

# Integrated Linked Data Generation of Heterogeneous raw Data



# Linked Data

semantically annotated using  
different vocabularies or ontologies  
and interlinked data representations

describe domain-level knowledge  
that is understandable by  
both humans and machines

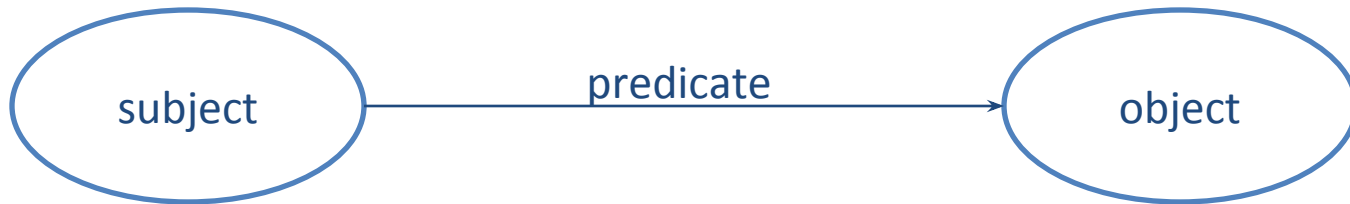
# Linked Data

derive from originally heterogeneous  
(semi-)structured data

published in the form of RDF datasets

# Resource Description Framework (RDF)

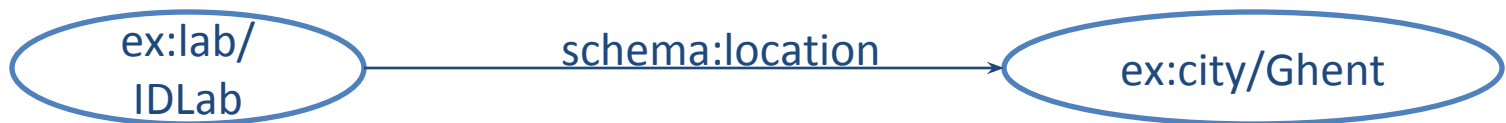
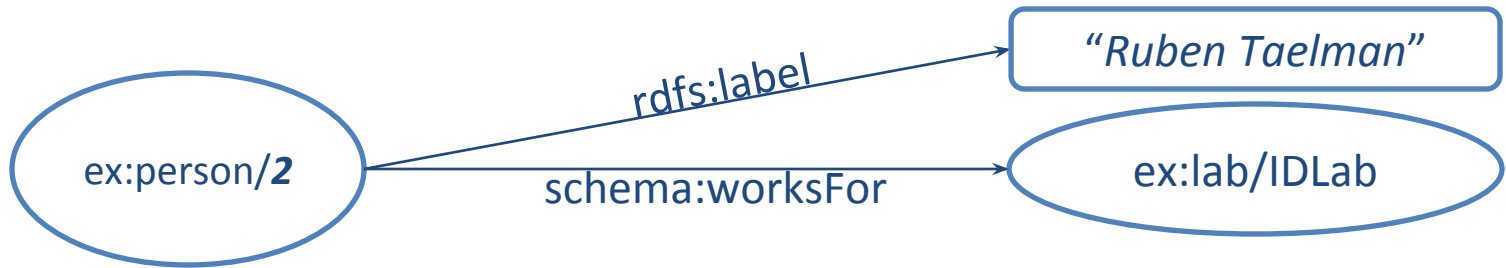
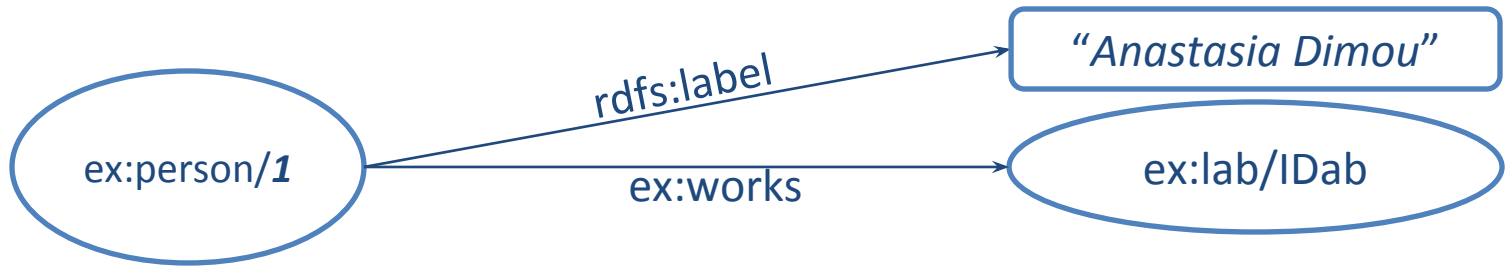
is the prevalent data model  
for describing **Linked Open Data**



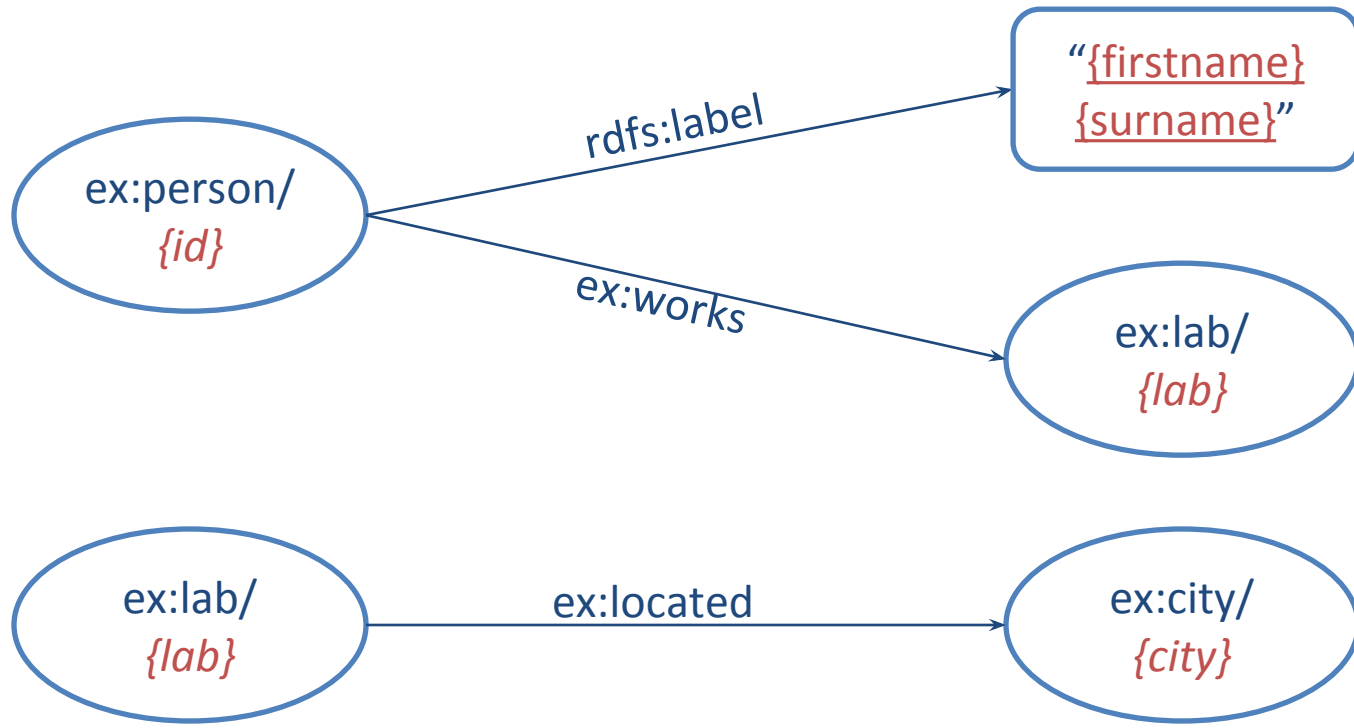
How is Linked Data derived  
from (semi-)structured data?

# How is Linked Data derived from tabular data?

<b>id</b>	<b>firstname</b>	<b>lastname</b>	<b>lab</b>	<b>city</b>
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent



sets of triples of a dataset have repetitive patterns



RDF dataset generation tools  
rely their implementation on repetitively  
applying those patterns to input data



Most mapping solutions

*provide **case-specific** solutions*

*OR*

map data on ***per-format** basis*

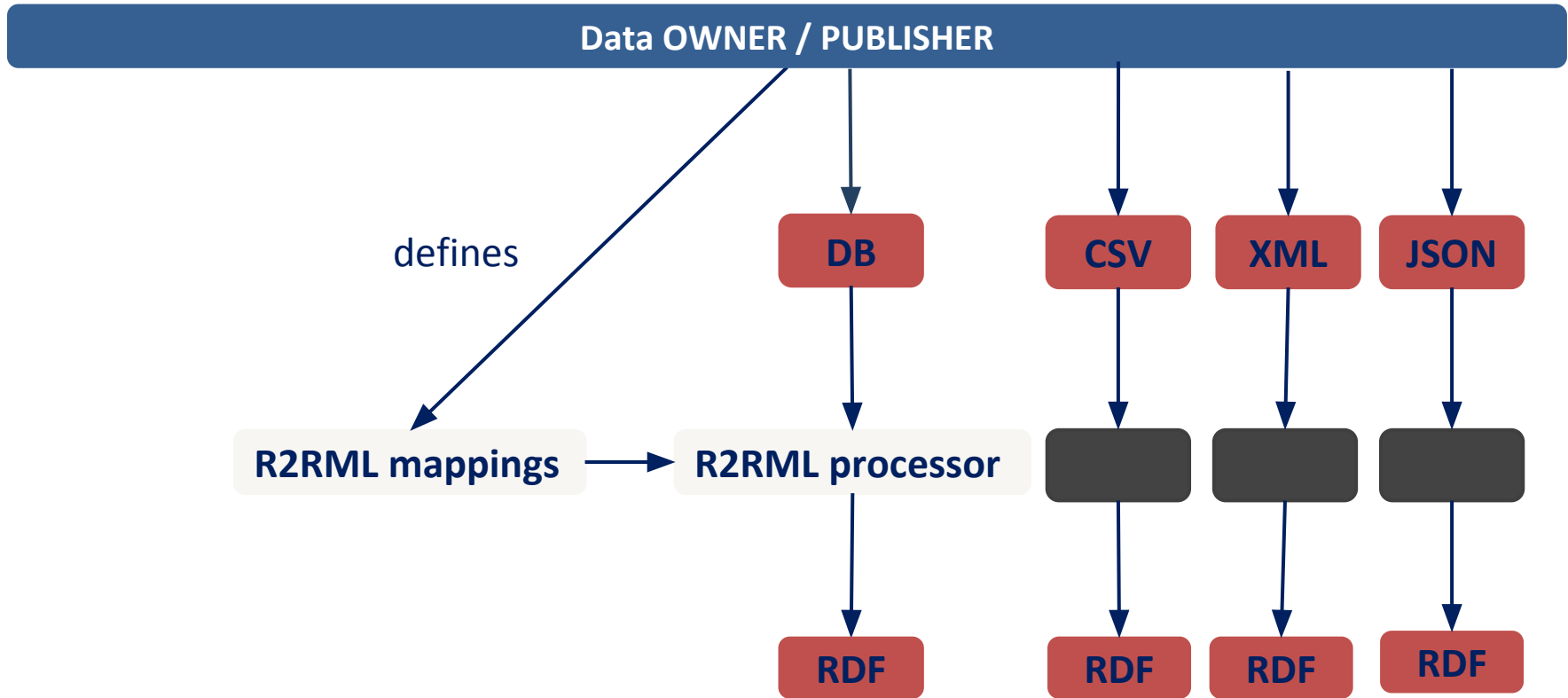
map data on ***per-source** basis*

# Relational Database to RDF Mapping Language

**R2RML**

## R2RML

W3C recommended mapping language  
for representing data in  
Relational Databases to RDF (RDB2RDF)



