RMLEditor: A Graph-Based Mapping Editor for Linked Data Mappings

Pieter Heyvaert

Anastasia Dimou, Aron-Levi Herregodts, Ruben Verborgh
Dimitri Schuurman, Erik Mannens and Rik Van de Walle

UNIVERSITEIT GENT

iMinds

How can I map existing data to Linked Data?

There are GUI tools for that.

Can non-Semantic Web experts work with these tools?

Well, if they are willing to read
the full specification
of the underlying technologies...

Our RMLEditor brings the editing of mappings to non-Semantic Web experts.

Overview

mapping process to generate Linked Data

data owners vs. Semantic Web experts during this process

existing tools help the mapping process

RMLEditor

- features
- user validation

Overview

mapping process to generate Linked Data

data owners vs. Semantic Web experts during this process

existing tools help the mapping process

RMLEditor

- features
- user validation

Knowledge is required to create mappings

Need to know what the data is about

know how the data is related to each other

Need to know how the mapping language works

mapping languages offer a declarative way to generate Linked Data

based on existing data

set of rules

improvement over custom scripts/tools

examples

- R2RML
- RML

Need to know the schemas

vocabularies

ontologies

to annotate the data

Focus on data and mapping language

Overview

mapping process to generate Linked Data

data owners vs. Semantic Web experts during this process

existing tools help the mapping process

RMLEditor

- features
- user validation

Data owners know their data

Data owners don’t know about mapping languages

data owners have limited background in the Semantic Web

need to learn the approach of the mapping language

need to learn the syntax of the mapping language

writing the mapping statements is error-prone

```
<#Triples_Mapping>
    rml:logicalSource <#DATA_EdgeInput> ;

    rr:subjectMap [
          rr:template "http://ex.com/{@node}" ;
          rr:class foaf:Person ] ;

    rr:predicateObjectMap [
          rr:predicate foaf:name;
          rr:objectMap [
                rr:parentTriplesMap <#LS_Mapping> ;
                rr:joinCondition [
                      rr:child "@title";
                      rr:parent "@id" ] ] ] ;
...
```

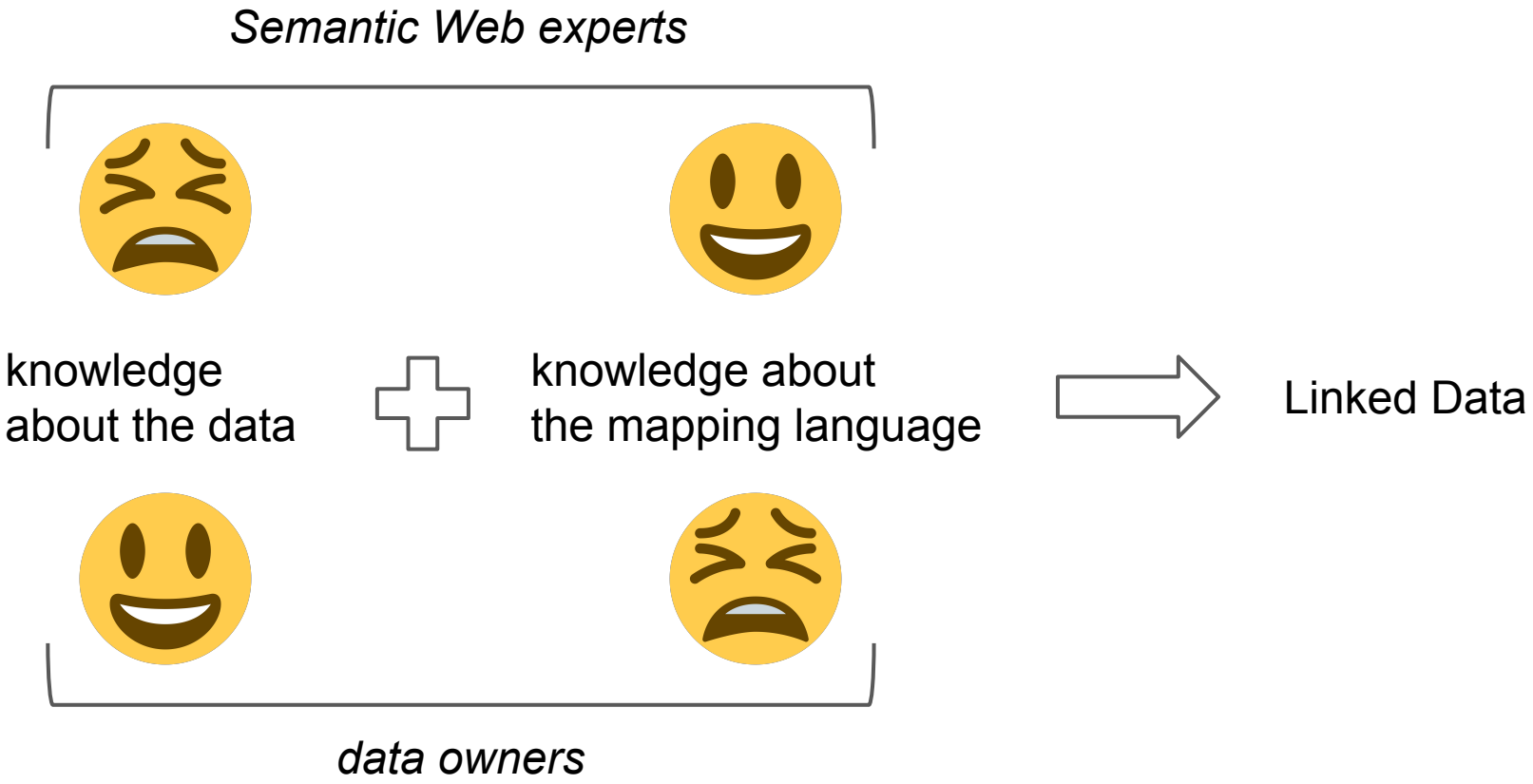
Data owners don’t want to use mapping languages

Semantic Web experts generate Linked Data

hired by the data owners

have knowledge of the mapping language

They don’t have knowledge about the data


Semantic Web experts
knowledge about the data
knowledge about the mapping language
Linked Data
data owners

Overview

mapping process to generate Linked Data

data owners vs. Semantic Web experts during this process

existing tools help the mapping process

RMLEditor

- features
- user validation

Tools to help the mapping process

knowledge about the data  ~~knowledge about the mapping language~~  Linked Data

Tools to help the mapping process

knowledge about the data  **use mapping tool**  Linked Data

GUIs are developed on mapping languages

no syntax errors

Better
but knowledge of the language is still required

subjectMap

parentTriplesMap

predicateObjectMap

Can you understand these terms without reading the full specification?

