Integrated mapping of heterogeneous sources to RDF with RML

Anastasia Dimou,
Miel Vander Sande, Ruben Verborgh, Erik Mannens
Ghent University – iMinds – Multimedia Lab

http://rml.io

iMinds

UNIVERSITEIT GENT

MULTIMEDIA LAB

Semantic Web enabled applications
rely on data represented as
Linked Open Data
semantically annotated
using **ontologies and vocabularies**

BUT most data exists in
formats other than RDF

Many languages, tools and approaches were proposed

to map data from different sources to their RDF representation

Existing mapping solutions map

per-format* and *per-source

→ *focus more on handling the source rather than modeling the domain*

OR

*provide **case-specific** solutions*

→ *better model the domain*

BUT NOT many
languages, tools and approaches

model a knowledge domain in RDF
relying on
multiple sources of heterogeneous formats
that complementarily describe
this knowledge domain

The mappings are…

- independently defined disregarding
  - possible prior definitions
  - links to other resources
  - (re)using same ontologies for similar data
- manually aligned/interlinked
  - by reconstructing the same URIs
  - by post-mapping interlinking
  - by matching ontologies

Mappings involve

reusing, combining and extending
existing ontologies and vocabularies

OR

dynamically, randomly and on-demand
creating **new** ontologies and vocabularies

A well-considered policy is required
when mapping data to RDF
in the context of a certain knowledge domain

that shifts the focus
from modeling the data of a source
to modeling the knowledge domain

RDF Mapping Language (RML)

generic and scalable language
for mapping heterogeneous resources into RDF
in an integrable and interoperable fashion

superset of the W3C standardized
R2RML mapping language

http://rml.io

consider available data sources
that represent a knowledge domain
reusing/generating an ontology/vocabulary
that can formally represent this domain

identify data sources required to
more accurately and complete
represent this knowledge domain
considering an ontology/vocabulary

