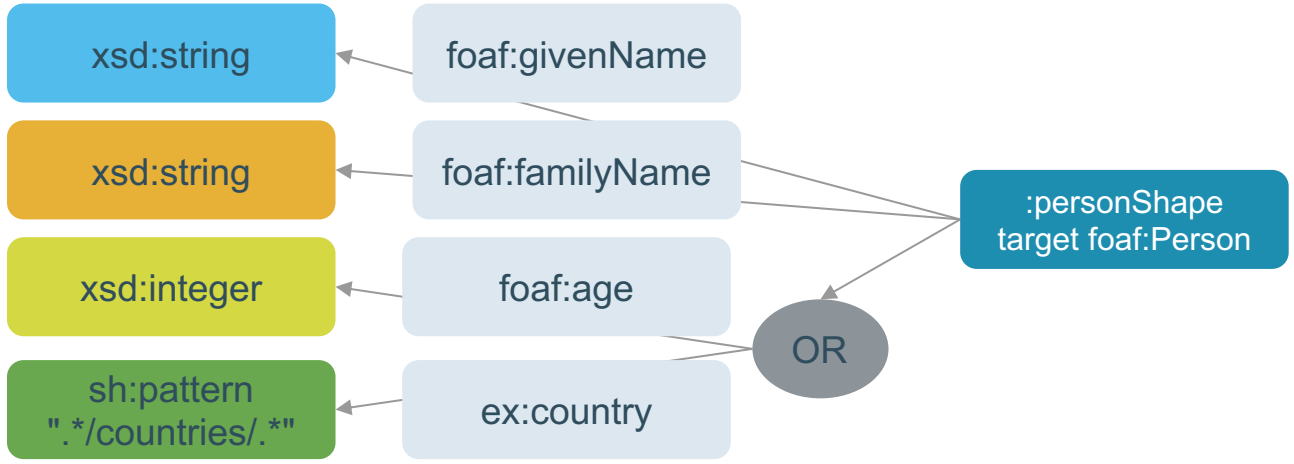
RML mapping rules & SHACL shapes



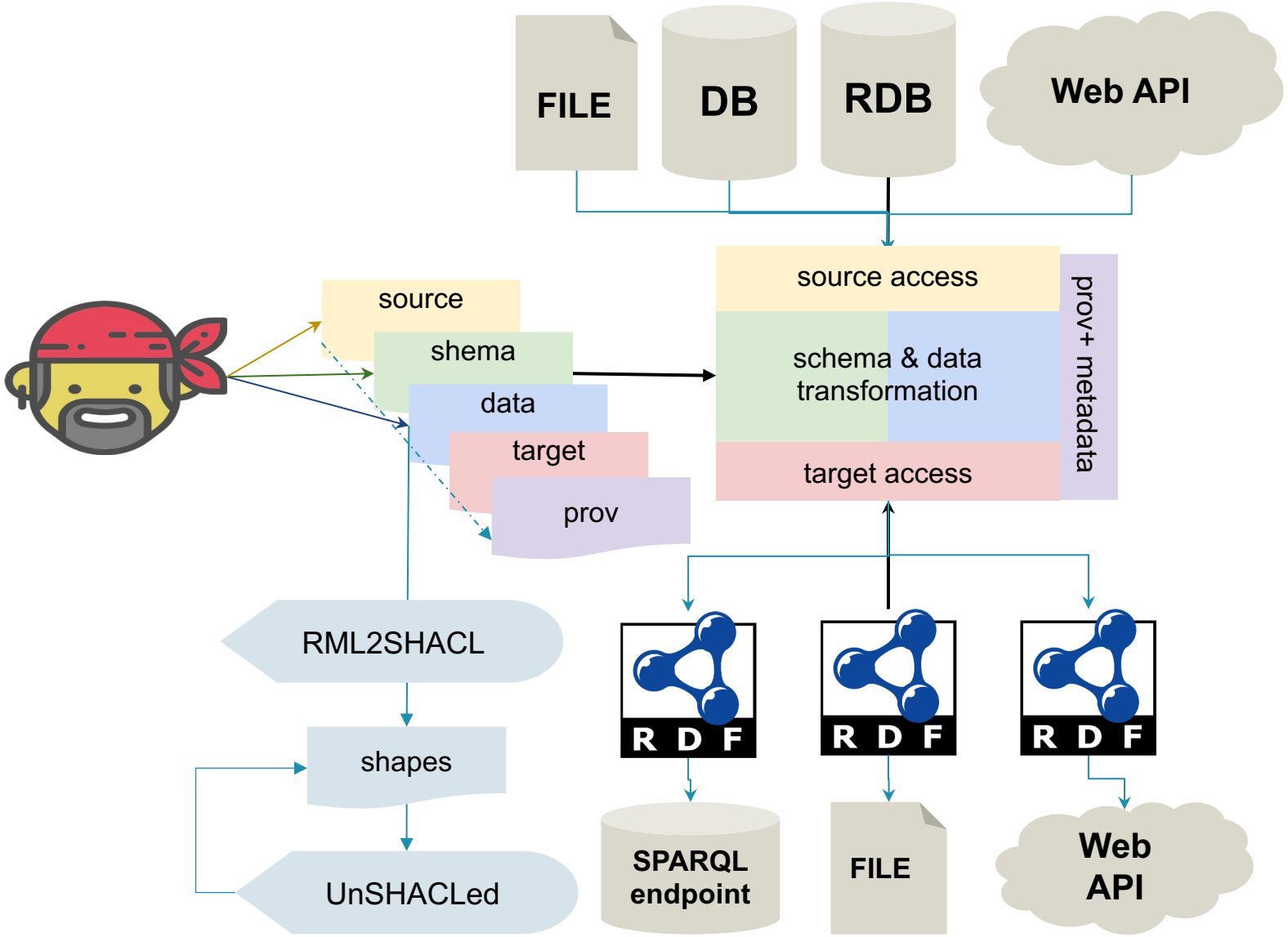
KU LEUVEN

RML mapping rules → SHACL shapes

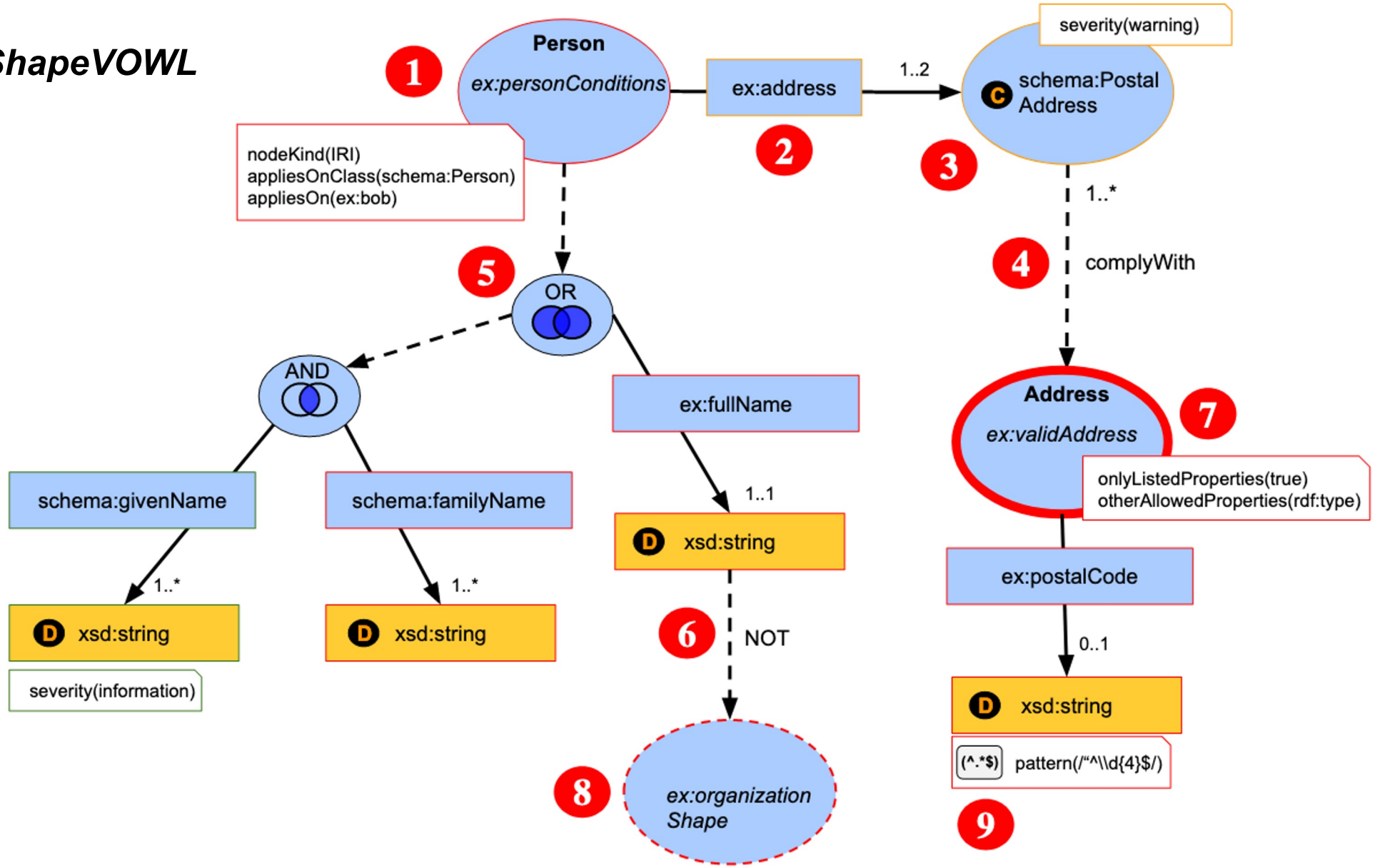




KU LEUVEN


FILE
DB
RDB
Web API
source access
prov+ metadata
source
shema
data
target
prov
schema & data transformation
target access
RML2SHACL
shapes
UnSHACLed
R D F
R D F
R D F
SPARQL endpoint
FILE
Web API