# **RDF Mapping Language**

**RML**

# RDF Mapping Language (RML)

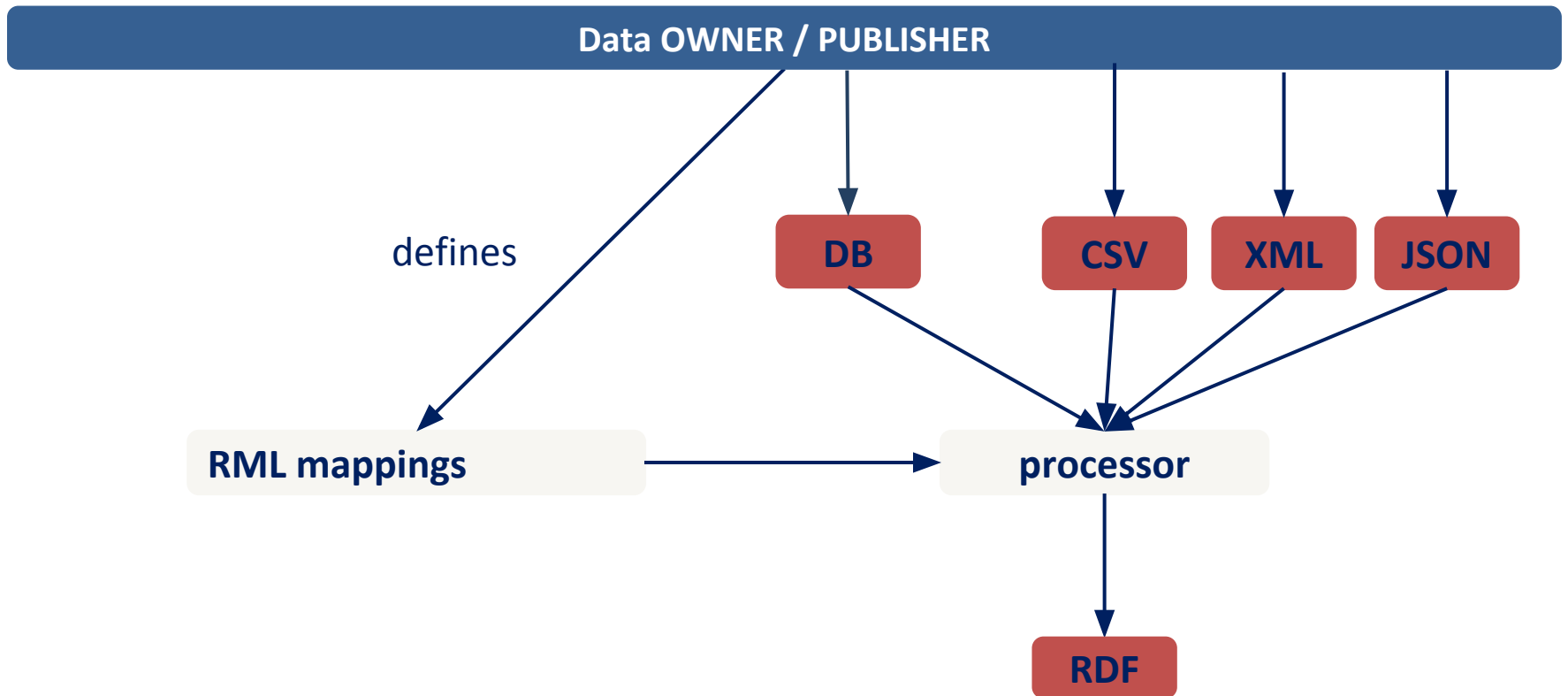


generic scalable mapping language  
for generating and interlinking  
RDF data from heterogeneous resources  
in an integrable and interoperable fashion

superset of the W3C standardized  
R2RML mapping language

<http://rml.io>

# Uniform way for integrated mapping of heterogeneous sources



# [R2]RML

Generating RDF Terms

Defining RDF Term Types

Defining Term Maps

Generating RDF Triples



# [R2]RML

## Generating RDF Terms

Defining RDF Term Types

Defining Term Maps

Generating RDF Triples

# RDF Terms

## IRIs

<http://example.com/person/1>

## Literals

“Anastasia Dimou”

## Blank Nodes

\_:7870

# Generating RDF Terms

column/reference-valued

template-valued

constant-valued

# Generating RDF Terms

column/reference-valued

template-valued

constant-valued

# Generating RDF Terms - R2RML column-valued

# Generating RDF Terms - R2RML

column-valued → default Literal

# Generating RDF Terms - R2RML

**column-valued** → default Literal

```
[] rr:column "firstname" .
```

# Generating RDF Terms - **RML**

**reference-valued** → default Literal

```
[] rml:reference "firstname" .
```



# Generating RDF Terms - RML

reference-valued → default Literal

[] rml:reference "firstname" .

id	firstname	lastname	lab	city
1	Anastasia	Dimou	IDLab	Ghent
2	Pieter	Heyvaert	IDLab	Ghent
3	Ruben	Verborgh	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

# Generating RDF Terms - RML

reference-valued → default Literal

```
[] rml:reference "firstname" .
```

"Anastasia"

"Pieter"

"Ruben"

# Generating RDF Terms

column/reference-valued

**template-valued**

constant-valued

# Generating RDF Terms template-valued

# Generating RDF Terms

template-valued → default IRI

# Generating RDF Terms

**template-valued** → default IRI

```
[] rr:template "http://example.com/person/{id}" .
```

# Generating RDF Terms

**template-valued** → default IRI

```
[] rr:template "http://example.com/person/{id}" .
```

```
<http://example.com/person/1>
```

```
<http://example.com/person/2>
```

```
<http://example.com/person/3>
```

```
<http://example.com/person/4>
```

# Generating RDF Terms

**template-valued** → default IRI

[] rr:template

“http://example.com/person/{firstname}\_{lastname}”



# Generating RDF Terms

**template-valued** → default IRI

```
[] rr:template
```

```
    "http://example.com/person/{firstname}_{lastname}"
```

```
<http://example.com/person/Anastasia_Dimou>
```

```
<http://example.com/person/Pieter_Heyvaert>
```

```
<http://example.com/person/Ruben_Verborgh>
```

```
<http://example.com/person/Ruben_Taelman>
```

# Generating RDF Terms

column/reference-valued

template-valued

**constant-valued**

# Generating RDF Terms

**constant-valued** → default IRI

```
[] rr:constant foaf:name .
```

```
[] rr:constant <http://example.com/name> .
```

foaf:name

<http://example.com/name>

# Generating RDF Terms

column/reference-valued

template-valued

constant-valued

Generating RDF Terms

**Defining RDF Term Types**

Defining Term Maps

Generating RDF Triples

## Generating RDF Terms

reference-valued → Literal

template-valued → IRI

constant-valued → IRI

## Defining RDF Term Types

rr:Literal

rr:IRI

rr:BlankNode

## Generating RDF Terms

reference-valued → default Literal

<term> rml:reference “URI”

“http://example.com/”

## Defining RDF Term Types

<term> rr:termType rr:IRI → IRI

<http://example.com/>

# Generating RDF Terms

**template-valued** → default IRI

<term> rr:template “{name} {surname}”

<Anastasia Dimou>

<Pieter Heyvaert>

<Ruben Verborgh>

<Ruben Taelman>



## Generating RDF Terms

**template-valued** → default IRI

<term> rr:template "{name} {surname}"

<Anastasia Dimou>

## Defining RDF Term Types

<term> rr:termType rr:Literal → Literal

"Anastasia Dimou"

# Generating RDF Terms

## Defining RDF Term Types

```
[ ] rr:termType rr:BlankNode → BlankNode
```

## Generating RDF Terms

reference-valued → Literal

template-valued → IRI

constant-valued → IRI

## Defining RDF Term Types

rr:Literal

rr:IRI

rr:BlankNode

# [R2]RML

Generating RDF Terms

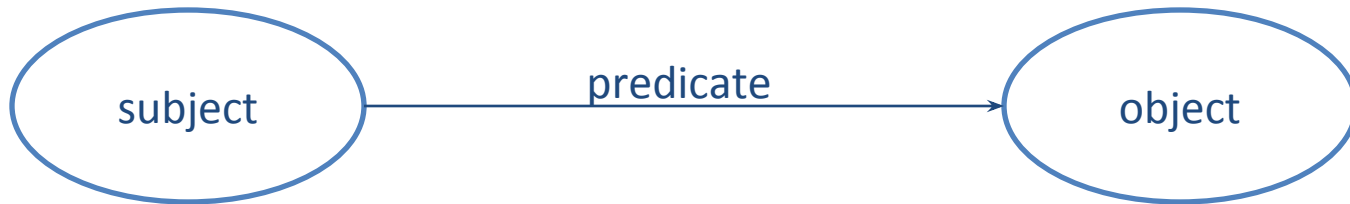
Defining RDF Term Types

**Defining Term Maps**

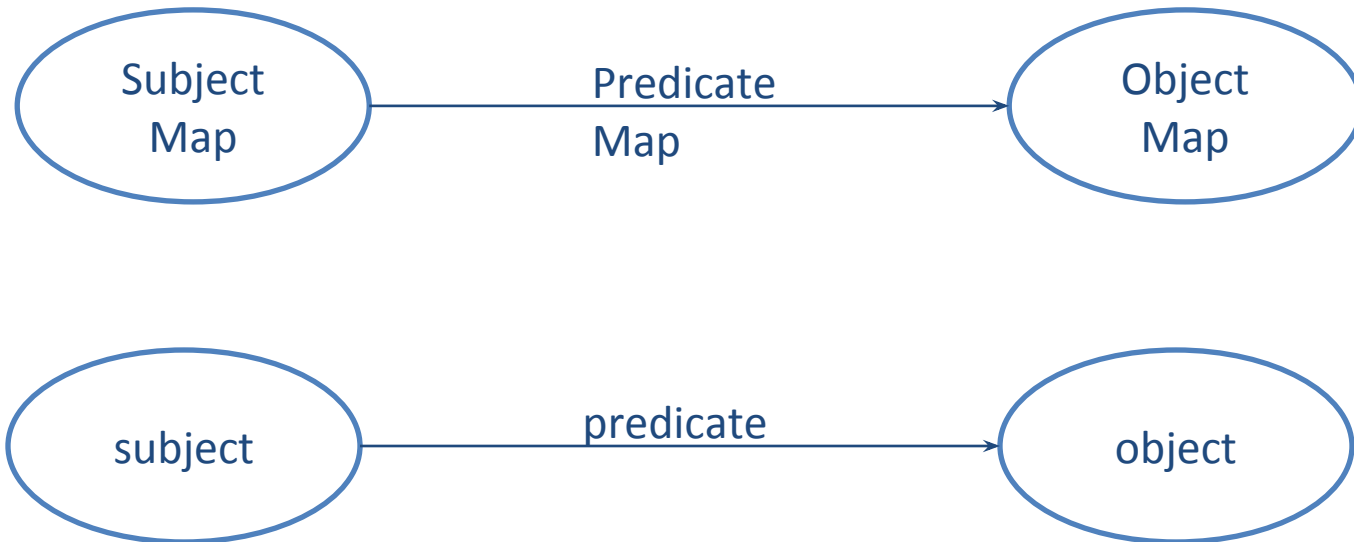
Generating RDF Triples

# Resource Description Framework (RDF)

is the prevalent data model  
for describing **Linked Open Data**



# [R2]RML



# Defining RDF Term Maps

## Subject Map

<SM> rr:template "http://example.com/{id}".

## Predicate Map

<PM> rr:constant rdfs:label.

## Object Map

<OM> rml:reference "firstname".

## Graph Map

# [R2]RML

Generating RDF Terms

Defining RDF Term Types

Defining Term Maps

Generating **RDF Triples**



# Generating RDF Triples

## Subject Map

<SM> rr:template “http://example.com/{id}”.

## Predicate Map

<PM> rr:constant rdfs:label.

## Object Map

<OM> rml:reference “firstname”.

## Generating **RDF Triples**

<TM> rr:subjectMap <SM>.

<SM> rr:template "http://example.com/{id}".

<TM> rr:predicateObjectMap <POM>.

<PM> rr:constant rdfs:label.

<POM> rr:predicateMap <PM>.

<OM> rml:reference "firstname".

<POM> rr:objectMap <OM>.

## Generating **RDF Triples**

<TM> rr:subjectMap <SM>.

<TM> rr:predicateObjectMap <POM>.

<POM> rr:predicateMap <PM>.

<POM> rr:objectMap **<OM>**..

# Generating RDF Triples

<SM> rr:template “http://example.com/{id}”.

<PM> rr:constant rdfs:label.

<OM> rml:reference “firstname”.

# Generating RDF Triples

<SM> rr:template “http://example.com/lab/{lab}”.

<PM> rr:constant schema:location.

<OM> rr:template “http://example.com/city/{city}”.

# Generating RDF Triples

<SM\_1> rr:template “http://example.com/lab/{lab}”.

<PM> rr:constant schema:location.

<SM\_2> rr:template “http://example.com/city/{city}”.

## Generating RDF Triples

<TM\_1> rr:subjectMap <SM\_1>.

<SM\_1> rr:template "http://example.com/lab/{lab}".

<TM\_1> rr:predicateObjectMap <POM>.

<PM> rr:constant schema:location.

<TM\_2> rr:subjectMap <SM\_2>.

<SM\_2> rr:template "http://example.com/city/{city}".

<POM> rr:objectMap

[ rr:parentTriplesMap <TM\_2>].

# Generating RDF Triples

<SM\_1> rr:template “http://example.com/lab/{lab}”.

<PM> rr:constant schema:location.

<SM\_2> rr:template “http://example.com/city/{city}”.