Tools require less knowledge of the language

knowledge about the data

use mapping tool

Linked Data

Goal:
tools require no knowledge of the language

knowledge about the data  **use mapping tool**  Linked Data

Overview

mapping process to generate Linked Data

data owners vs. Semantic Web experts during this process

existing tools help the mapping process

RMLEditor

- features
- user validation

Used by multiple projects and companies

including

Overview

mapping process to generate Linked Data

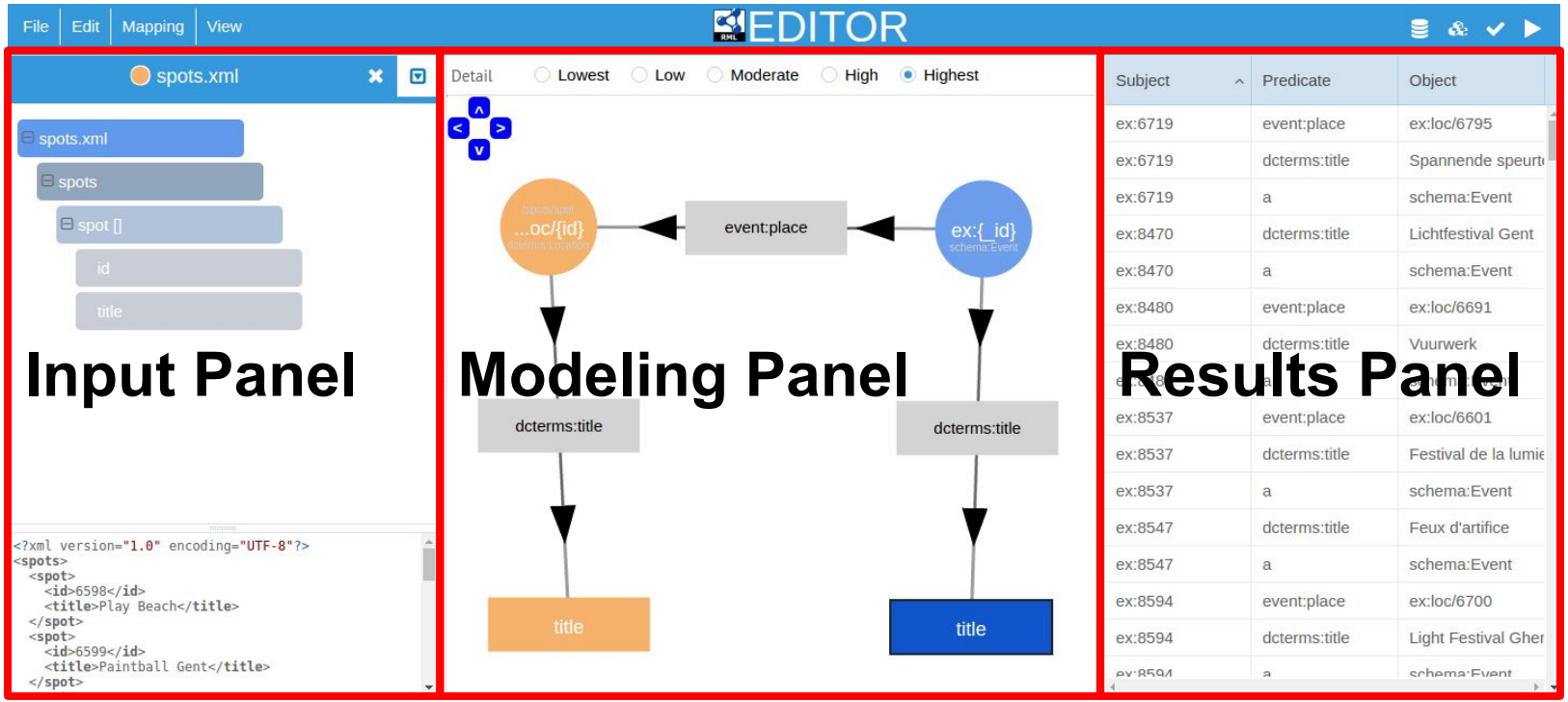
data owners vs. Semantic Web experts during this process