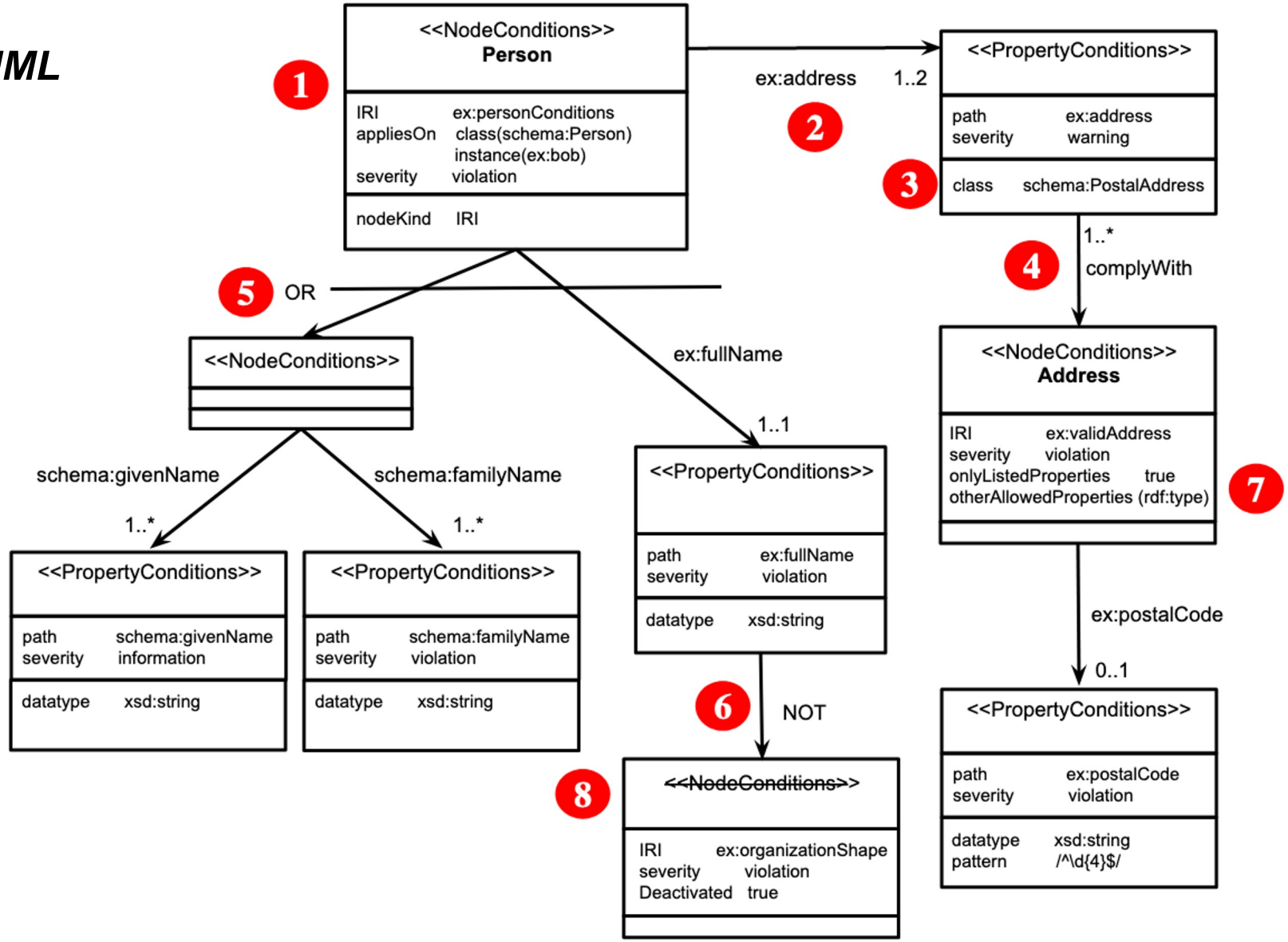
ShapeVOWL



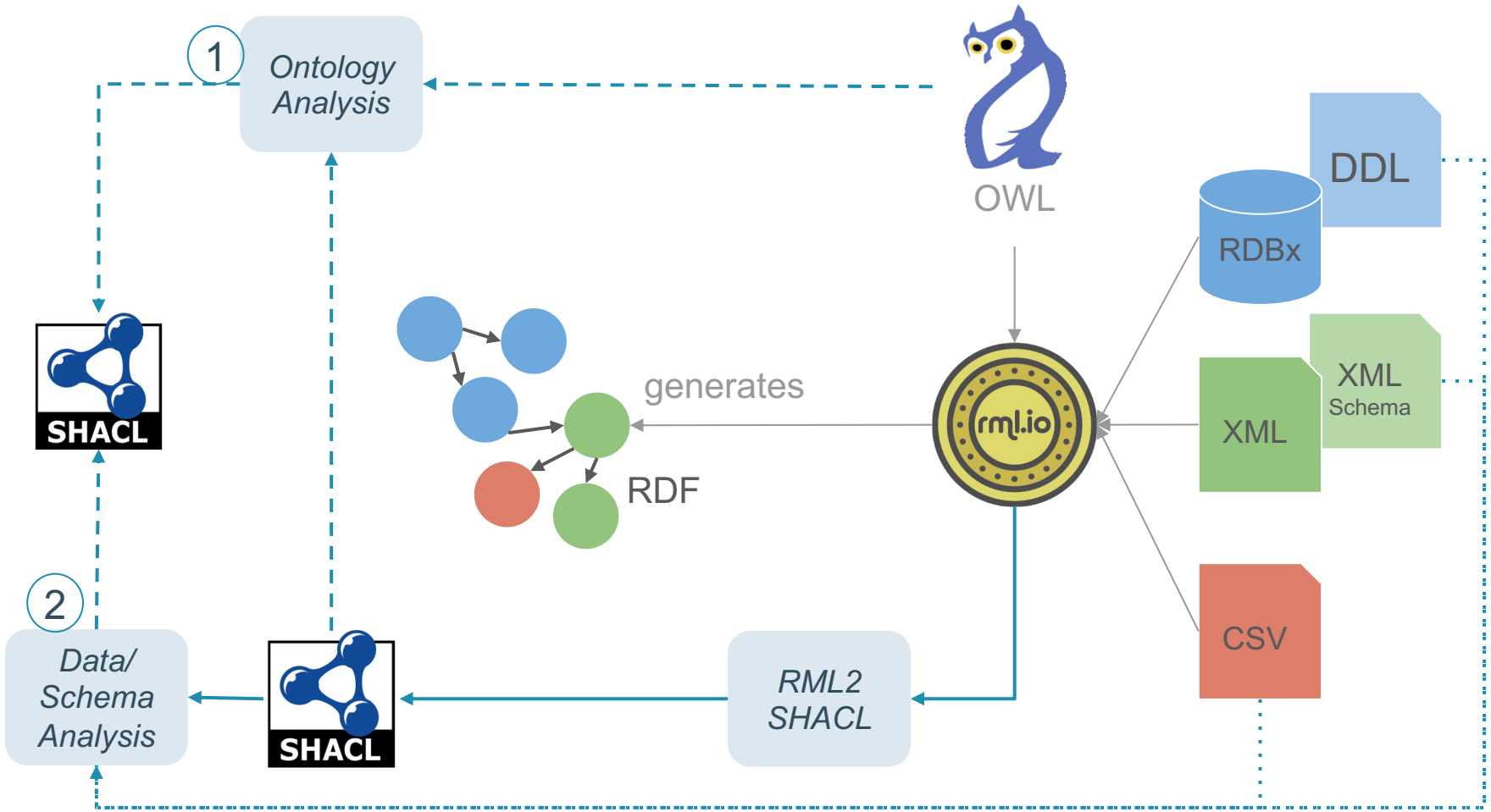
Visual Notations for Viewing and Editing RDF Constraints with UnSHACLed
S. Lieber, B. De Meester, P. Heyvaert, F. Brückmann, R. Wambacq, E. Mannens, R. Verborgh, A. Dimou (2021) SWJ
doi.org/10.3233/SW-210450

KU LEUVEN

ShapeUML

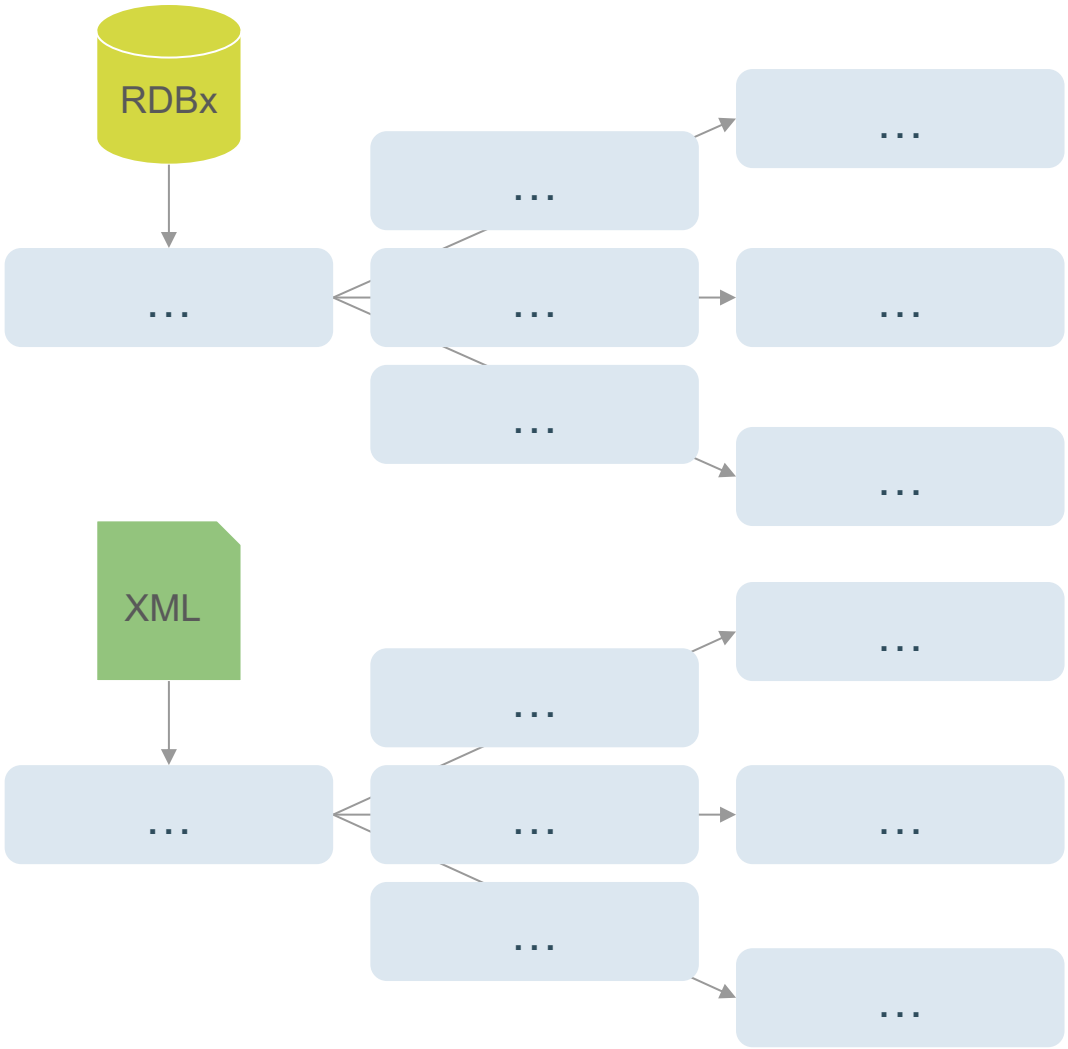

<<NodeConditions>>
Person
IRI ex:personConditions
appliesOn class(schema:Person)
instance(ex:bob)
severity violation
nodeKind IRI
1
ex:address
1..2
2
<<PropertyConditions>>
path ex:address
severity warning
3
class schema:PostalAddress
1..*
4
complyWith
<<NodeConditions>>
Address
IRI ex:validAddress
severity violation
onlyListedProperties true
otherAllowedProperties (rdf:type)
7
ex:postalCode
0..1
<<PropertyConditions>>
path ex:postalCode
severity violation
datatype xsd:string
pattern /^\d{4}$/
5
OR
<<NodeConditions>>
schema:givenName
1..*
<<PropertyConditions>>
path schema:givenName
severity information
datatype xsd:string
schema:familyName
1..*
<<PropertyConditions>>
path schema:familyName
severity violation
datatype xsd:string
ex:fullName
1..1
<<PropertyConditions>>
path ex:fullName
severity violation
datatype xsd:string
6
NOT
8
<<NodeConditions>>
IRI ex:organizationShape
severity violation
Deactivated true