# Generating RDF Triples

<b>id</b>	<b>firstname</b>	<b>lastname</b>	<b>lab</b>	<b>city</b>
1	Anastasia	Dimou	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

<b>cityId</b>	<b>cityName</b>	<b>country</b>
A	Brussels	Belgium
B	Ghent	Belgium

## Generating RDF Triples

<TM\_1> rr:subjectMap <SM\_1>.

<SM\_1> rr:template "http://example.com/lab/{lab}".

<TM\_1> rr:predicateObjectMap <POM>.

<PM> rr:constant schema:location.

<TM\_2> rr:subjectMap <SM\_2>.

<SM\_2> rr:template "http://other.com/city/{cityId}".

<POM> rr:objectMap

[ rr:parentTriplesMap <TM\_2>;  
rr:joinCondition <JC> ].

# Generating RDF Triples

<b>id</b>	<b>firstname</b>	<b>lastname</b>	<b>lab</b>	<b>city</b>
1	Anastasia	Dimou	IDLab	Ghent
4	Ruben	Taelman	IDLab	Ghent

<b>cityId</b>	<b>cityName</b>	<b>country</b>
A	Brussels	Belgium
B	Ghent	Belgium

**rr:joinCondition [**  
**rr:parent “city” ; rr:child “cityName”]**

# [R2]RML

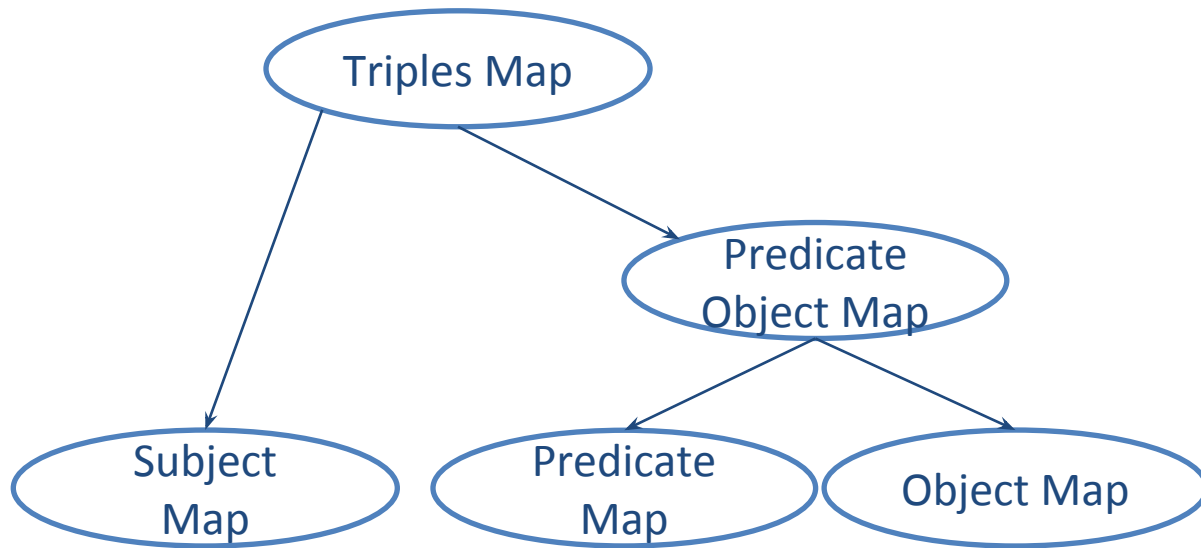
Generating RDF Terms

Defining RDF Term Types

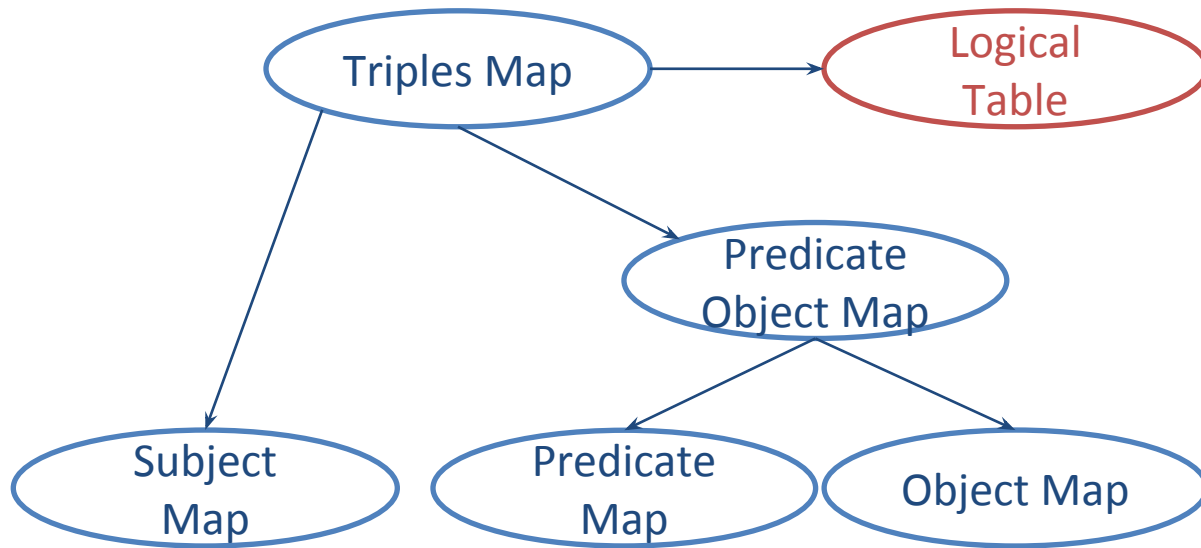
Defining Term Maps

Generating RDF Triples

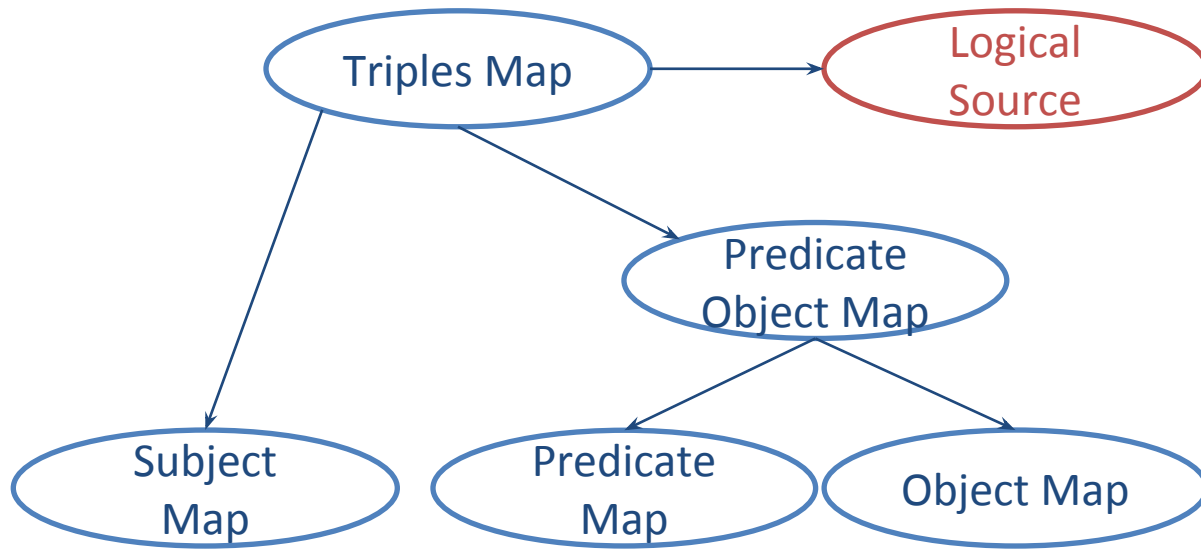
# [R2]RML



# R2RML



# RML



## Specifying data

which data form a data input

how to reference data input extracts

## Accessing & Retrieving data

data input from original source(s)



# Support data in Heterogeneous Structures and Formats

## *tabular-structured*

tables in DBs or CSV files ...

## *hierarchical-structured*

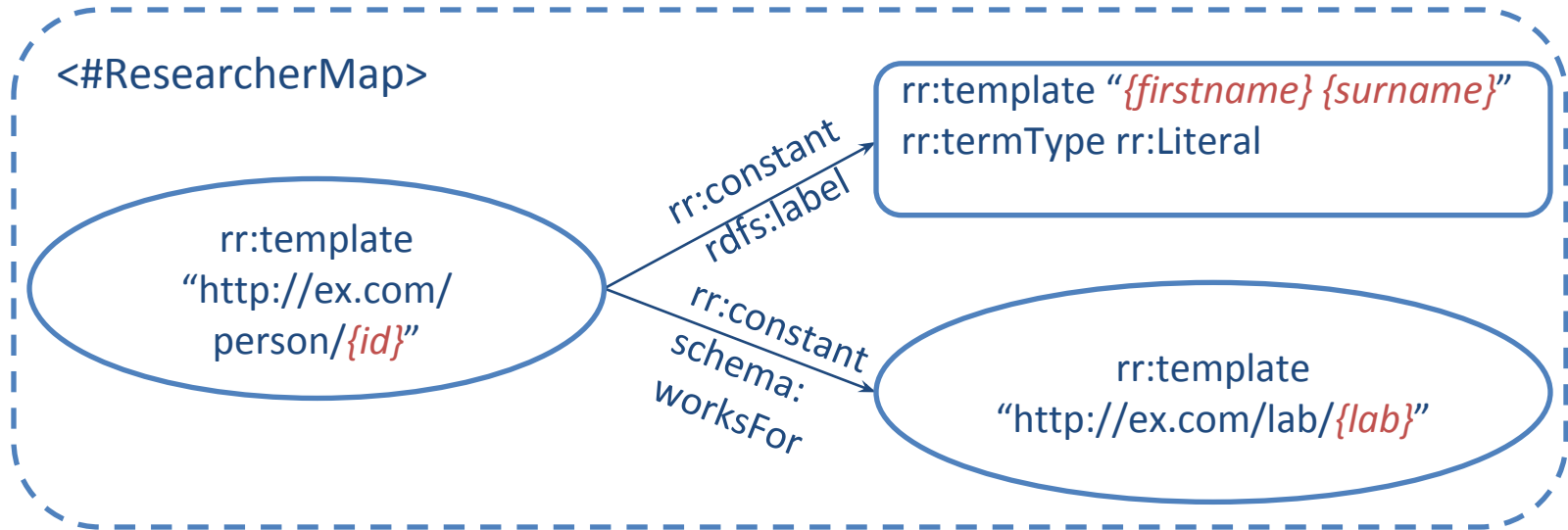
JSON or XML ...

## *(semi-)structured*

HTML ...

....