uniform mapping definitions

to describe mapping rules for heterogeneous sources

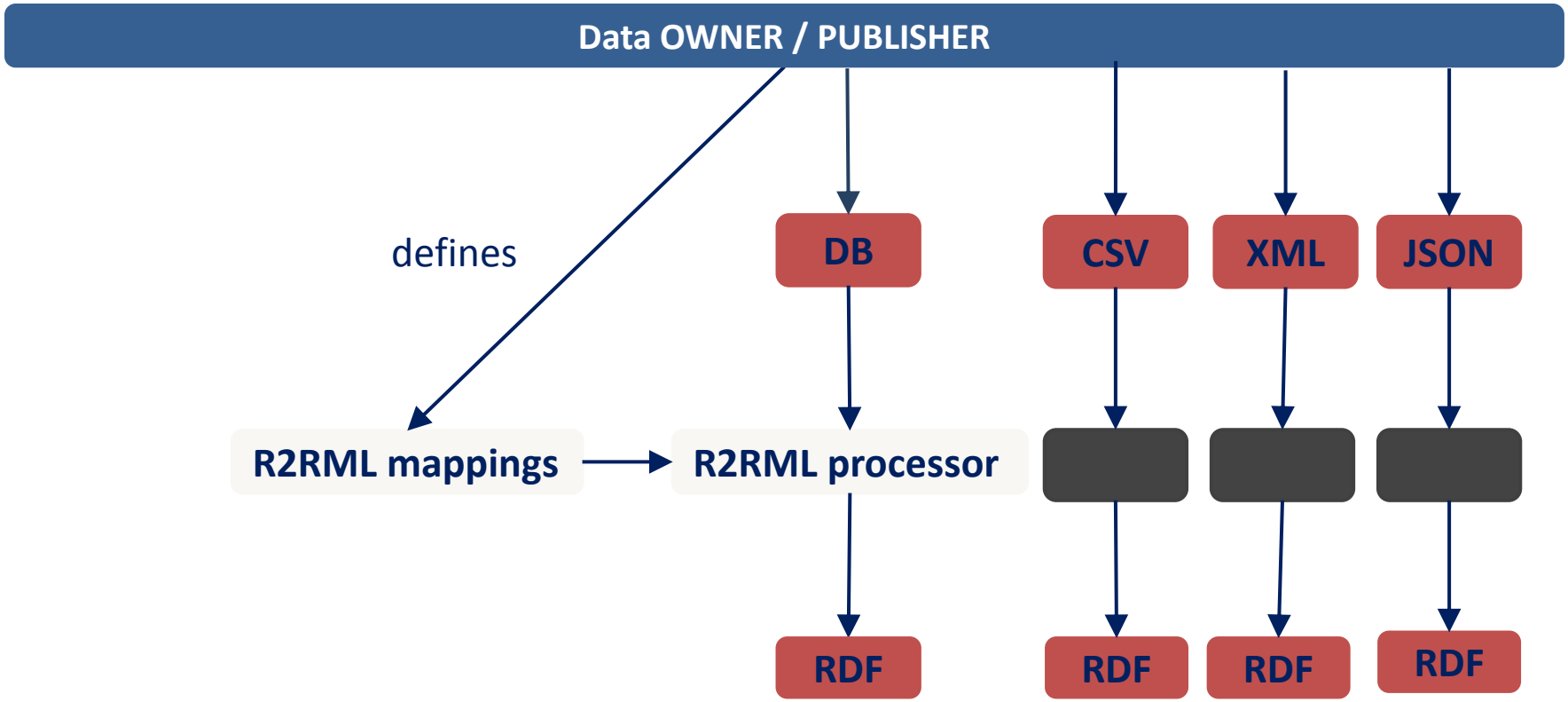
interoperable mapping definitions

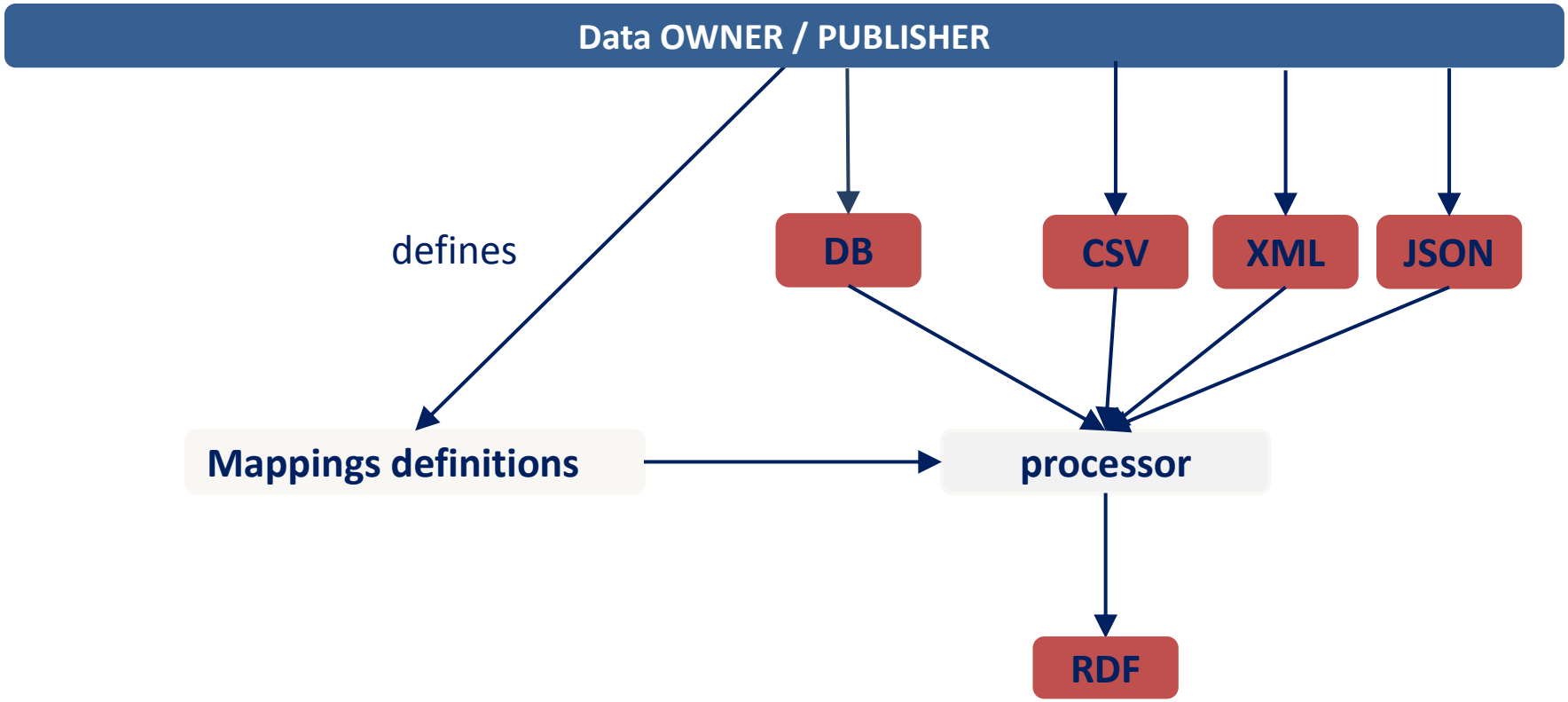
that would allow the re-use of mapping rules
across different implementations