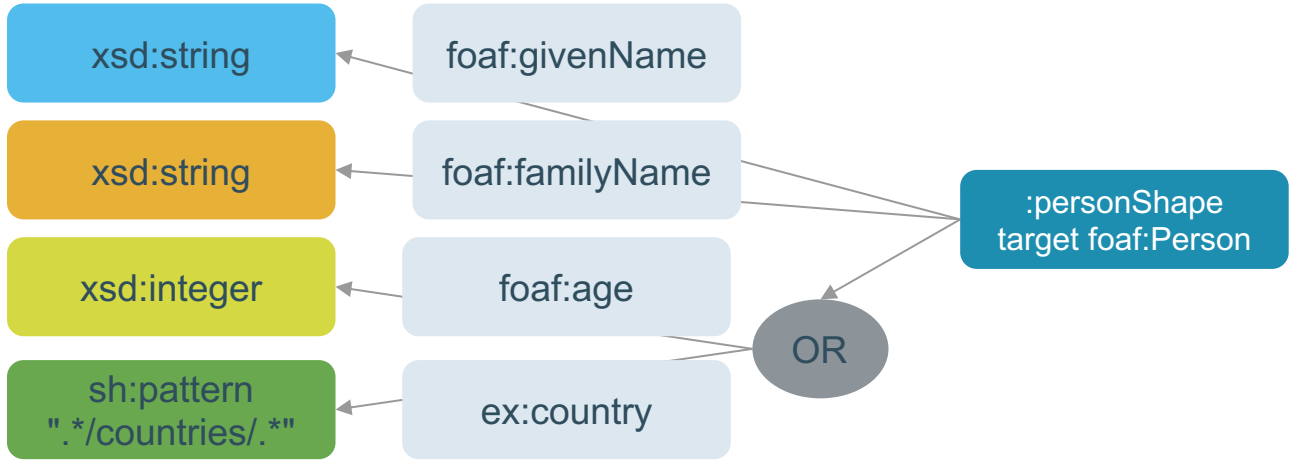
Enriching SHACL shapes



Extra input:
a) *well-known data schema*
b) *analysis of the data*
c) *analysis of the ontology*

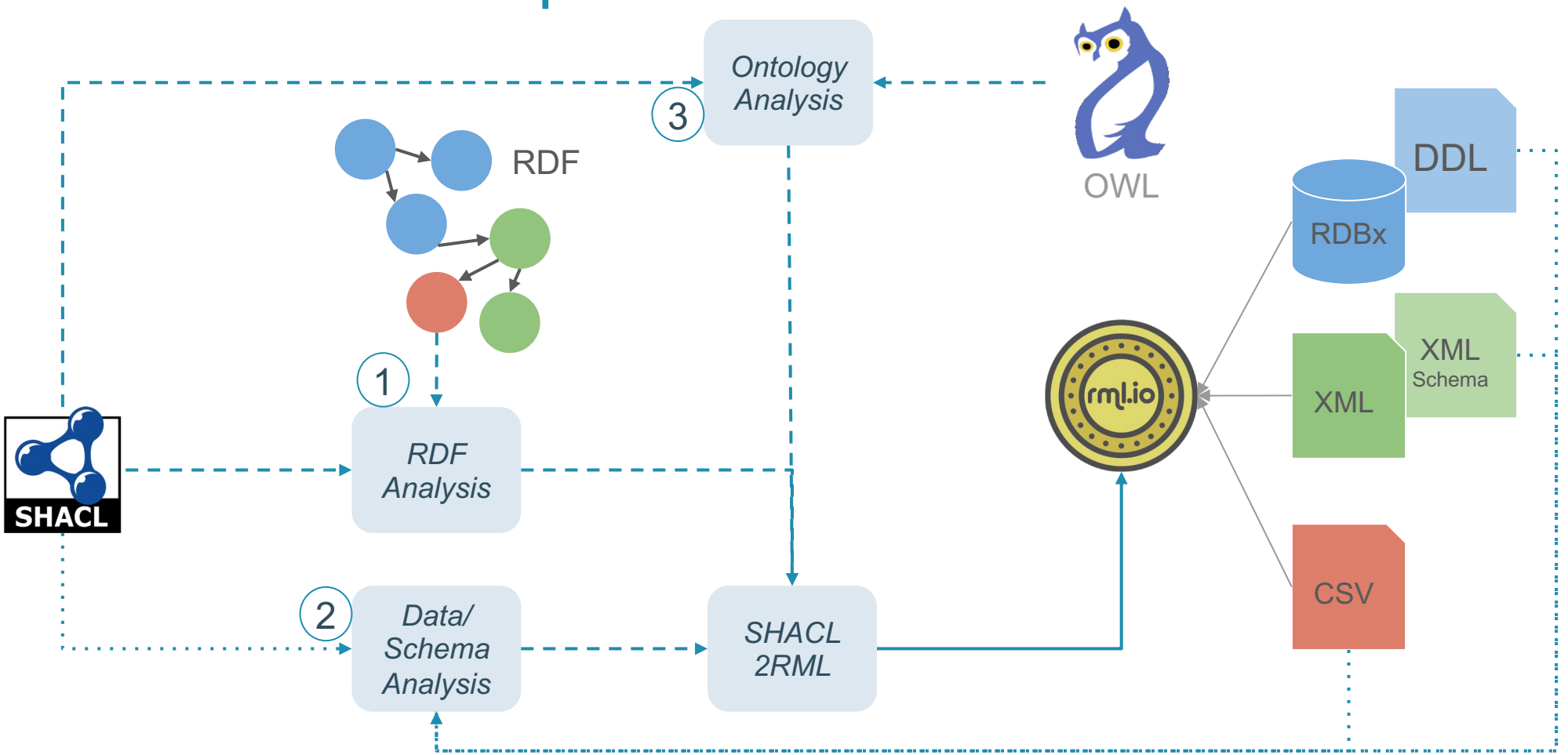
RML mapping rules ← SHACL shapes





KU LEUVEN

Generating RML mapping rules from SHACL shapes



Input:
a) RDF graph
b) SHACL shapes
c) ontology
d) data schema
Output:
RML mapping rules

KU LEUVEN

What did I do so far?

Mapping Languages & RML

Mapping rules & User support

Mapping rules & RDF graphs validation

Mapping rules & SHACL shapes

Implementations & benchmarks

RML processors

materialisation

RMLMapper Java (github.com/RMLio/rmlmapper-java)
CARML Java (github.com/carml/carml)
RocketRML JavaScript (github.com/semantifyit/RocketRML)
Morph-xR2RML Scala (github.com/frmichel/morph-xr2rml)
RMLStreamer Flink (github.com/RMLio/RMLStreamer)
Chimera Camel (github.com/cefriel/chimera)
SDM-RDFizer heuristic-based planning (github.com/SDM-TIB/SDM-RDFizer)
FunMap function-free planning (github.com/SDM-TIB/FunMap)
MapSDI deduplication-based optimizations (github.com/SDM-TIB/MapSDI)
Morph-KGC mapping partitions (github.com/oeg-upm/morph-kgc)
GeoTriples geospatial data (github.com/LinkedEOData/GeoTriples)

DB2triples (github.com/antidot/db2triples)
R2RML Parser (github.com/nkons/r2rml-parser)
XSPARQL (xsparql.sourceforge.net/)

homogeneous data sources

heterogeneous data sources

Morph-RDB (github.com/oeg-upm/morph-rdb)
Ontop (github.com/ontop/ontop)

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

virtualisation

Declarative RDF graph generation from heterogeneous (semi-)structured data: a Systematic Literature Review.
D. Van Assche, T. Delva, G. Haesendonck, P. Heyvaert, B. De Meester, A. Dimou. 2022. JWS https://doi.org/10.1016/j.websem.2022.100753

Choose your implementation

Choose yourself the best tool for your needs!

http://rml.io/test-cases/
http://rml.io/implementation-report/

Benchmarks:

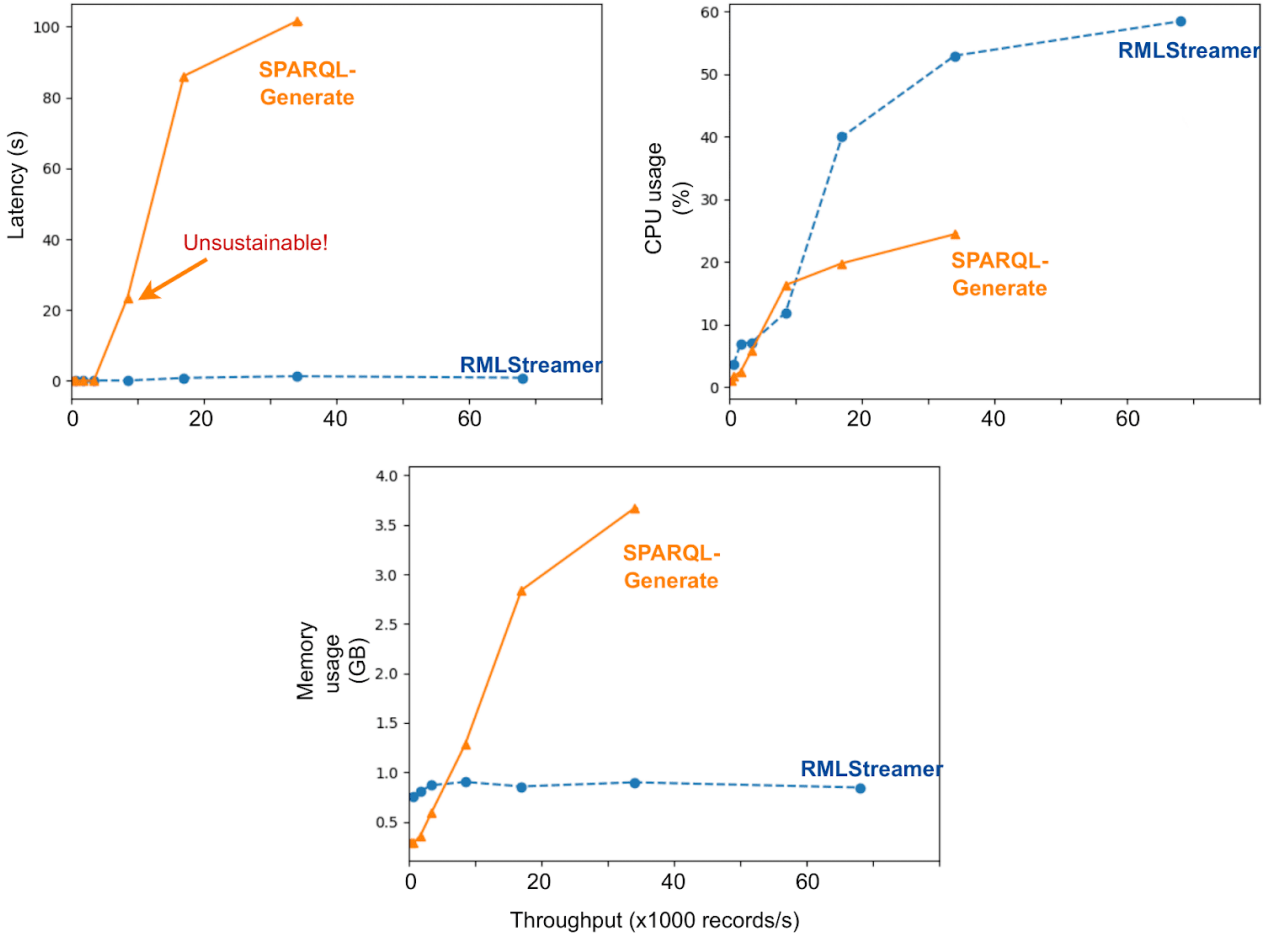
GTFS: evaluate tools generating RDF graphs with RML mapping rules (https://github.com/oeg-upm/gtfs-bench)

RODI:
test the quality of (semi-)automatically generated mapping rules (https://github.com/chrpin/rodi)

Test Case	RMLMapper	CARML	RocketRML	SDM-RDFizer	RMLStreamer	Chimera	Morph-KGC
RMLTC0000-CSV	passed	passed	passed	passed	passed	passed	passed
RMLTC0000-JSON	passed	passed	passed	passed	passed	passed	failed
RMLTC0000-MYSQL	passed	inapplicable	inapplicable	passed	inapplicable	inapplicable	passed
RMLTC0000-POSTGRESQL	passed	inapplicable	inapplicable	passed	inapplicable	inapplicable	passed
RMLTC0000-SPARQL	passed	inapplicable	inapplicable	inapplicable	inapplicable	inapplicable	inapplicable
RMLTC0000-SQLSERVER	passed	inapplicable	inapplicable	passed	inapplicable	inapplicable	inapplicable

Conformance test-cases for the RDF Mapping Language. P. Heyvaert et al. KGSWC2019 doi.org/10.1007/978-3-030-21395-4_12
GTFS-Madrid-Bench: A benchmark for virtual knowledge graph access in the transport domain Chaves-Fraga et al. 2020 (JWS) doi.org/10.1016/j.websem.2020.100596