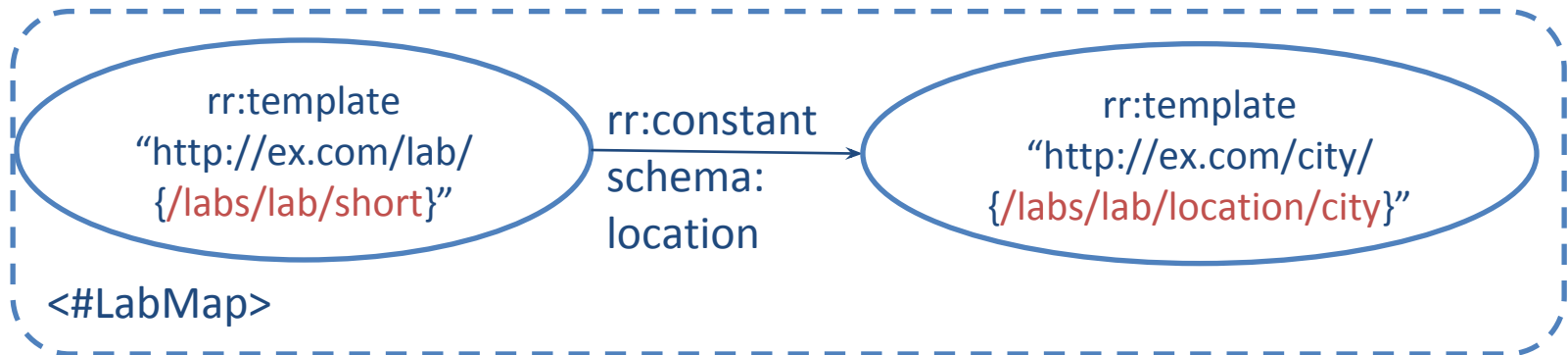
# support tabular-structured data



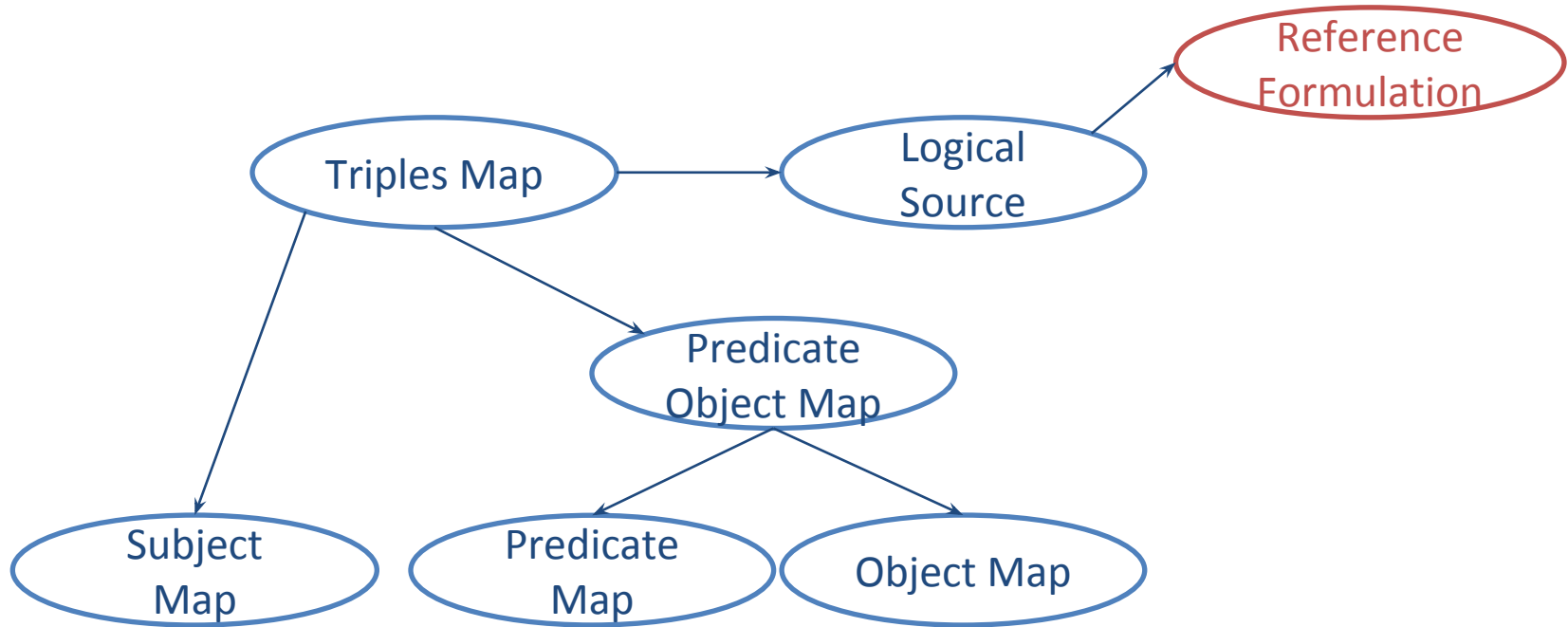
id	firstname	surname	lab
1	Anastasia	Dimou	IDLab
2	Pieter	Heyvaert	IDLab
3	Ruben	Verborgh	IDLab
4	Ruben	Taelman	IDLab

# support hierarchical-structured data

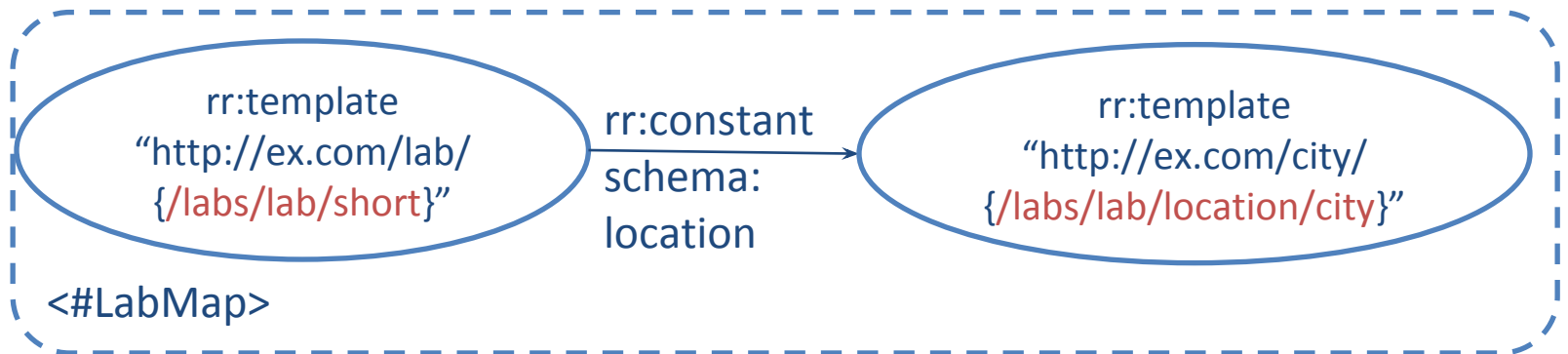
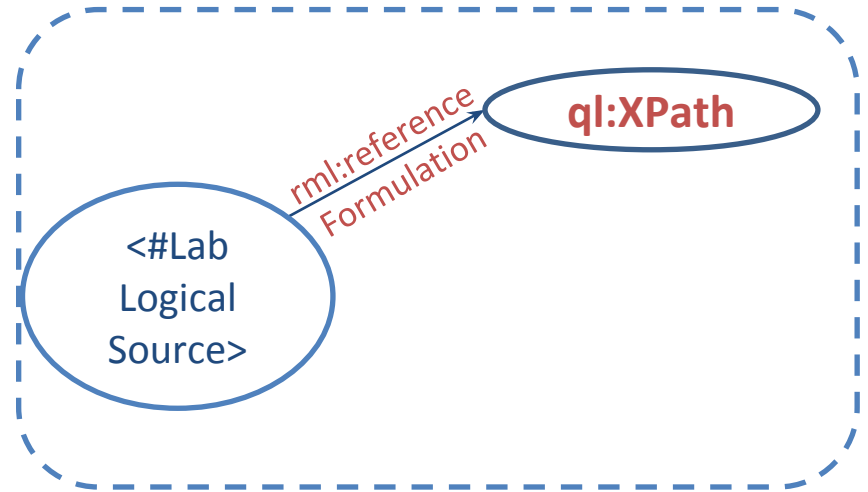
```
<labs>  
  <lab>  
    <short>IDLab</short>  
    <location>  
      <city>Ghent</city>  
    </location>  
  </lab>  
  <lab> .... </lab>  
  ...  
</labs>
```



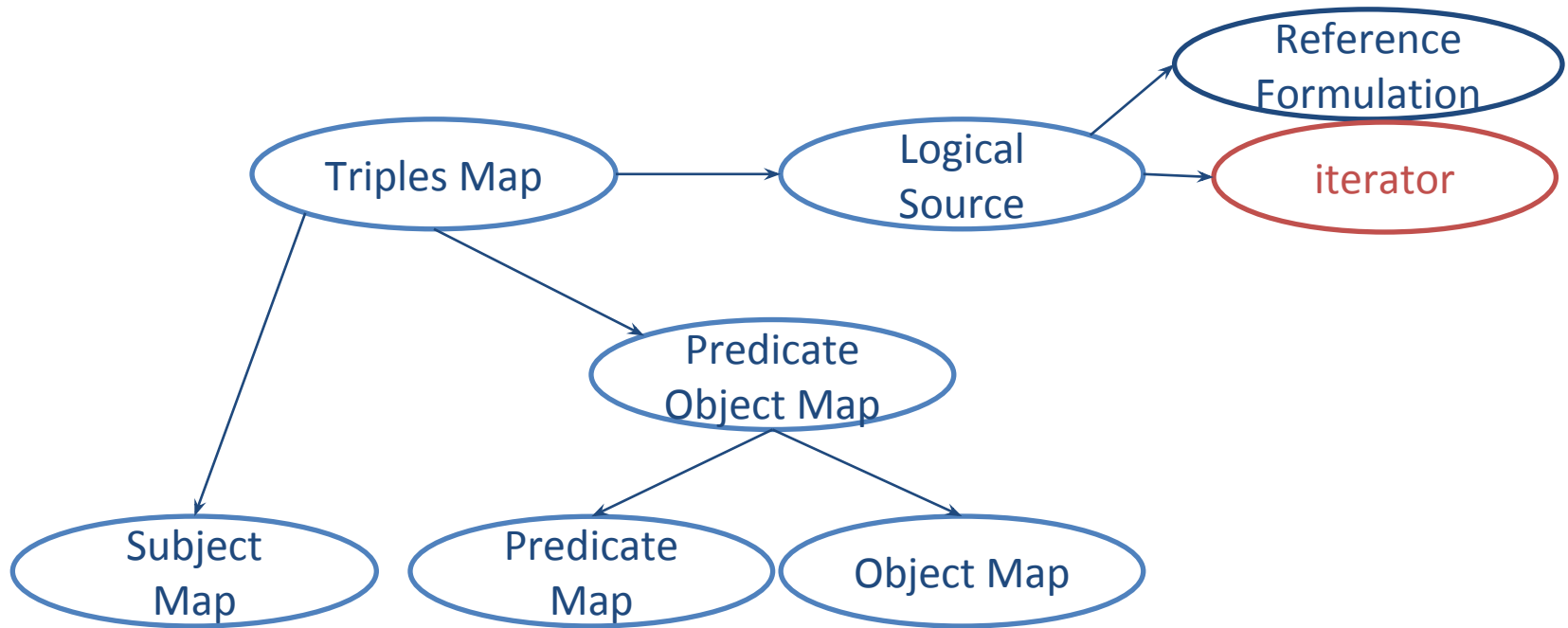
# RDF Mapping Language (RML)



```
<labs>
  <lab>
    <short>IDLab</short>
    <location>
      <city>Ghent</city>
    </location>
  </lab>
  <lab> .... </lab>
  ...
</labs>
```