existing tools help the mapping process

RMLEditor

- **features**
- user validation

Three panels



Independent of mapping language

graph visualization to hide it

uses RML under the hood

mapping language can be replaced


ex:{ID}
schema:Movie
dcterms:title
Title

Map multiple data sources at the same time

data is spread across multiple data sources

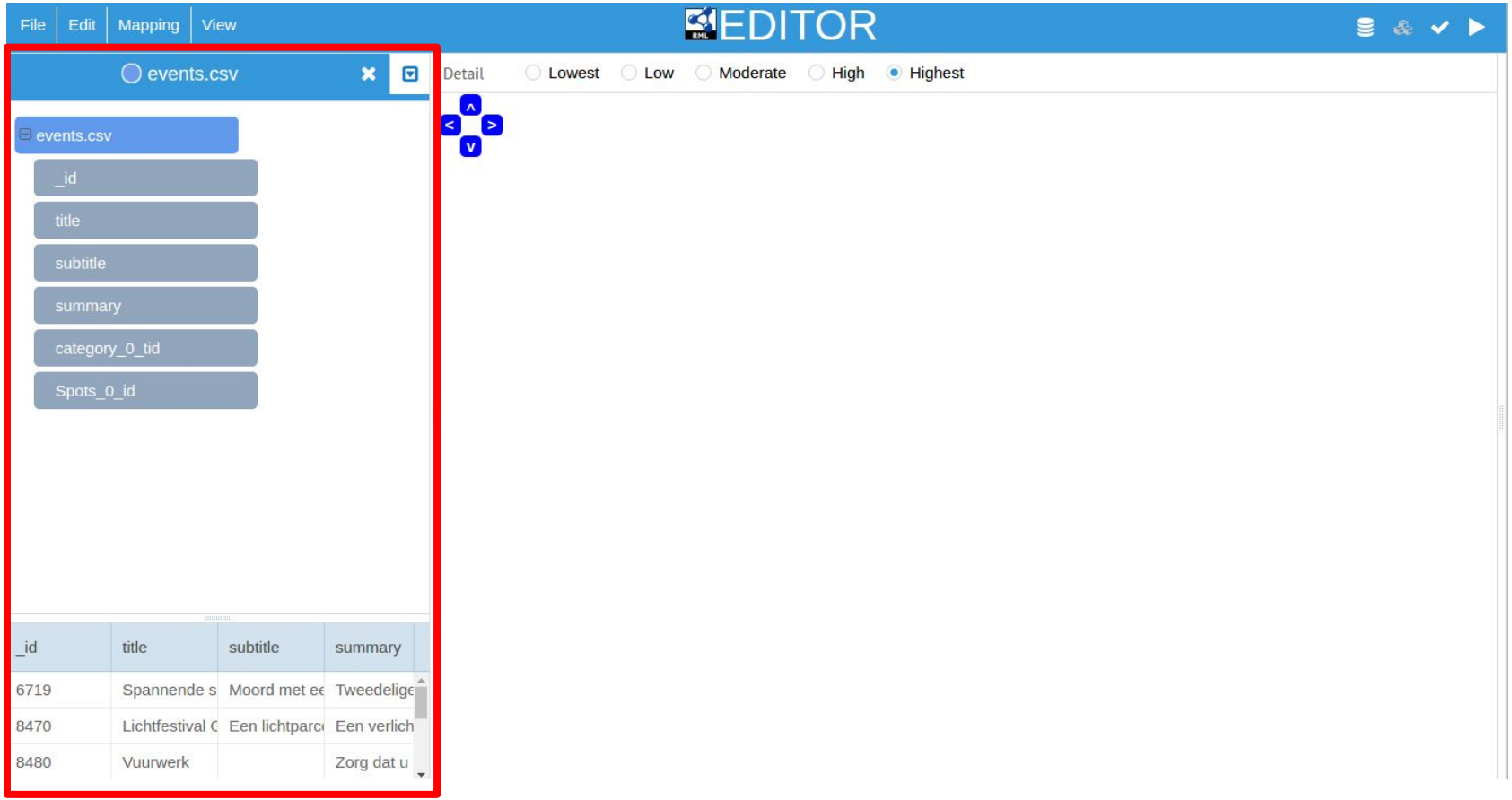
Map data sources in different formats

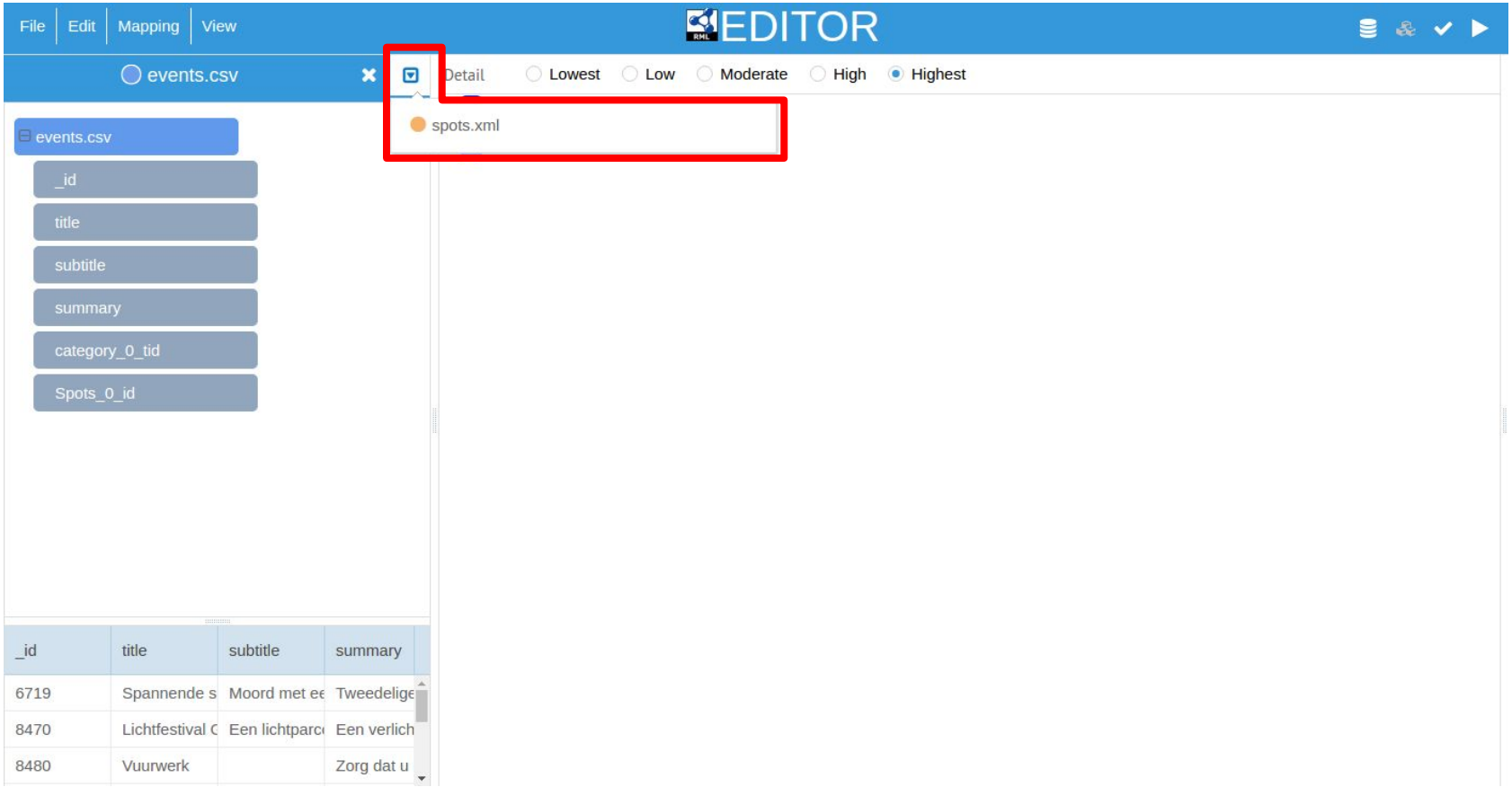
databases

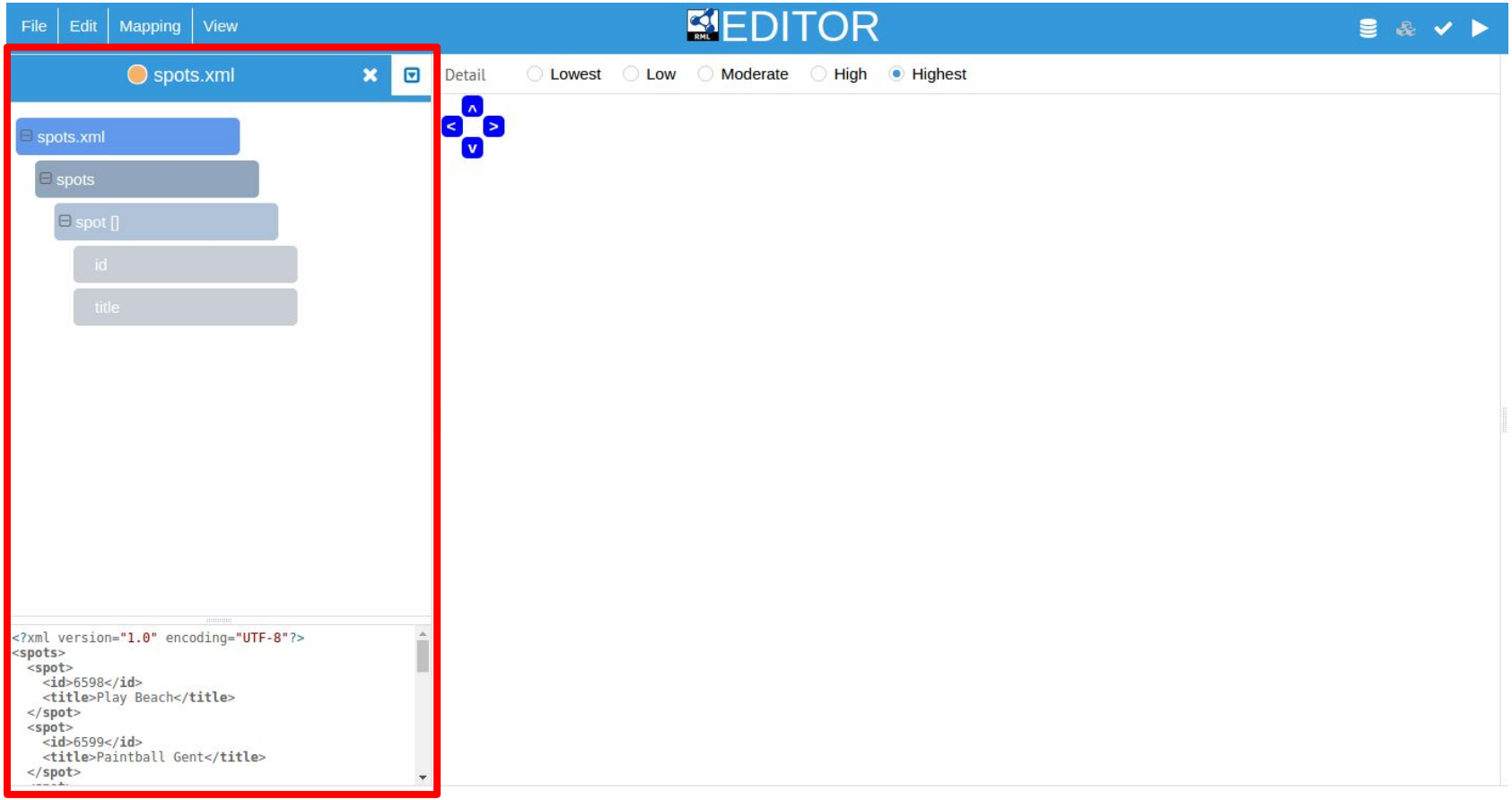
CSV files

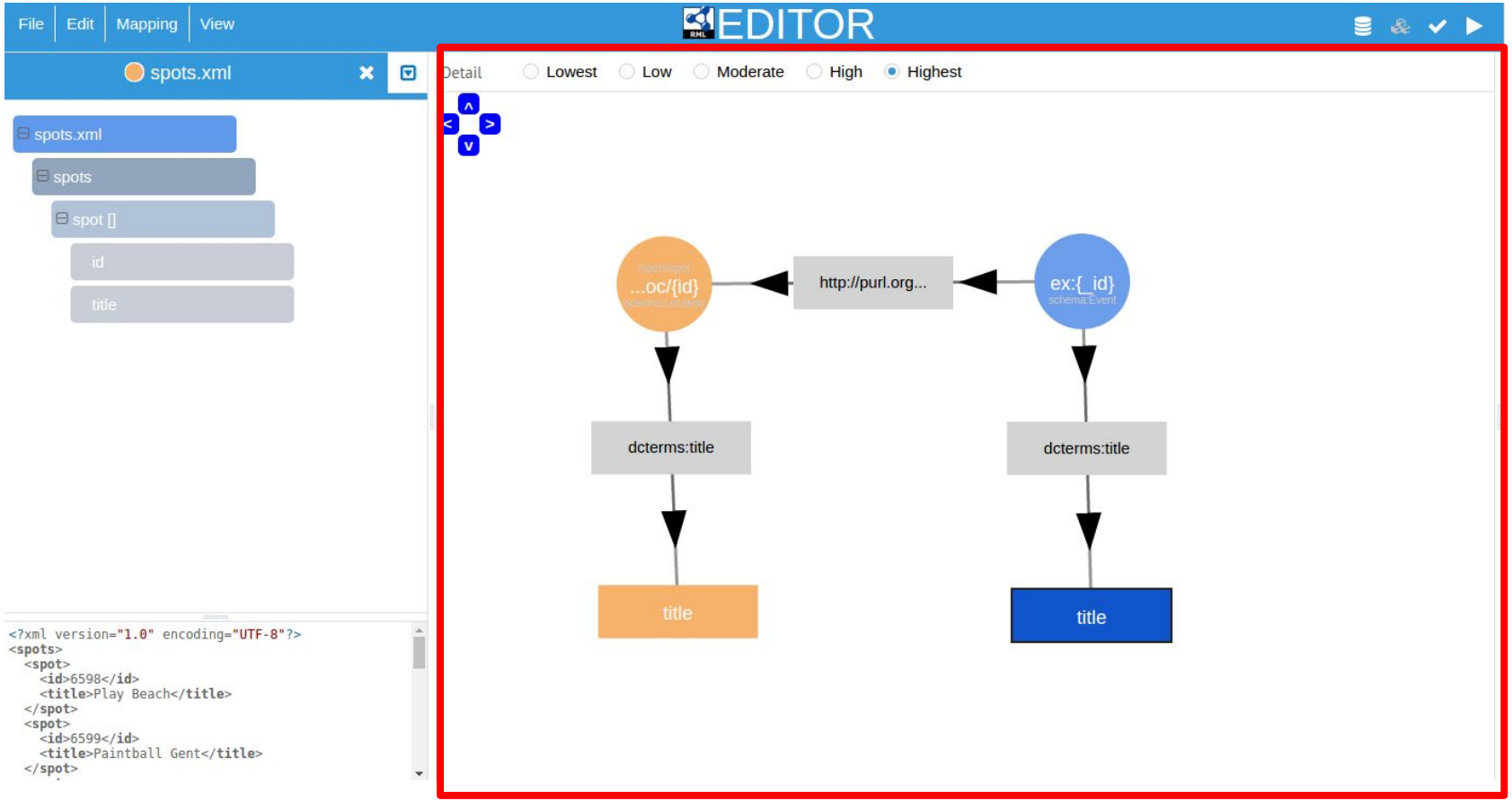
XML files

JSON files


File