RMLStreamer-SISO

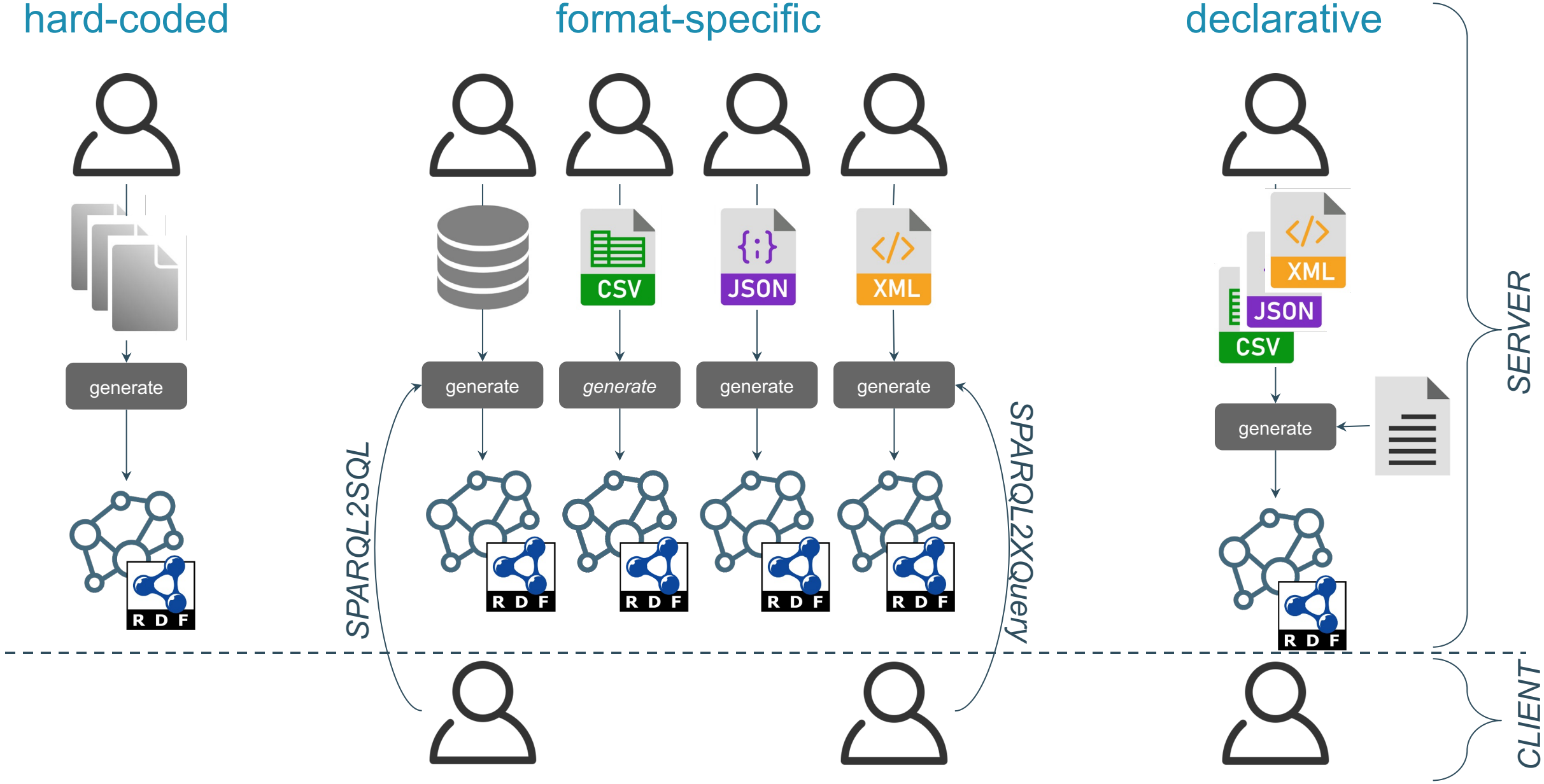


shameless advertisement:

Tuesday, 25 October 15:10-16:10

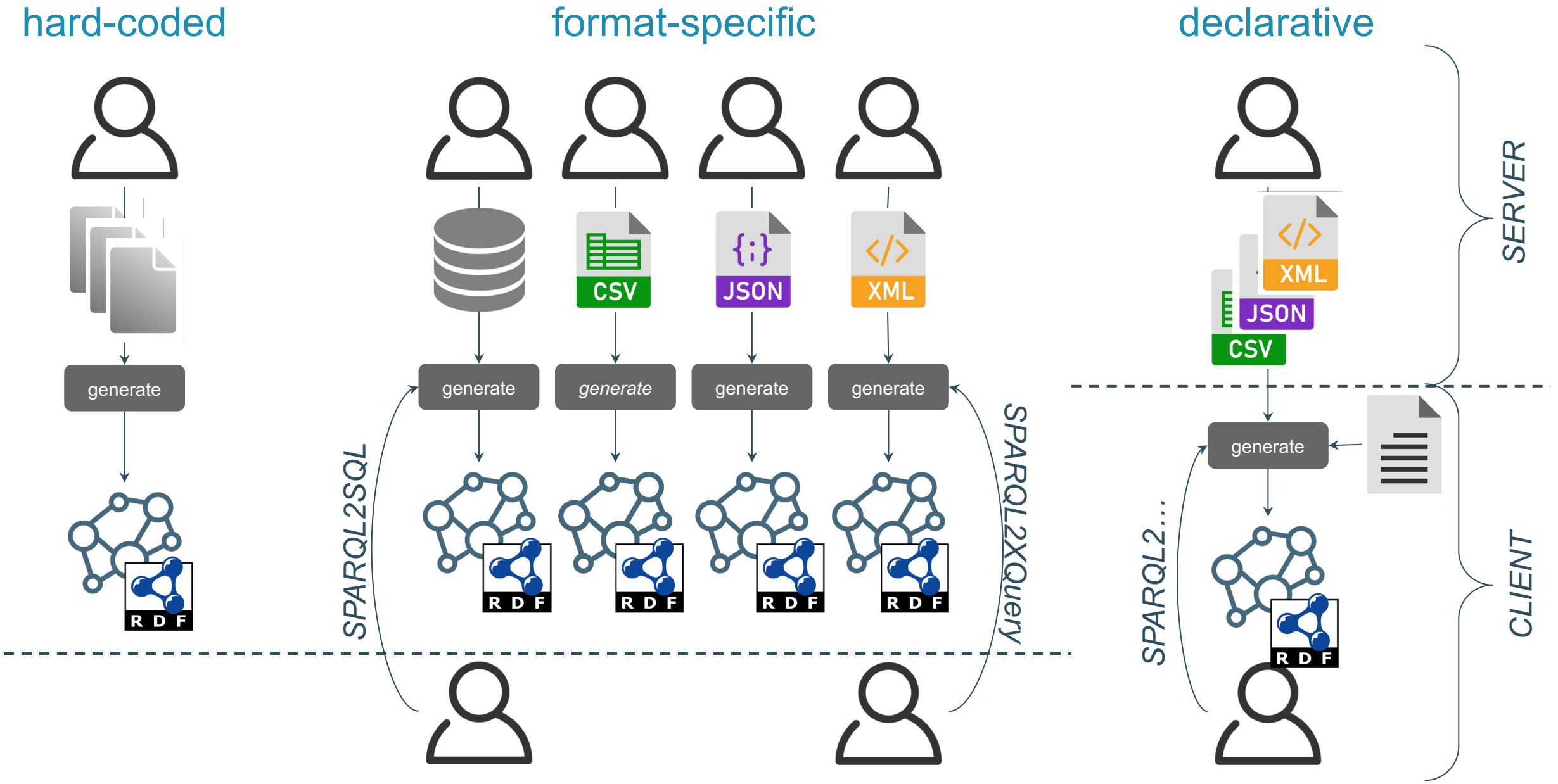
RMLStreamer-SISO: An RDF Stream Generator from Streaming Heterogeneous Data.
Sitt Min Oo, Gerald Haesendonck, Ben De Meester, Anastasia Dimou ISWC2022
doi.org/10.1007/978-3-031-19433-7_40