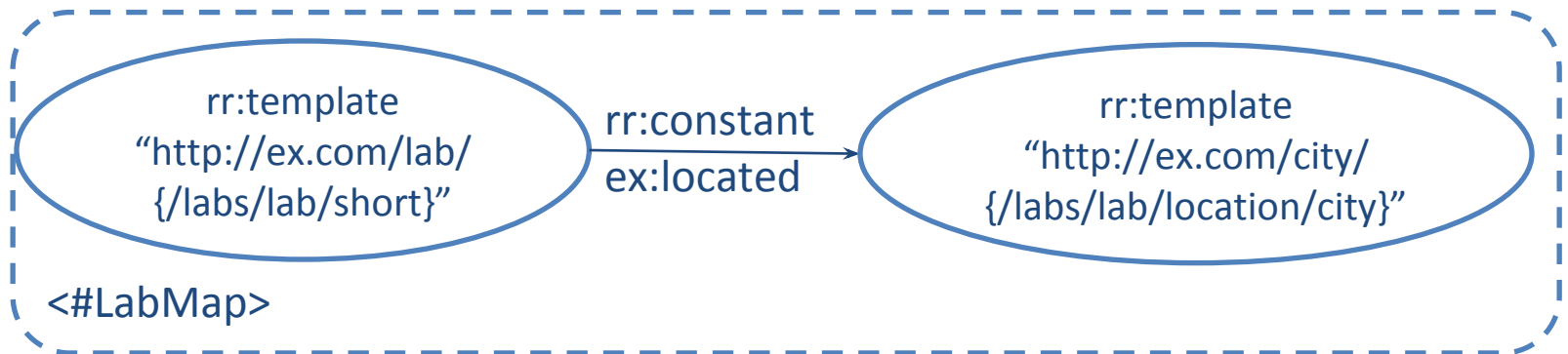
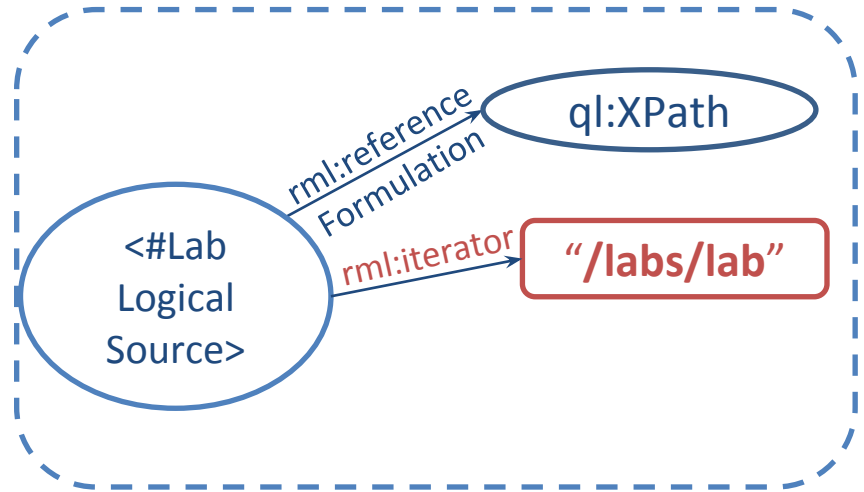
# RDF Mapping Language (RML)



```

<labs>
  <lab>
    <short>IDLab</short>
    <location>
      <city>Ghent</city>
    </location>
  </lab>
  <lab> .... </lab>
  ...
</labs>

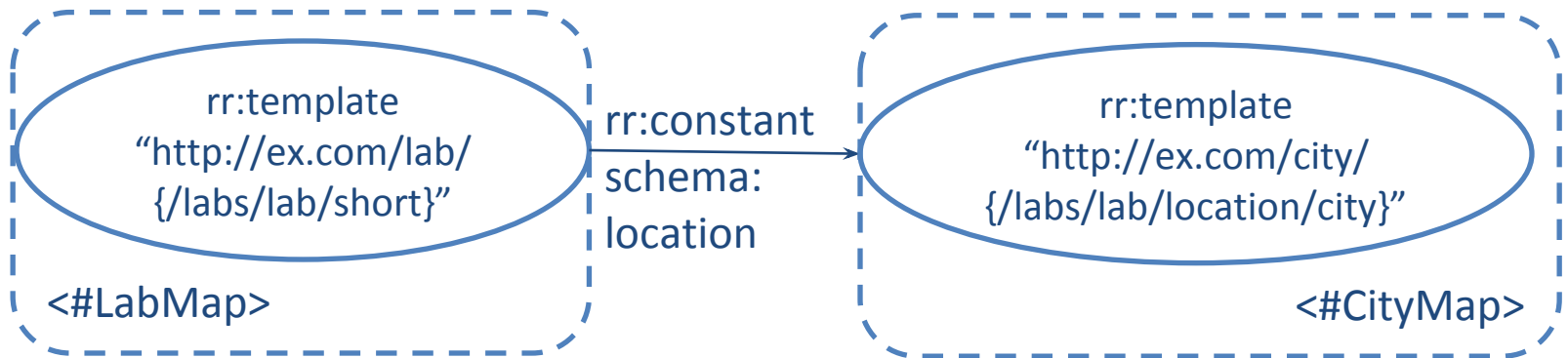
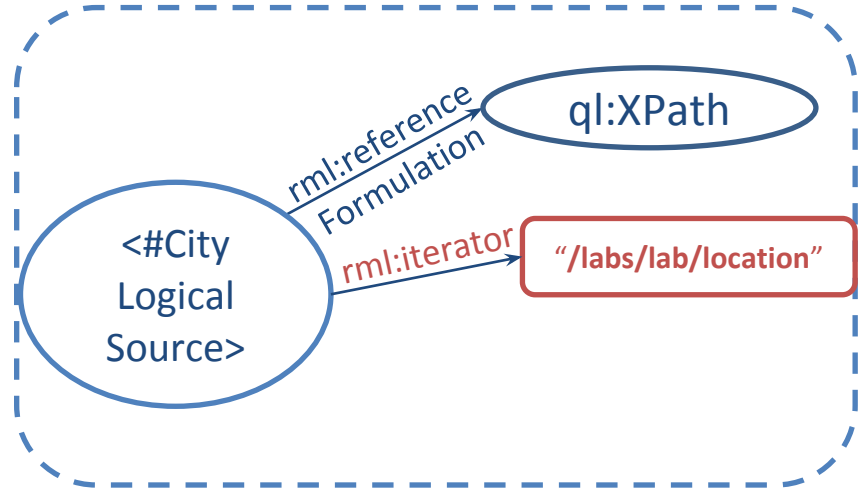
```



```

<labs>
  <lab>
    <short>IDLab</short>
    <location>
      <city>Ghent</city>
    </location>
  </lab>
  <lab> .... </lab>
  ...
</labs>

```





## Specifying data

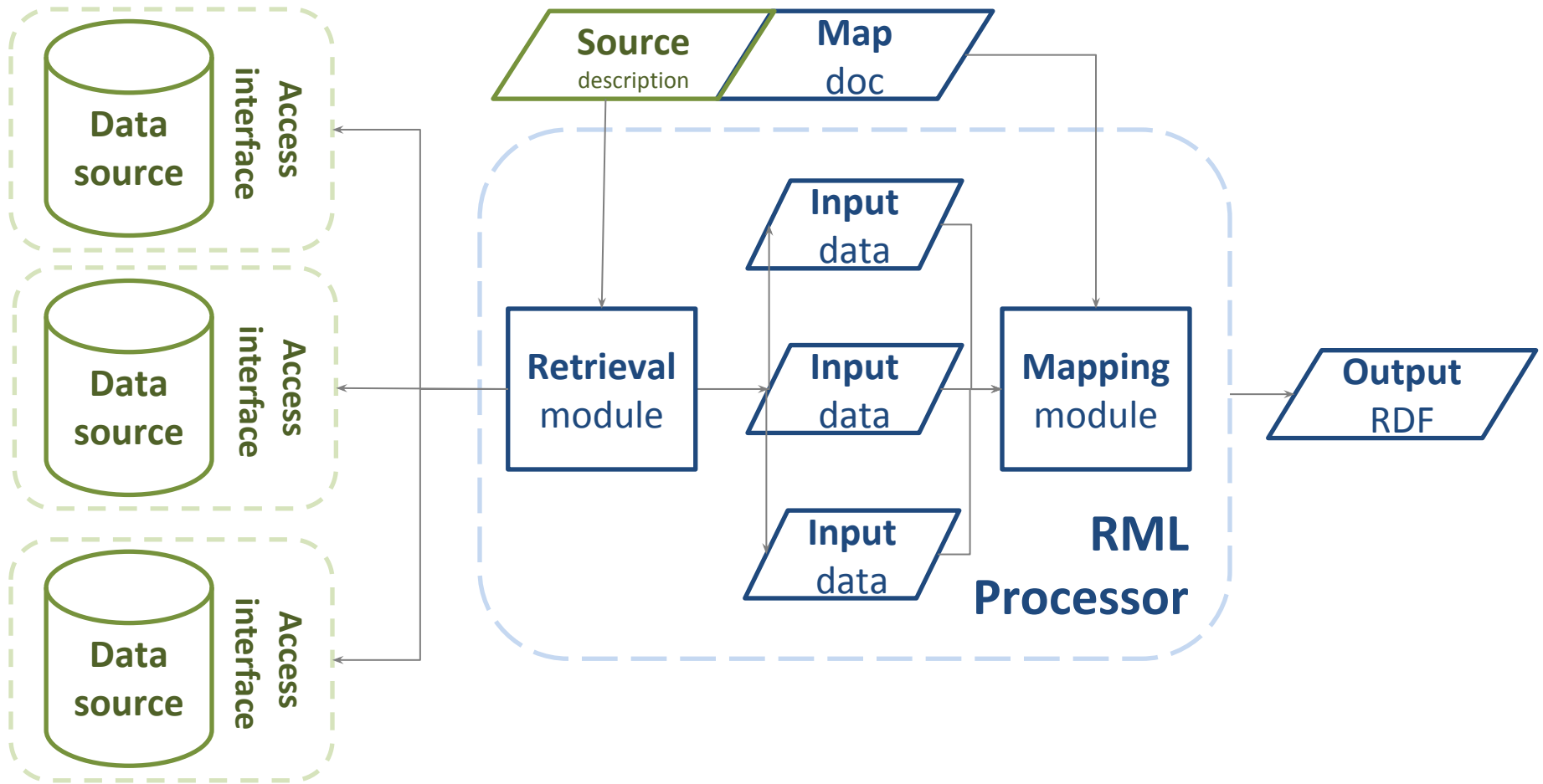
which data form a data input

how to reference data input extracts

## Accessing & Retrieving data

data input from original source(s)

# Where does this data originally come from?



# Support different Locations and Access Interfaces

Local File(s)

Database connectivity

Web source(s)

RDF source(s)

# Dataset and Service Vocabularies

advertising in machine-interpretable fashion

how to access the underlying data

## Dataset and Service Vocabularies

advertising in machine-interpretable fashion

how to access the underlying data

can also be used in combination with RML

to retrieve the data input to be mapped

from its original source

# Support different Locations and Access Interfaces

Local File(s)

Database connectivity

D2RQ

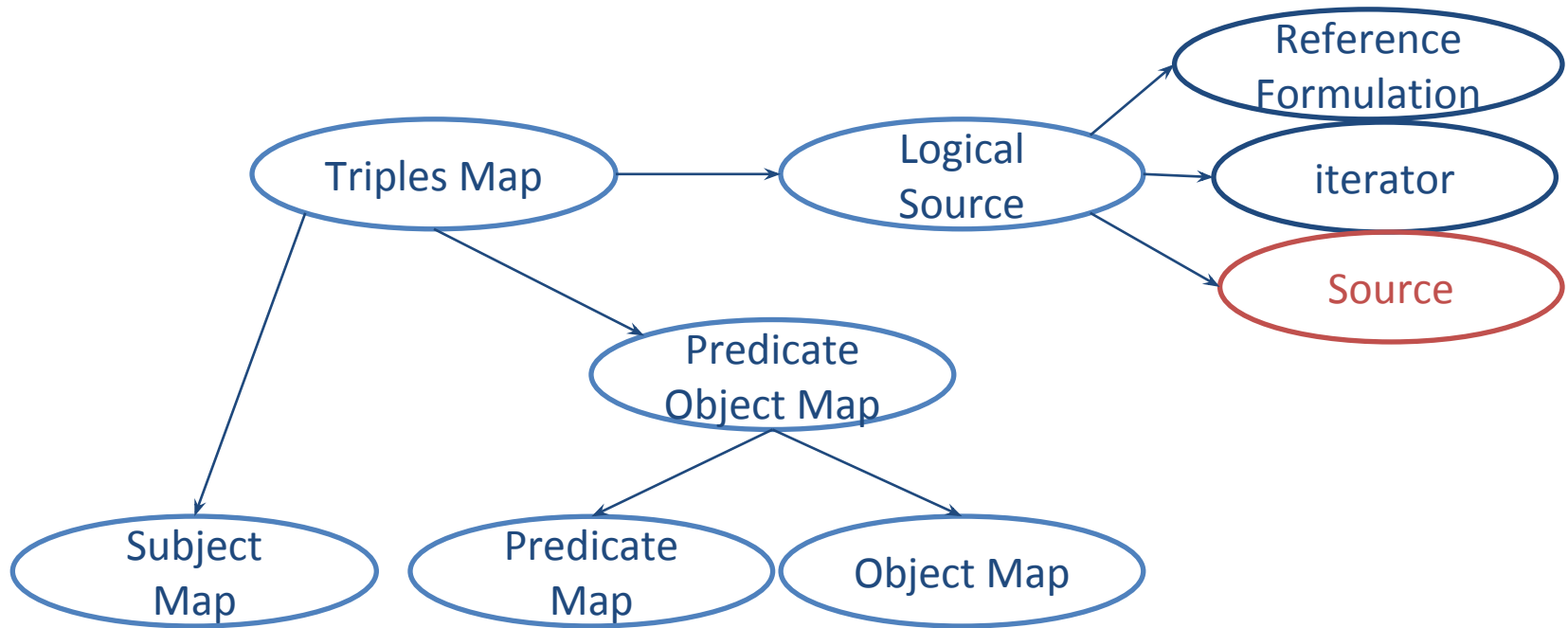
Web source(s) (Web API/service)

DCAT, CSVW, Hydra, VOiD (Dataset)