reusable mapping definitions

that would allow the re-use of mapping rules
for representing data in the same or different formats

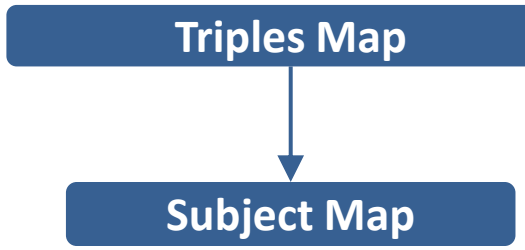
→reflects on the usage of
(combinations of) ontologies/vocabularies


Data OWNER / PUBLISHER
defines
DB
CSV
XML
JSON
R2RML mappings
R2RML processor
RDF
RDF
RDF
RDF


Data OWNER / PUBLISHER
defines
DB
CSV
XML
JSON
Mappings definitions
processor
RDF

defining an RML subject

NAME	BIRTH_DATE	DEATH_DATE
Robert Theodore McCall	1919-12-23	2010-02-26
Ronald Anderson	1929-12-06	



<#ArtistMapping>

rr:subjectMap [
rr:template “http://ex.com/{NAME}” ;
rr:class me:Person];

<http://ex.com/Robert+Theodore+McCall> a me:Person

defining an RML predicate & object

NAME	BIRTH_DATE	DEATH_DATE
Robert Theodore McCall	1919-12-23	2010-02-26
Ronald Anderson	1929-12-06	

Triples Map

Predicate Object Map

Predicate Map

Objectt Map

```
<#ArtistMapping>
    rr:predicateObjectMap [
    rr:predicate me:birth_date;
    rr:objectMap [
        rr:column "BIRTH_DATE" ]  ];
```

<http://ex.com/Robert+Theodore+McCall> me:birth_date “1919-12-23”