KU LEUVEN



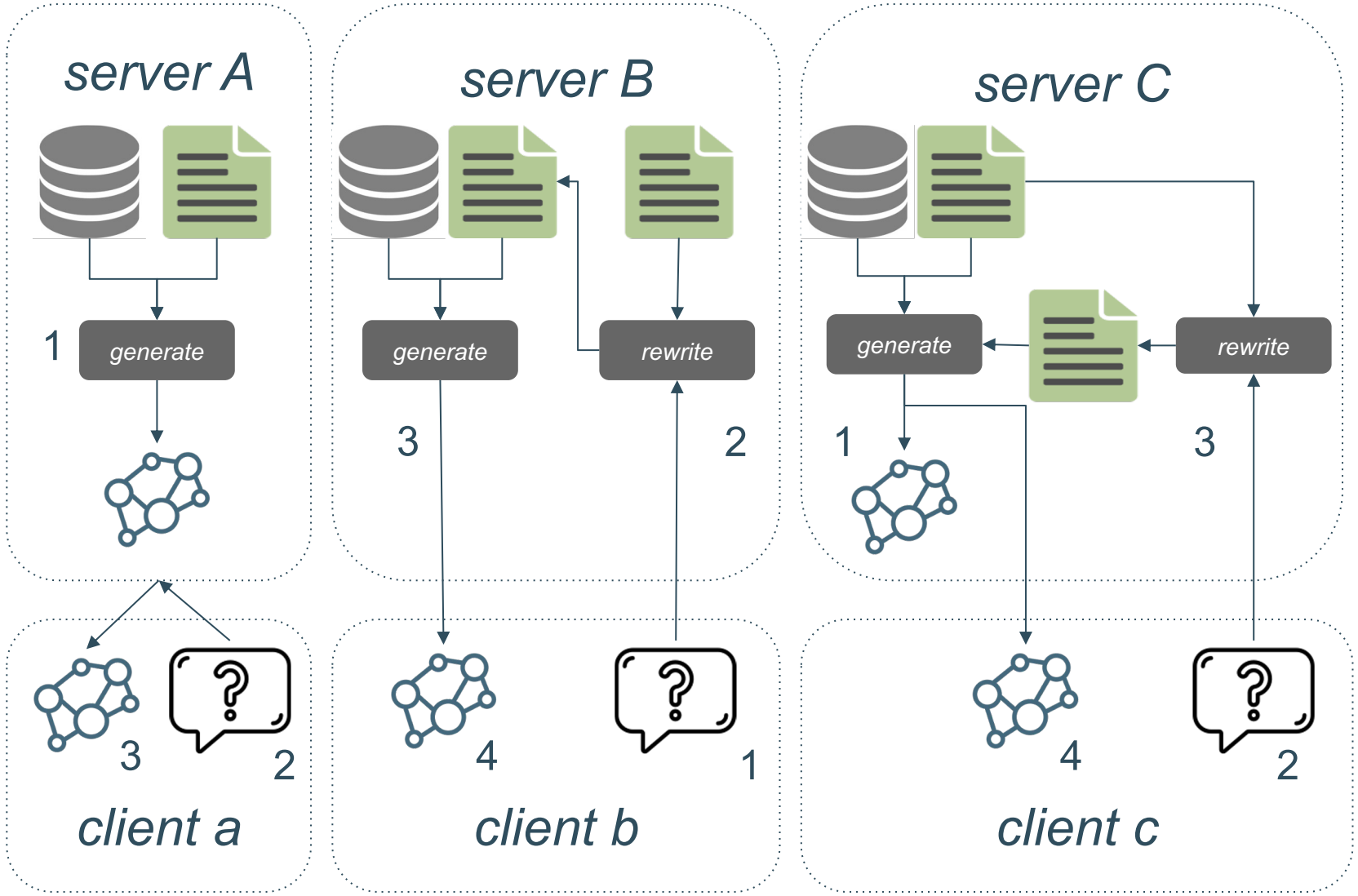
Declarative RDF graph generation from heterogeneous (semi-)structured data: a Systematic Literature Review.
D. Van Assche, T. Delva, G. Haesendonck, P. Heyvaert, B. De Meester, A. Dimou. 2022. JWS
https://doi.org/10.1016/j.websem.2022.100753

KU LEUVEN



Balancing RDF generation from heterogeneous data sources.
D. Van Assche. 2022. ISWC2022 PhD Consortium
https://doi.org/10.1007/978-3-031-11609-4_40

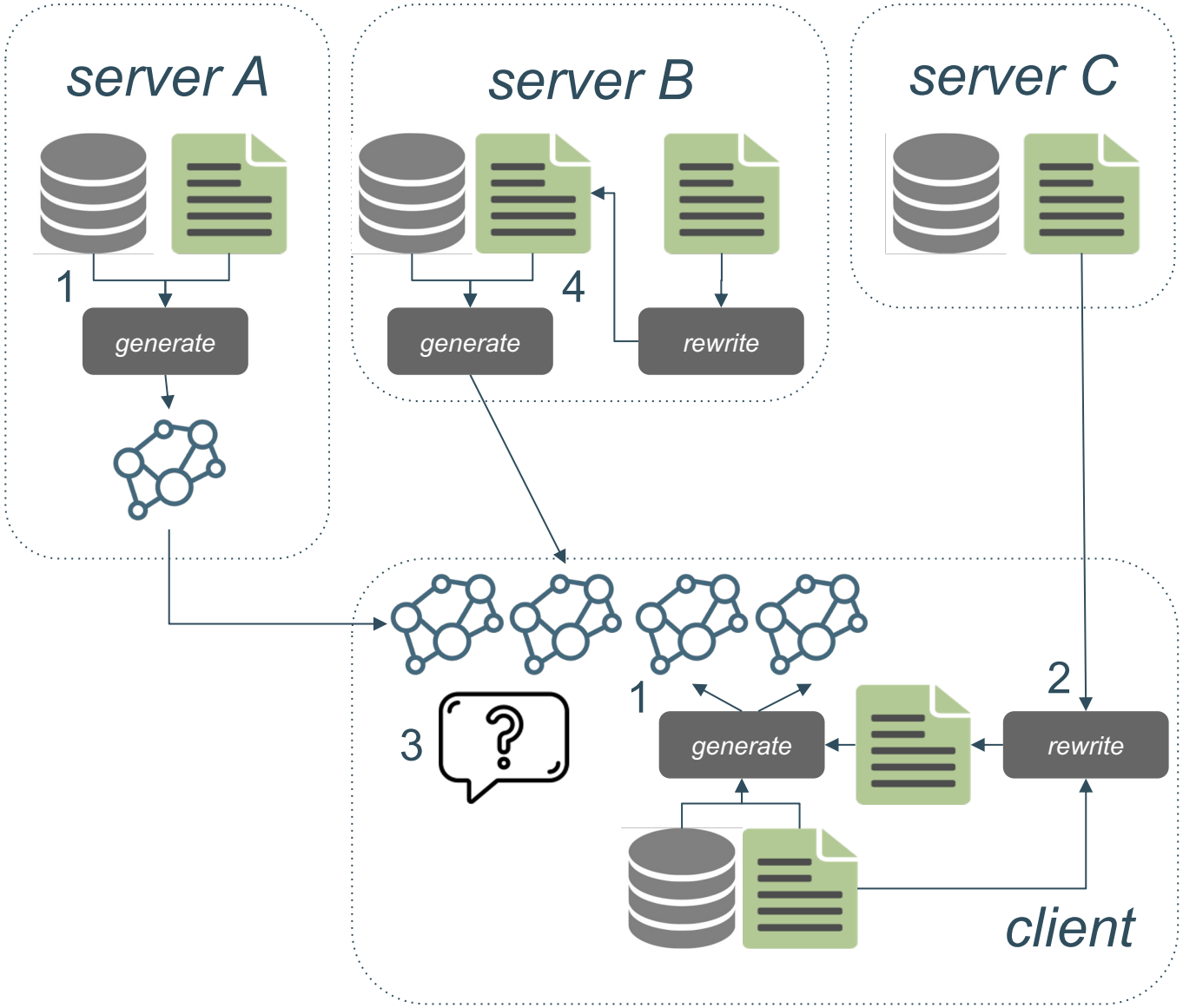
KU LEUVEN



knowledge graph materialisation

virtual knowledge graphs generation over heterogeneous data

hybrid materialisation & virtualisation for knowledge graph generation



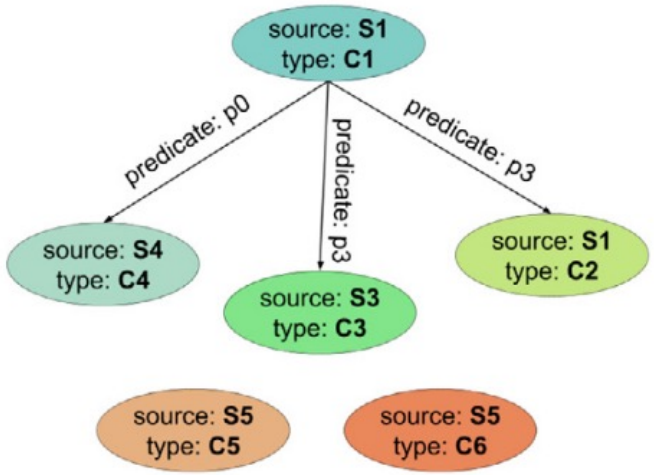
server A materializes its knowledge graph