Edit
Mapping
View
EDITOR
events.csv
Detail
Lowest
Low
Moderate
High
Highest
events.csv
_id
title
subtitle
summary
category_0_tid
Spots_0_id
_id
title
subtitle
summary
6719
Spannende s
Moord met ee
Tweedelige
8470
Lichtfestival C
Een lichtparc
Een verlich
8480
Vuurwerk
Zorg dat u


File
Edit
Mapping
View
EDITOR
events.csv
Detail
Lowest
Low
Moderate
High
Highest
spots.xml
events.csv
_id
title
subtitle
summary
category_0_tid
Spots_0_id
_id
title
subtitle
summary
6719
Spannende s
Moord met ee
Tweedelige
8470
Lichtfestival C
Een lichtparc
Een verlich
8480
Vuurwerk
Zorg dat u


File
Edit
Mapping
View
EDITOR
spots.xml
Detail
Lowest
Low
Moderate
High
Highest
spots.xml
spots
spot []
id
title
<?xml version="1.0" encoding="UTF-8"?>
<spots>
<spot>
<id>6598</id>
<title>Play Beach</title>
</spot>
<spot>
<id>6599</id>
<title>Paintball Gent</title>
</spot>


File
Edit
Mapping
View
EDITOR
spots.xml
spots.xml
spots
spot []
id
title
Detail
Lowest
Low
Moderate
High
Highest
...oc/{id}
http://purl.org...
ex:{_id}
schema:Event
dcterms:title
dcterms:title
title
title
<?xml version="1.0" encoding="UTF-8"?>
<spots>
<spot>
<id>6598</id>
<title>Play Beach</title>
</spot>
<spot>
<id>6599</id>
<title>Paintball Gent</title>
</spot>