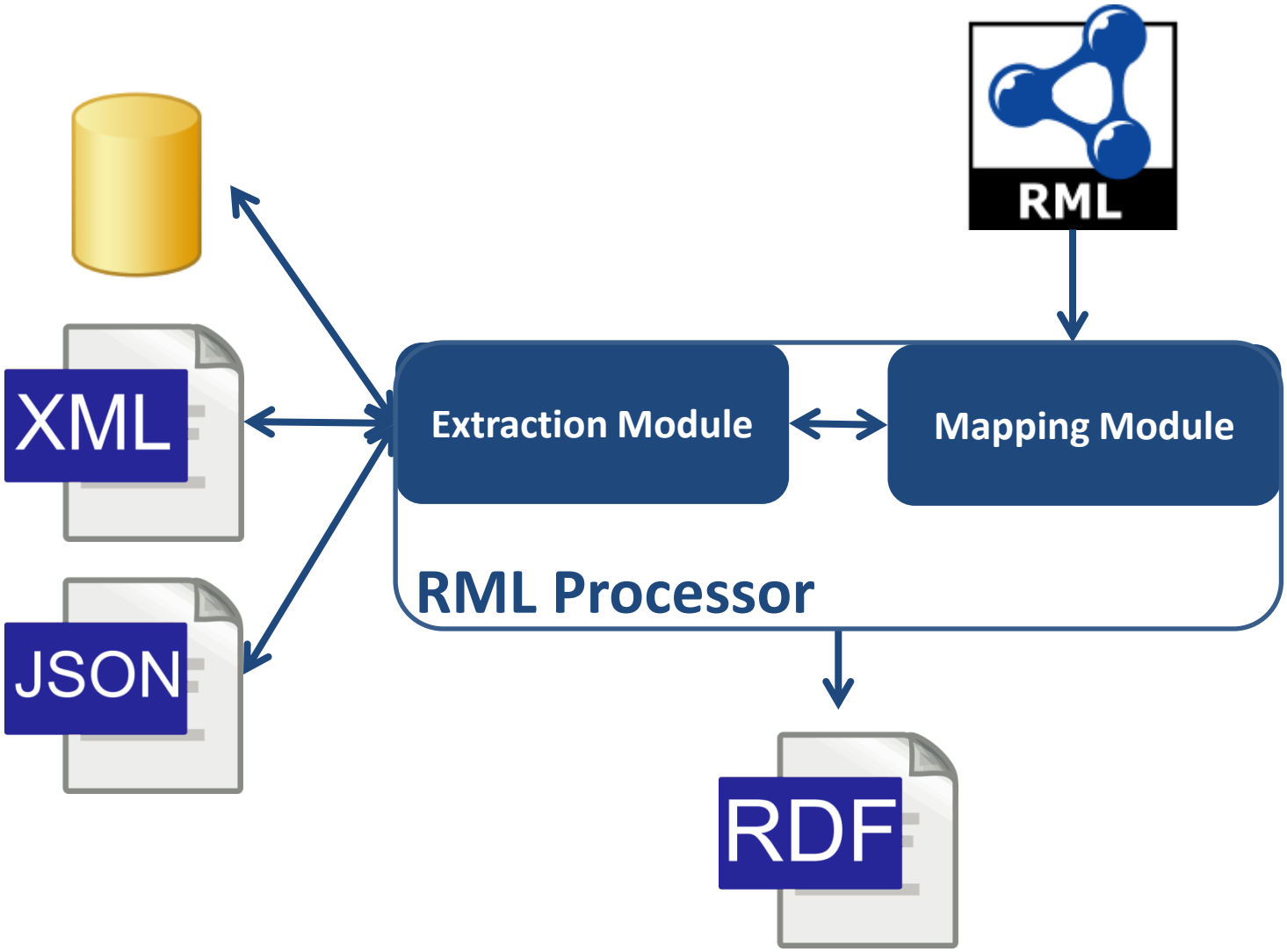
```
<span class="CEURVOLNR">
Vol-1056</span>

<h3>
<span class="CEURLOCTIME">
Montpellier, France, May 26, 2013
</span>.
</h3>
```

```
<#CEURMapping>
  rr:subjectMap [
    rr:template
"http://ex.com/{span.CEURVOLNR}/" ];

  rr:predicateObjectMap [
    rr:predicate bibo:presentedAt;
    rr:objectMap [
          rml:reference "span. CEURLOCTIME";
          rml:process “^([\s\w]*),”
      ] ] ] .
```

ex:Vol-1056/ bibo:presentedAt **“Montpellier”**


RML
XML
JSON
Extraction Module
Mapping Module
RML Processor
RDF

Enables Linked Open Data, required to realize the envisaged Semantic Web.

rml.io

Anastasia.Dimou@ugent.be **@natadimou**

iMinds

UNIVERSITEIT GENT

MULTIMEDIA LAB