server B generates a virtual knowledge graph

server C does not generate a knowledge graph

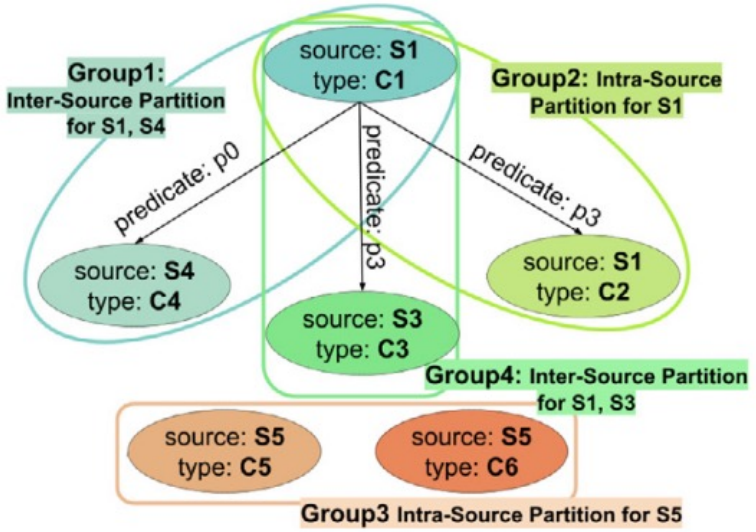
client materialises its own knowledge graph
& the knowledge graph for server C's data
receives a virtual knowledge graph from server B
receives a subgraph from server A
combines all subgraphs to perform its tasks

Mapping rules partition



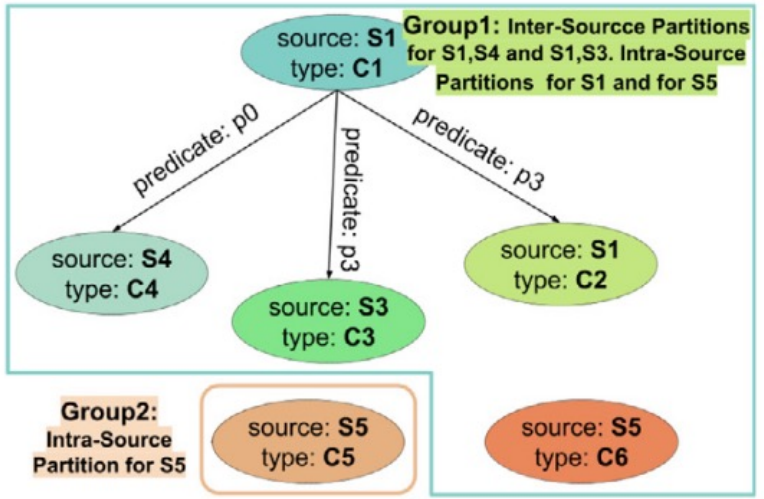
No Partitioning

RMLMapper: Five-hour timeout (No results)
RocketRML: Out of memory (No results)
SDM-RDFizer: 441.18 sec (100% results)
Morph-KGC: 42.32 sec (100% results)



Optimized Partitioning

RMLMapper: Five-hour timeout (92.65% results)
RocketRML: 133.82 sec (100% results)
SDM-RDFizer: 211.02 sec (100% results)
Morph-KGC: 36.81 sec (100% results)



Random Partitioning

RMLMapper: Five-hour timeout (5.41% results)
RocketRML: Out of memory (5.41% results)
SDM-RDFizer: 369.27 sec (100% results)
Morph-KGC: 40.21 sec (100% results)

Scaling up knowledge graph creation to large and heterogeneous data sources
E. Iglesias, S. Jozashoori, M.-E. Vidal (2022 JWS) doi.org/10.1016/j.websem.2022.100755

What did I do so far?

Mapping Languages & RML

Mapping rules & User support

Mapping rules & RDF graphs validation

Mapping rules & SHACL shapes

Implementations & benchmarks

What do we still miss?

Many RML implementations by now – focus on ***materialisation implementations***
still ***missing virtualisation & hybrid*** *(materialisation+virtualisation) solutions*

Many RML ***implementations for static data***
still ***missing solutions for dynamic/streaming & versioned data***

Community ***converges on mapping language*** *for heterogeneous semi-structured data*
still ***missing automated solutions*** *for mapping rules definition*

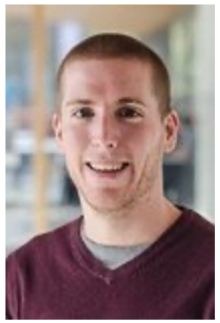
Only very ***few GUIs*** *to support the humans*
still ***missing thorough research on methodologies for knowledge graphs construction & human-in-the-loop***

Preliminary benchmarks
still missing more fine-grained ***benchmarking for all aspects of knowledge graph construction***

Preliminary quality assessment solutions
still ***missing systematic solutions for quality assessment*** *& feedback loop*

Thank you! former team @UGhent/imec

presentation based on works with
my former team at Ghent University
and my current team at KULeuven
and
many references to different members
of the Knowledge Graph Construction
Community Group

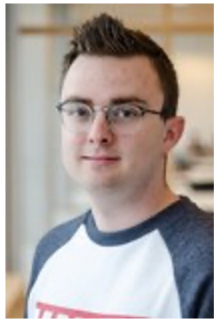


Dr Pieter Heyvaert

Dr Ben De Meester

Dr Sven Lieber

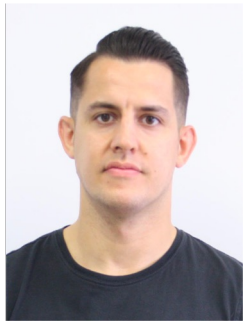
Ioannis Chrysakis



Dylan Van Assche

Sitt Min Oo

Gerald Haesendonck



Gertjan De Mulder

Thomas Delva

Thank you! team @KULeuven

prof Anastasia Dimou
anastasia.dimou@kuleuven.be
@natadimou

Dr David Chaves Fraga
postdoctoral fellow
(KU Leuven & UPM)
cleansing & benchmarking KGs

Ioannis Chrysakis
joint PhD student
(UGhent & KU Leuven)
KGs for dataspaces interoperability & privacy

Dylan Van Assche
PhD student (UGhent)
trade offs of KGs materialisation & virtualisation

Ioannis Dasoulas
PhD student (KU Leuven)
(June 2022 - now)
fusion of KGs & ML

Xuemin Duan
PhD student (KU Leuven)
(Sep 2022 - now)
automated KG generation

?

(to start Dec 2022)
KGs for manufacturing

?

(to start Jan 2023)
KG & Decision Making

KU LEUVEN

Knowledge Graphs Construction W3C Community Group

w3.org/community/kg-construct/

github.com/kg-construct/

>150 members world-wide

new RML specifications

(work in progress)

https://kg-construct.github.io/rml-core/

https://kg-construct.github.io/rml-target-source-spec/

https://w3id.org/kg-construct/collections-containers/

https://kg-construct.github.io/rml-star-spec/

Knowledge Graph Construction Workshop

https://kg-construct.github.io/workshop/

1st edition theme: mapping languages
2nd edition theme: human-in-the-loop
3rd edition: automated KG construction
4th edition (2023): benchmarking implementations?

networks on Knowledge Graphs

graph theory
data management
knowledge representation
programming languages

Knowledge Graphs for Data Integration (KG4DI)
w3id.org/kg4di/
Scientific Research Network
kick off: *November 30th*

Distributed Knowledge Graph COST Action (DKG)

DKG – cost-dkg.eu

>100 members from >30 countries in Europe

Are Knowledge Graphs Ready for the Real World? Challenges and Perspective

dagstuhl.de/24061

February 2024

The adoption of Semantic Web technologies depends on the availability of knowledge graphs

It's on our hands to investigate how to efficiently generate high quality knowledge graphs

KU LEUVEN LEUVEN.AI

Aligning heterogeneous semi-structured data and knowledge graphs

Anastasia Dimou
📨 anastasia.dimou@kuleuven.be
🐦 @natadimou