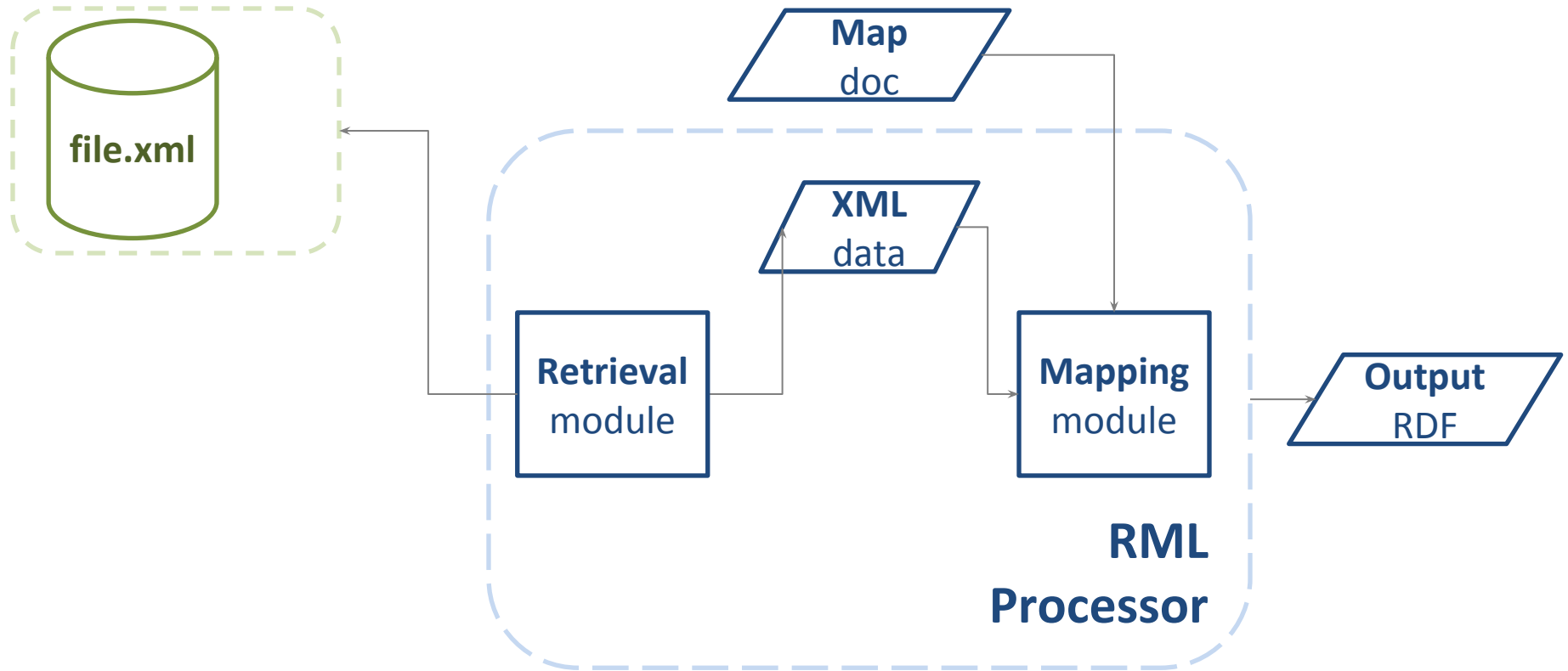
RDF source(s)

VOiD (Endpoint), SPARQL-SD

# RDF Mapping Language (RML)



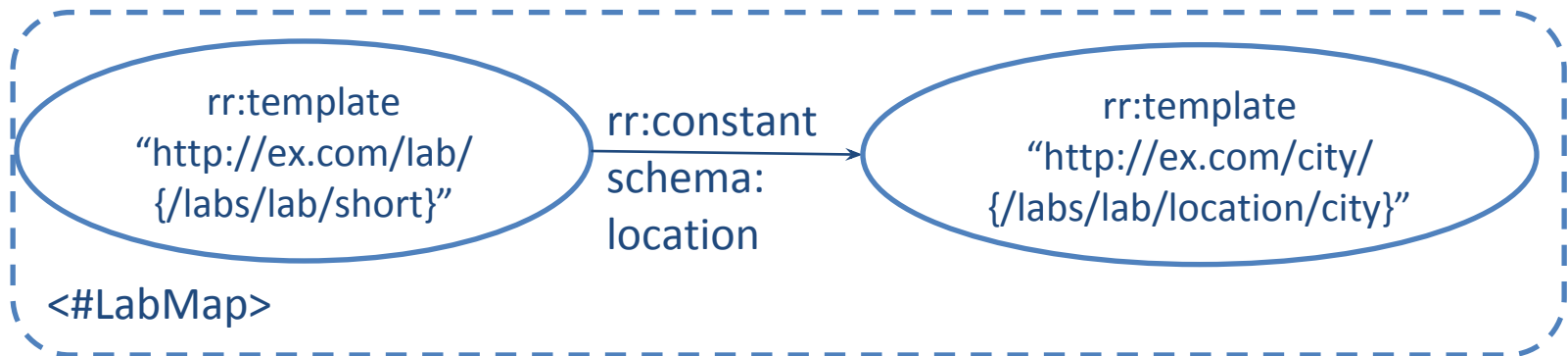
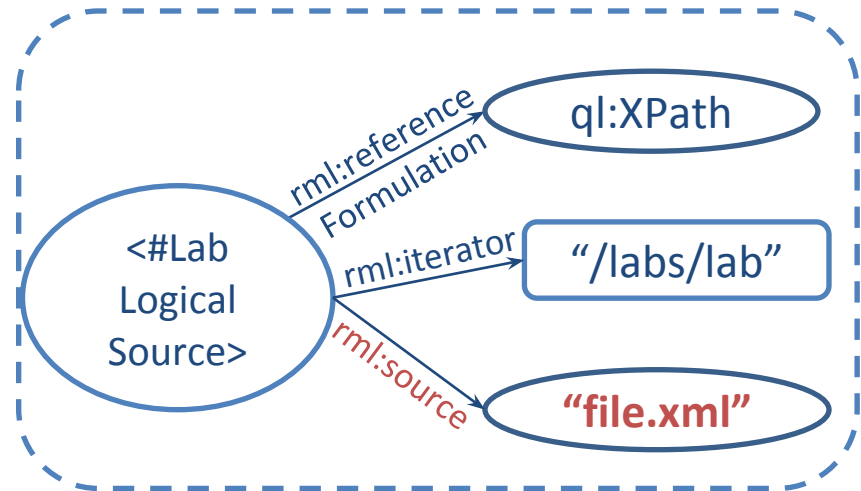
# Support Local File(s)



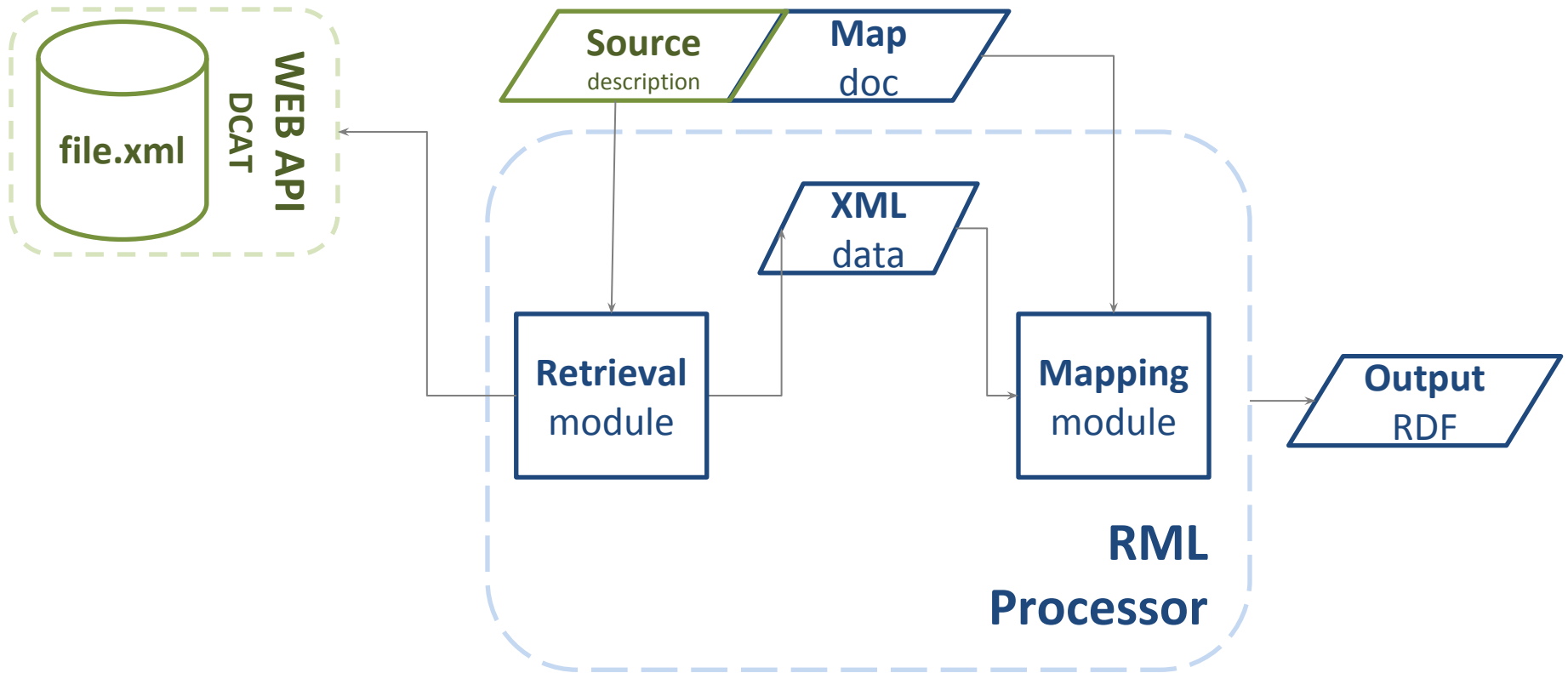


# Support Local File(s)

```
<labs>
  <lab>
    <short>IDLab</short>
    <location>
      <city>Ghent</city>
    </location>
  </lab>
  <lab> .... </lab>
  ...
</labs>
```

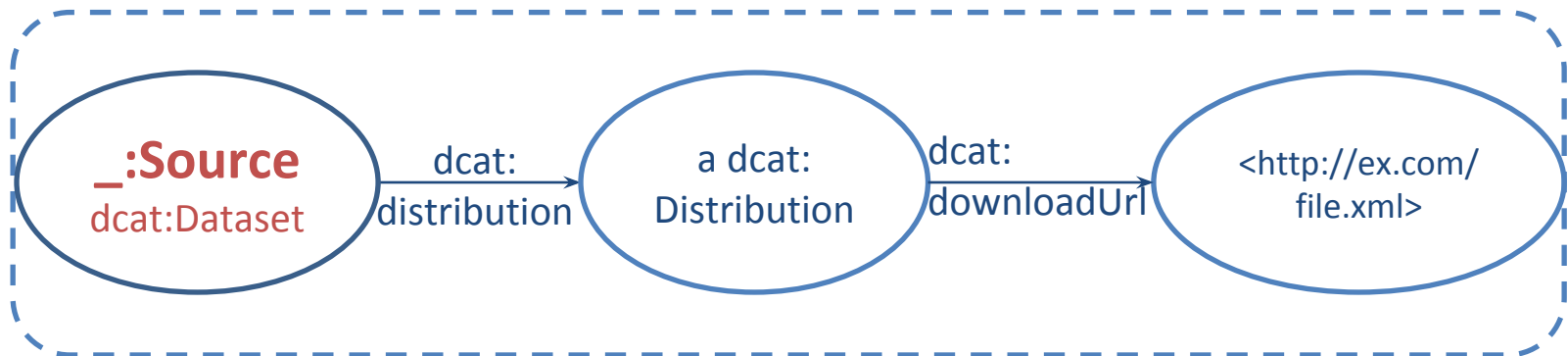
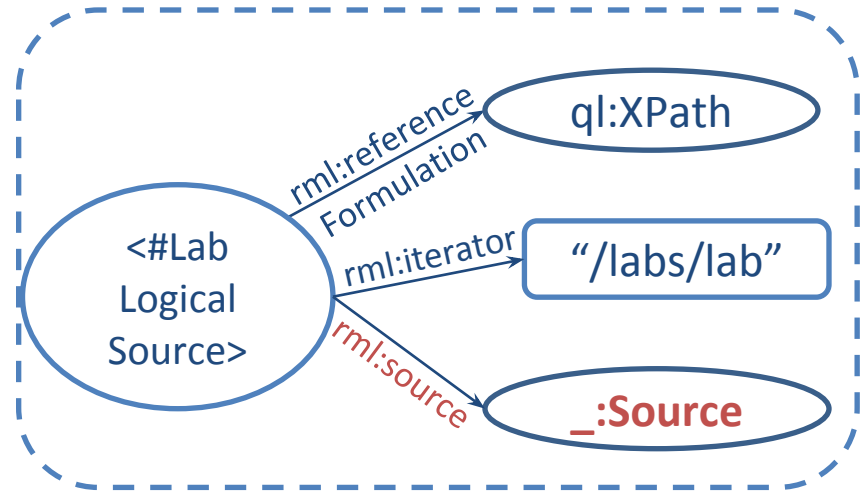