Use multiple vocabularies and ontologies

not restricted to a single schema

can be changed during mapping process

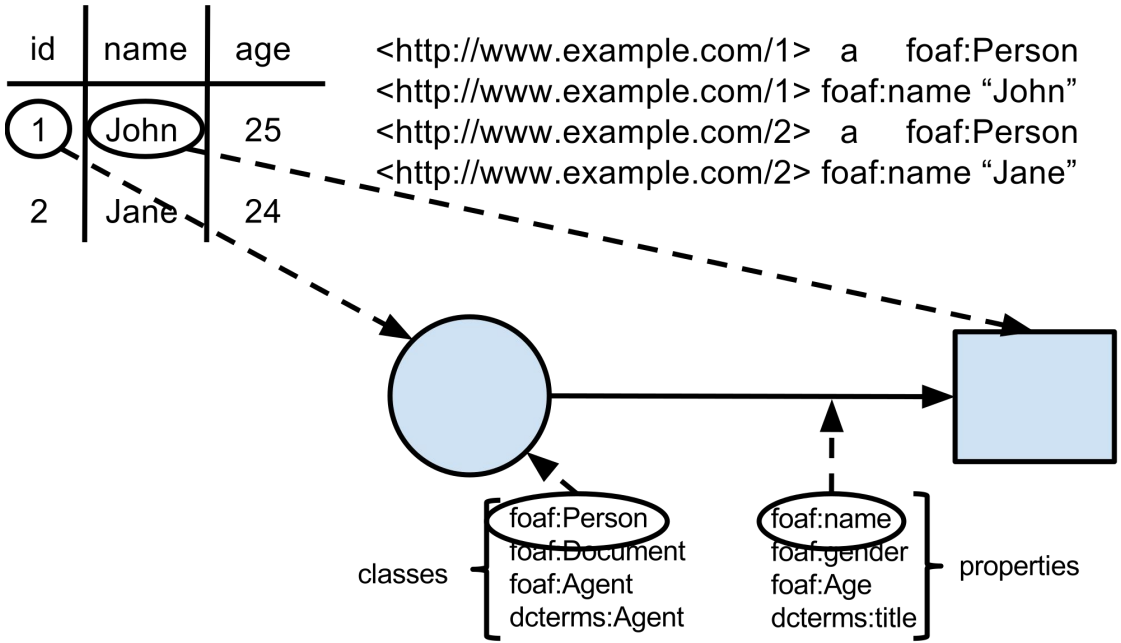
Allow multiple modeling approaches

certain use cases benefit from using specific approach

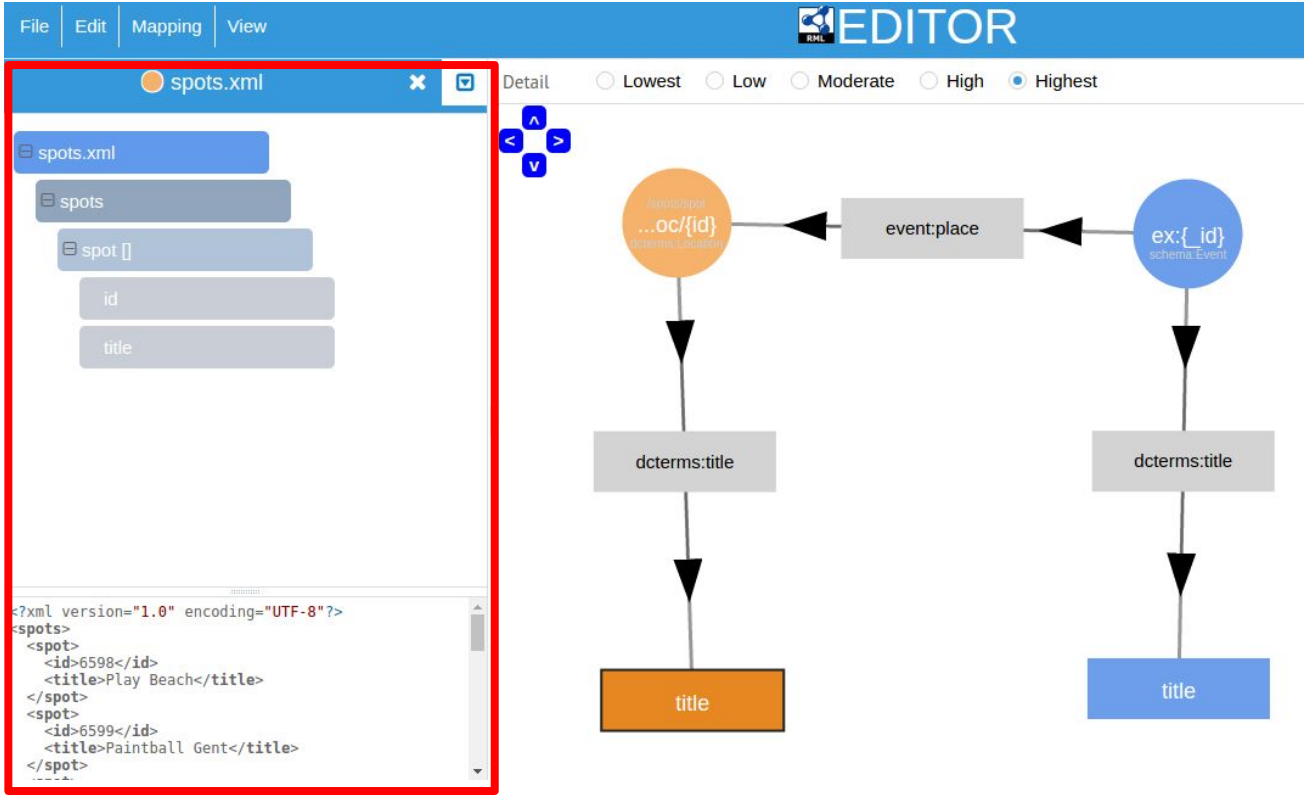
data-driven

schema-driven

Data-driven
start with data to model knowledge



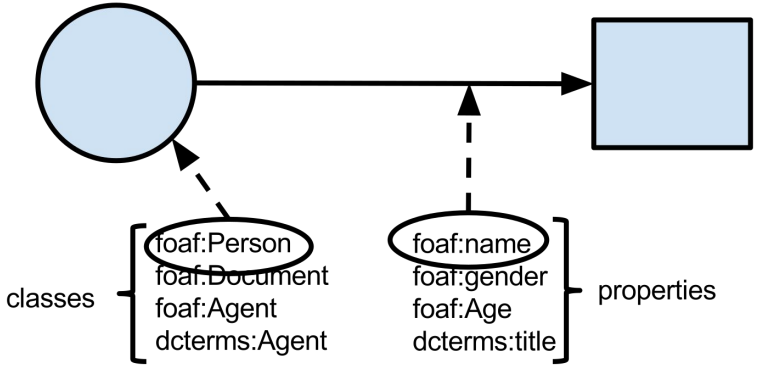
Data-driven via RMLEditor



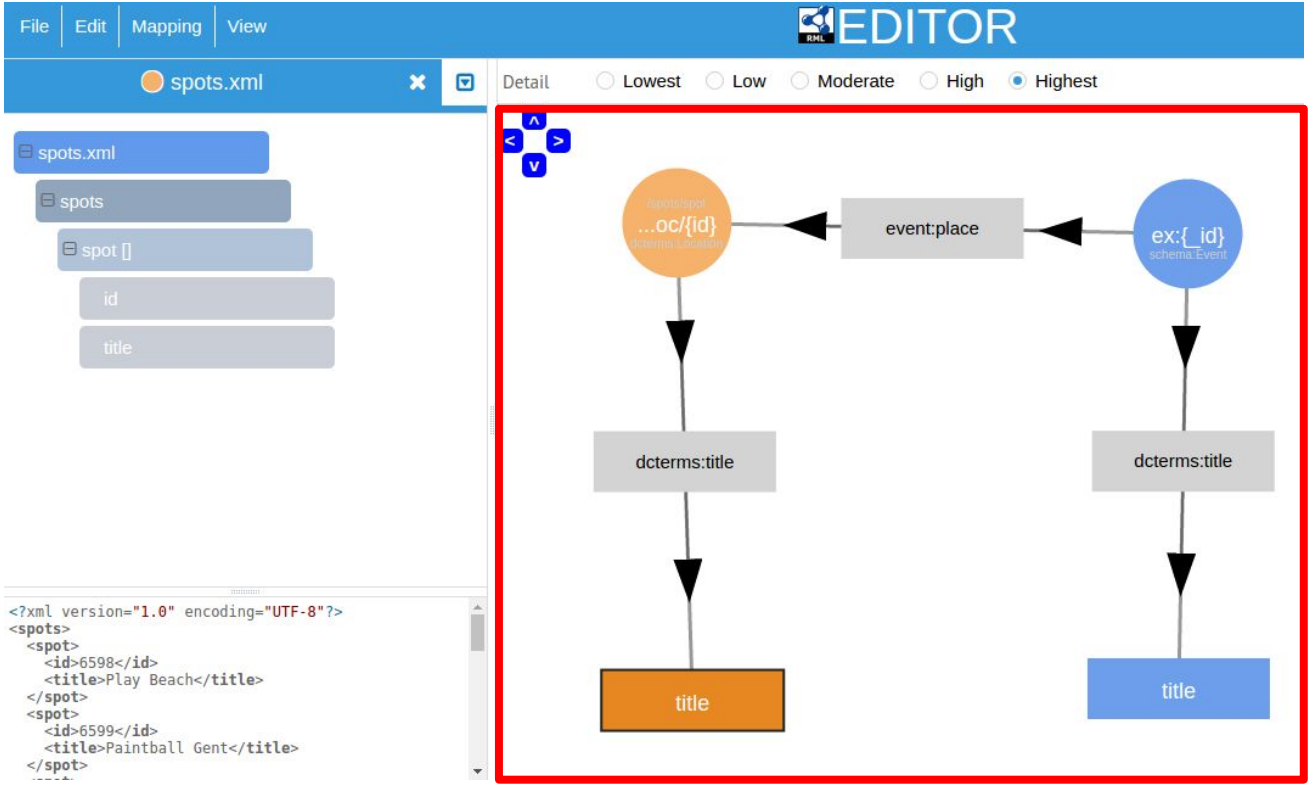
Schema-driven
start with schema to model knowledge