# Support file(s) published on the Web

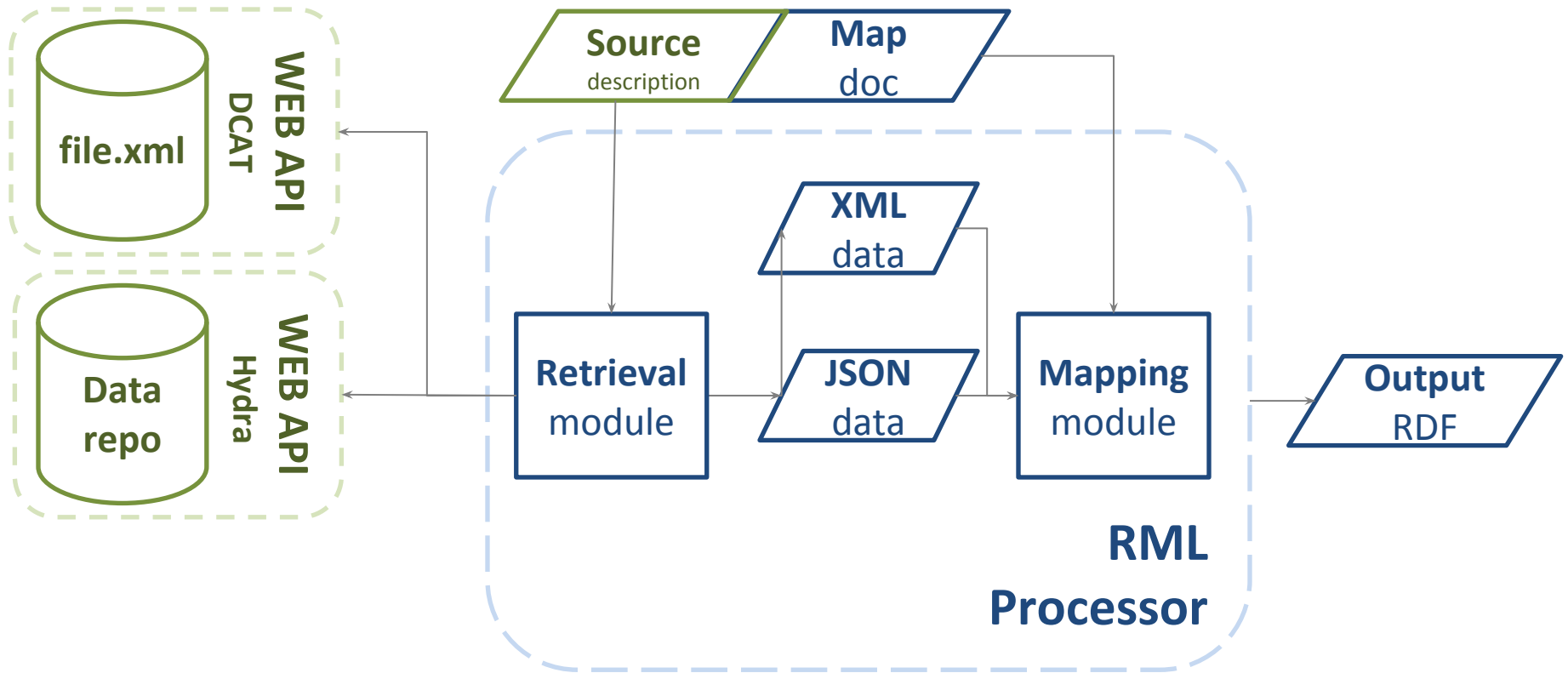


# Support dataset on the Web (DCAT)

```
<labs>
  <lab>
    <short>IDLab</short>
    <location>
      <city>Ghent</city>
    </location>
  </lab>
  <lab> .... </lab>
  ...
</labs>
```

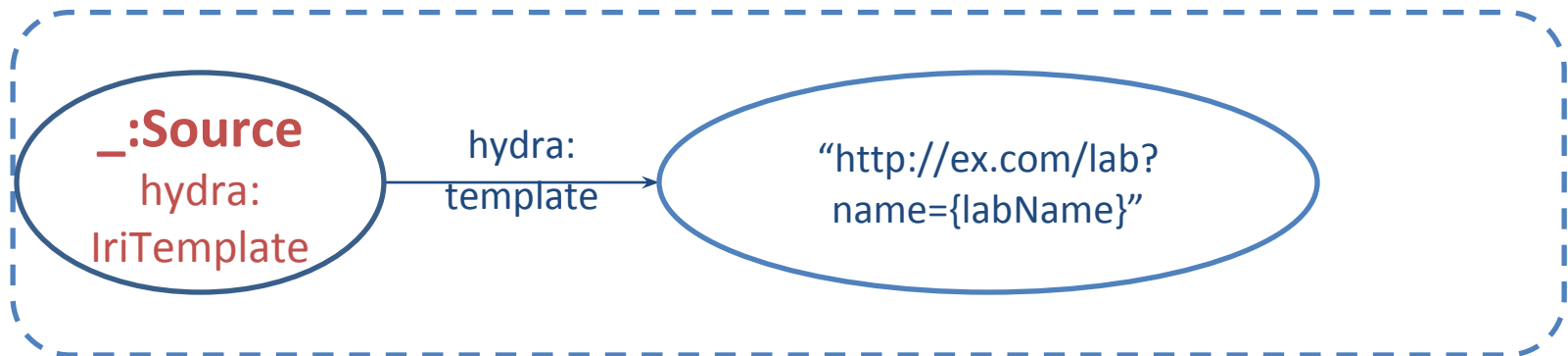
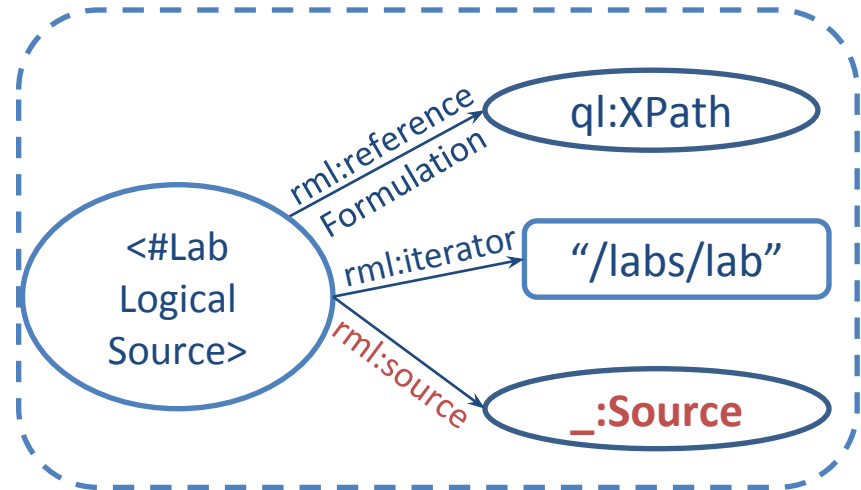


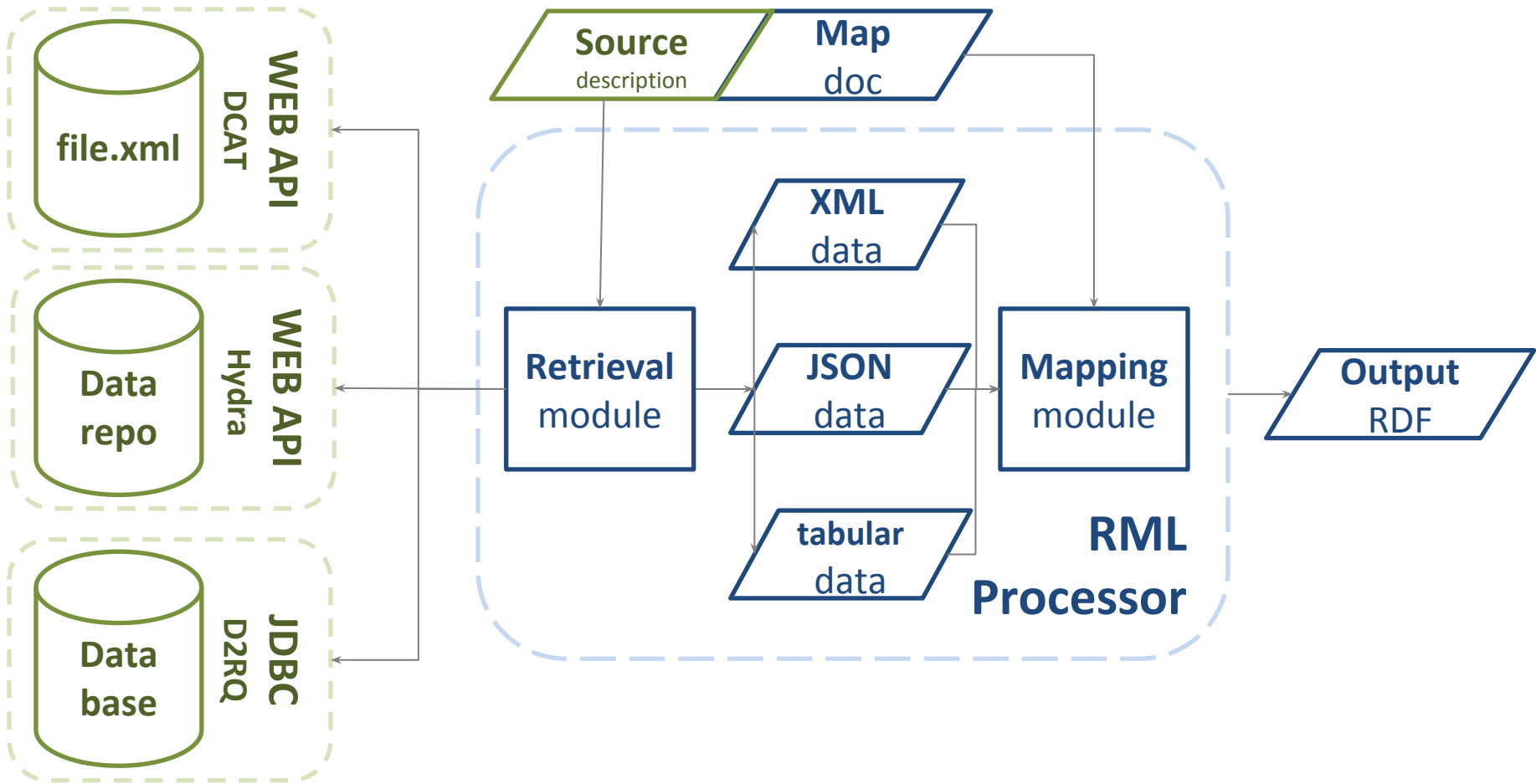
# Support data derived from a Web API



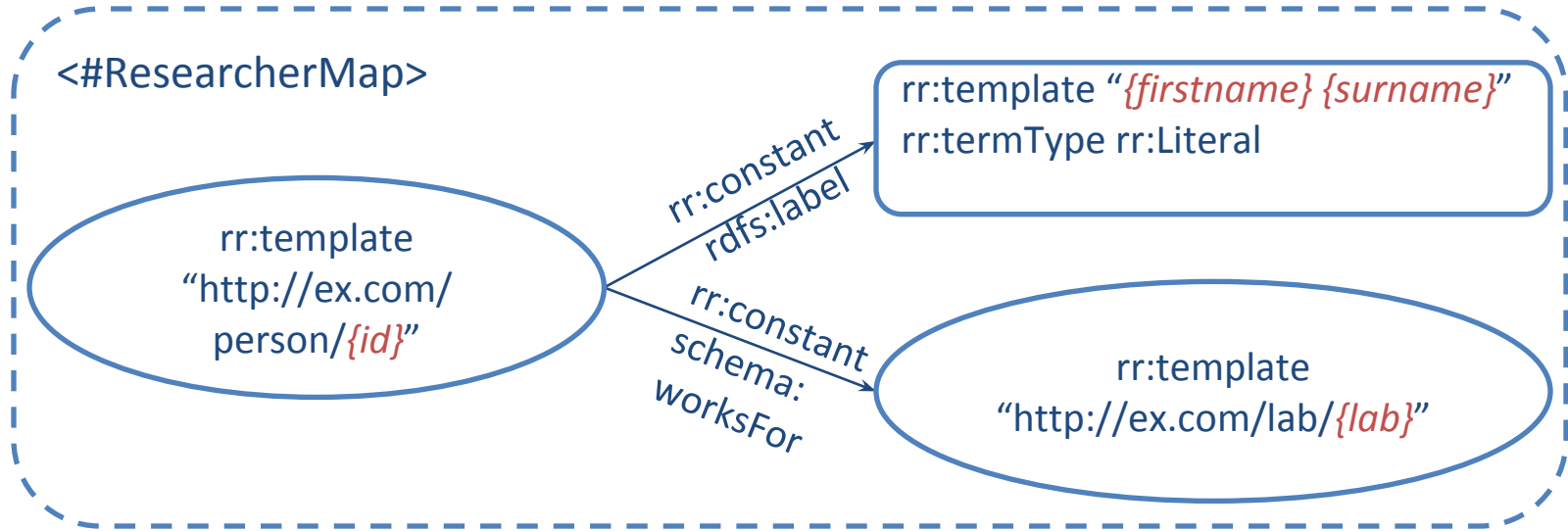
# Support data from a Web API (Hydra)

```
<labs>
  <lab>
    <short>IDLab</short>
    <location>
      <city>Ghent</city>
    </location>
  </lab>
  <lab> .... </lab>
  ...
</labs>
```