id	name	age
1	John	25
2	Jane	24

<http://www.example.com/1> a foaf:Person
<http://www.example.com/1> foaf:name “John”
<http://www.example.com/2> a foaf:Person
<http://www.example.com/2> foaf:name “Jane”



Schema-driven via RMLEditor



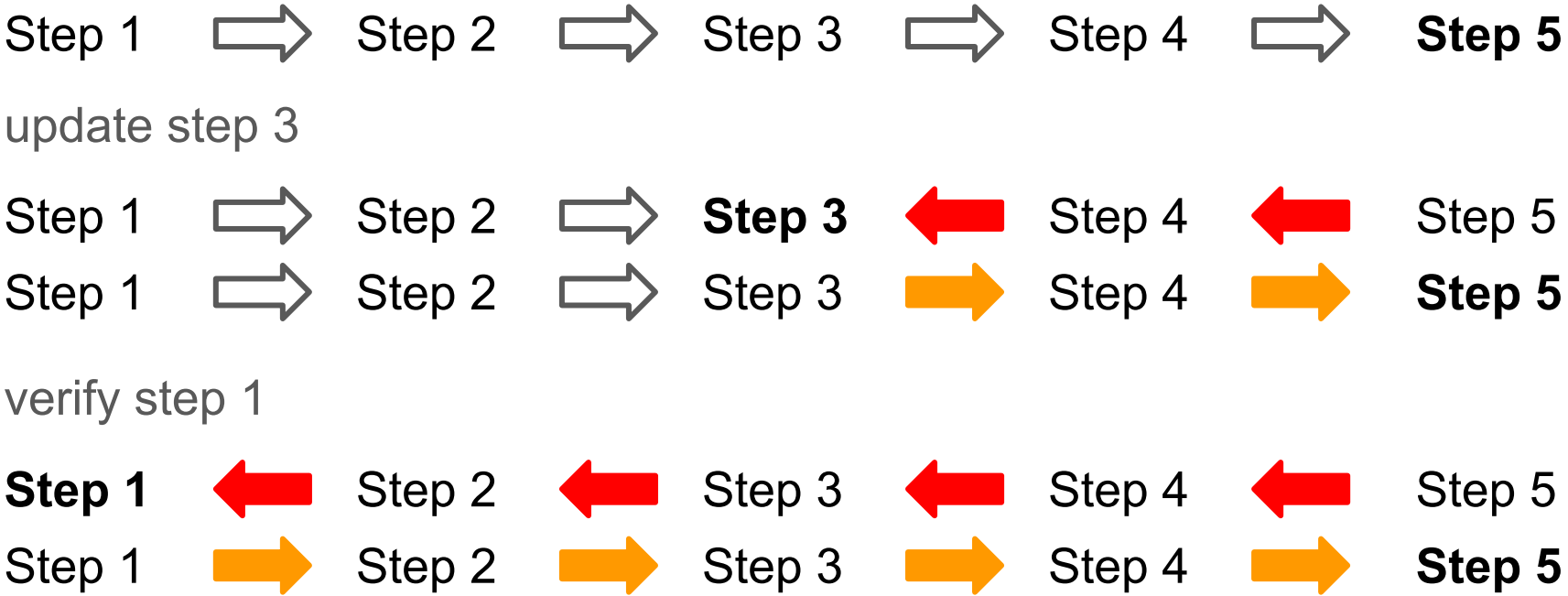
Allow non-linear workflows

keep overview of mapping model and relationships

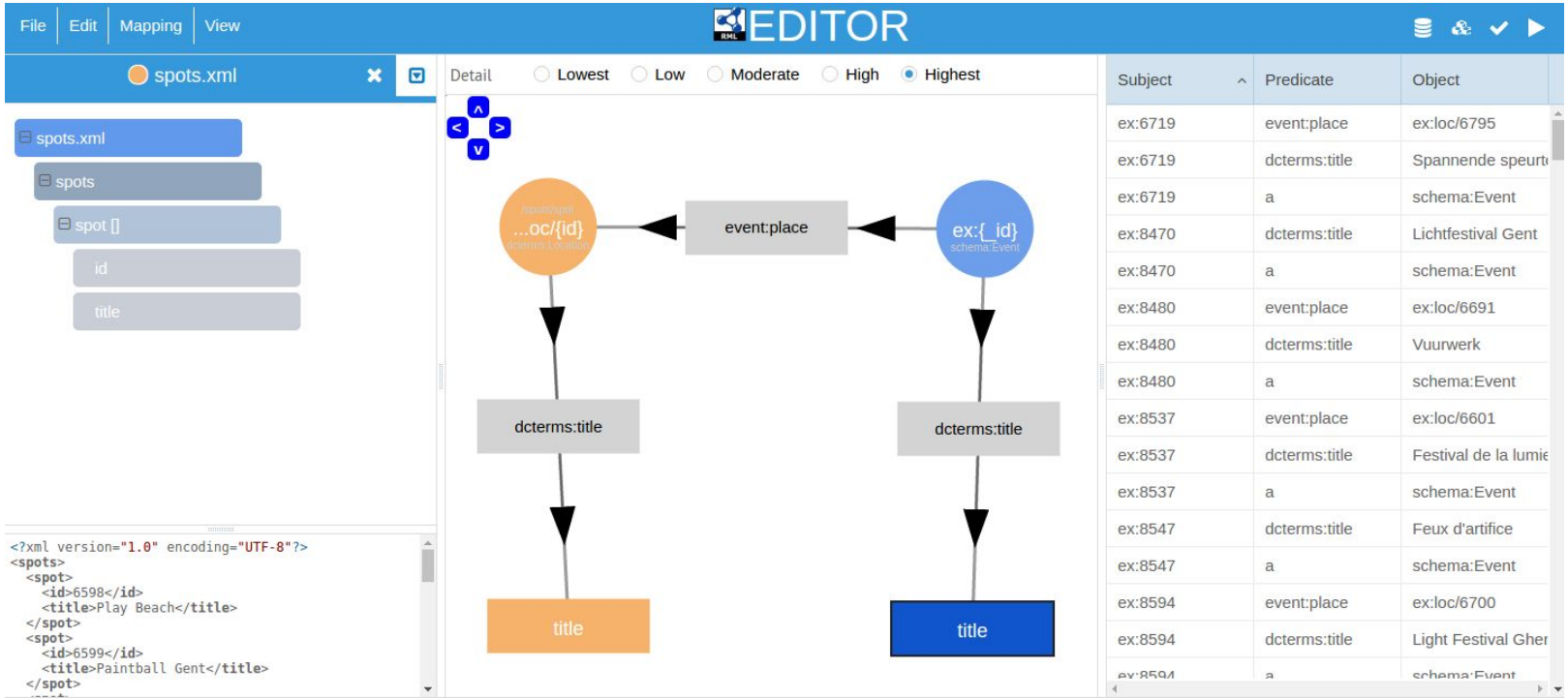
linear workflows force user’s steps

updating details in specific step is cumbersome

Linear workflows require to retrace steps to update specific step



Non-linear workflows via panels



Export mappings

mappings can be executed outside the RMLEditor

use reusability of mapping language

in the format specified by the mapping language

not tied to the visualization

Summary features

three panels
- allow multiple modeling approaches
- allow non-linear workflows

independent of mapping language

map multiple data sources at the same time
- possibly in different data formats

use multiple vocabularies and ontologies

export mappings