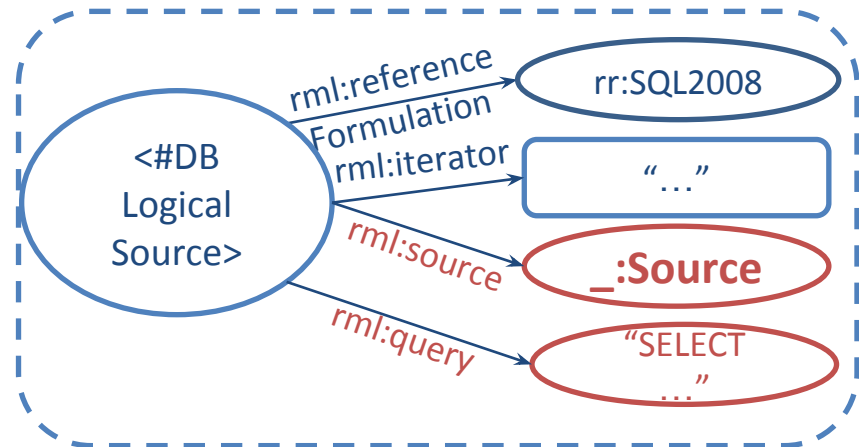




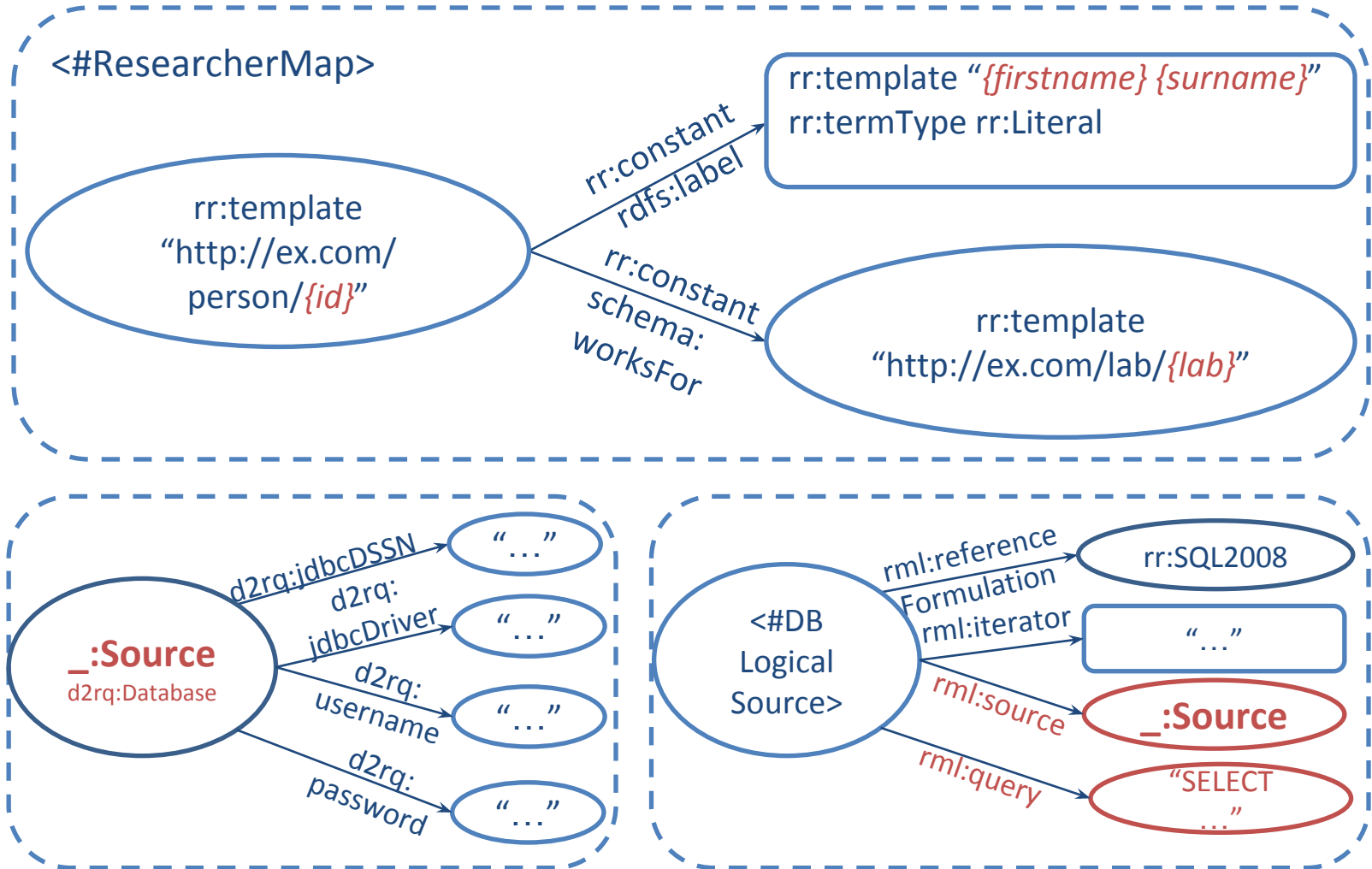
# Support tabular-structured data



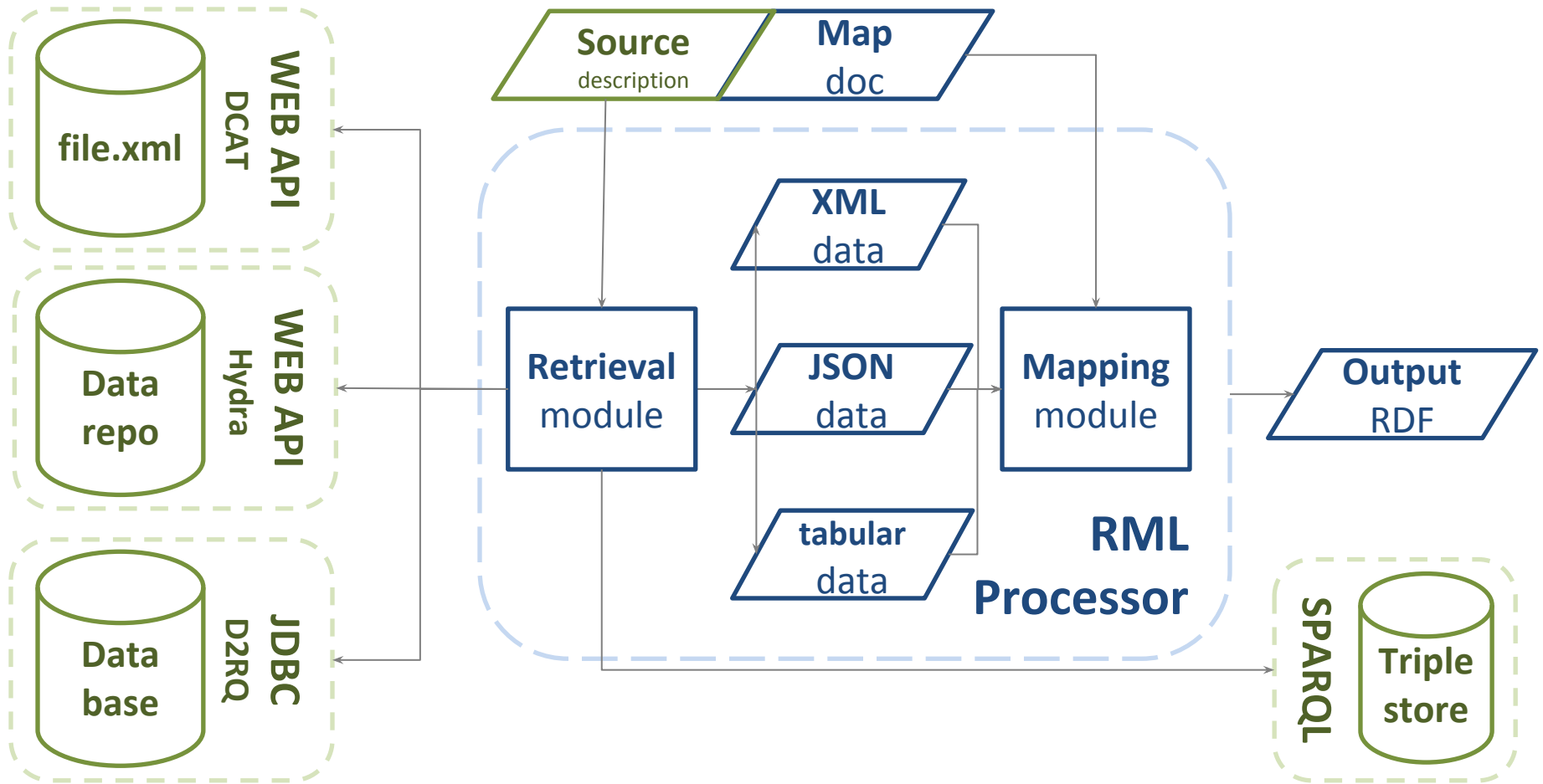
id	firstname	surname	lab
1	Anastasia	Dimou	IDLab
2	Pieter	Heyvaert	IDLab
2	Ruben	Verborgh	IDLab
3	Ruben	Taelman	IDLab

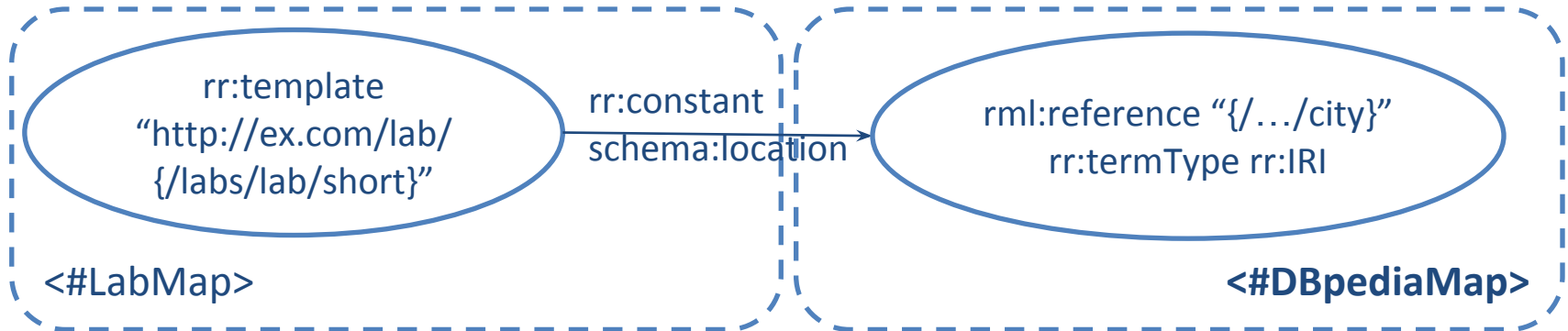
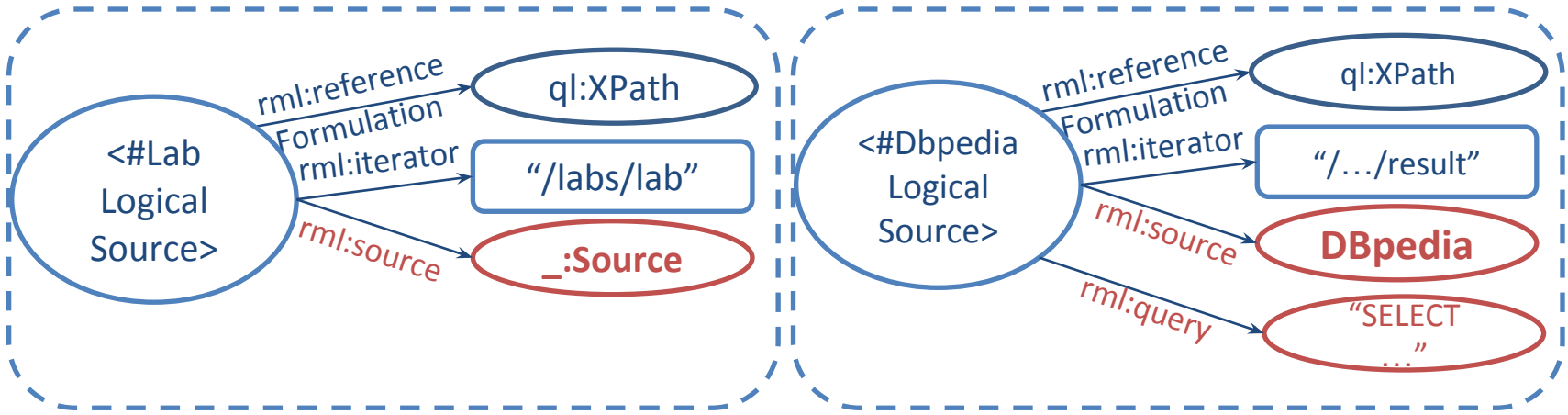


# Support tabular-structured data









## Specifying data

which data form a data input

how to reference data input extracts

## Accessing & Retrieving data

data input from original source(s)

R2RML, <https://www.w3.org/TR/r2rml/>

RML, <http://rml.io/>

# Integrated Linked Data Generation of Heterogeneous raw Data