Overview

mapping process to generate Linked Data

data owners vs. Semantic Web experts during this process

existing tools help the mapping process

RMLEditor

- features
- **user validation**

(Non)-Semantic Web experts as participants

10 Semantic Web experts

- 5 with experience in Linked Data publishing
- 5 without experience in Linked Data publishing

5 non-Semantic Web experts

Each participant completed two use cases

data-driven: start with the data

schema-driven: start with the schema

experts with experience in LD publishing
also did the use cases by directly writing RML statements

Important results

supported approaches

graph visualization

interlinking between data sources

Data-driven and schema-driven approaches supported

choice driven by personal preference

Graph visualization helps editing

participants found it beneficial

only limited number of use cases
done by SW experts were incomplete or inaccurate

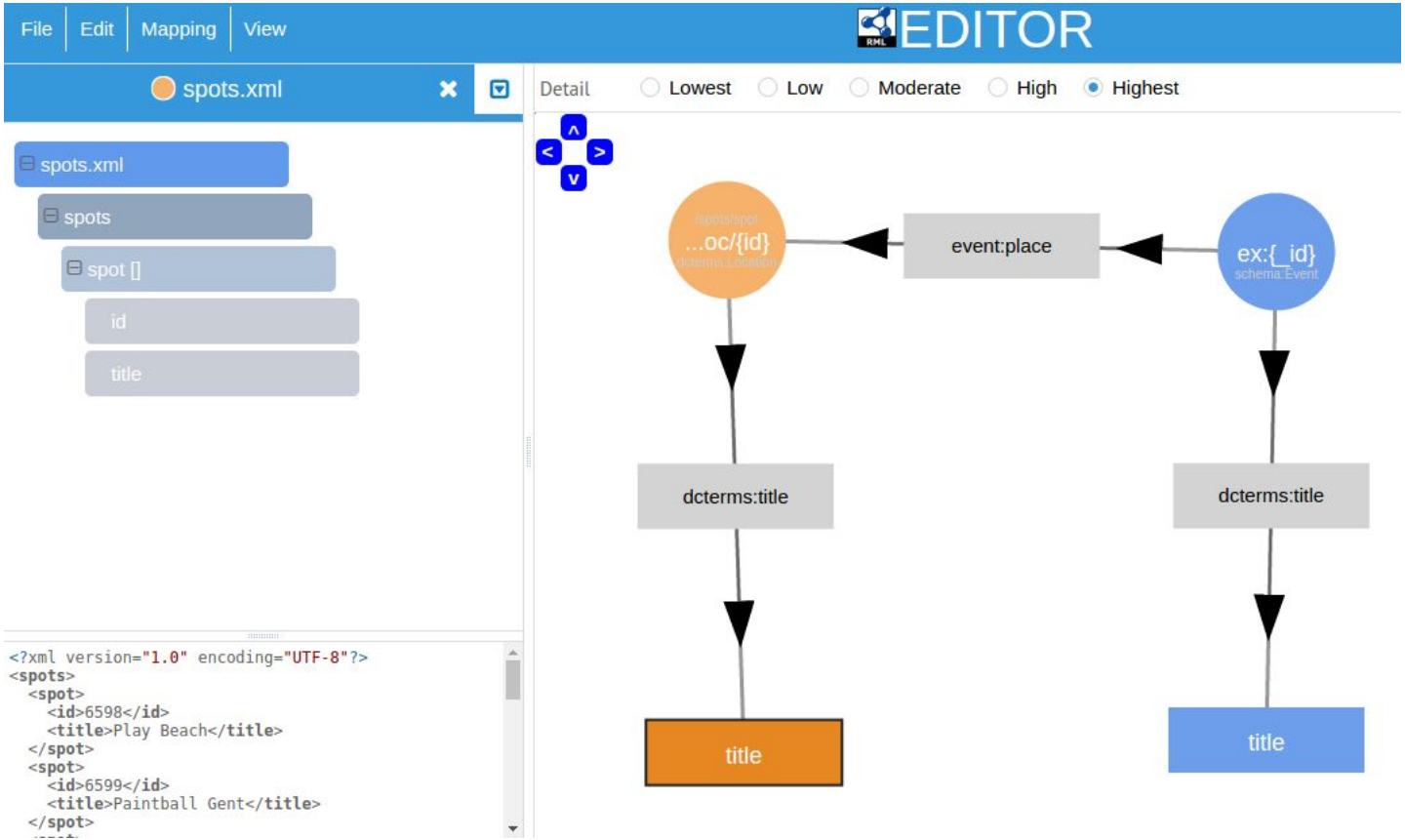
non-Semantic Web experts are able to define mappings

Improved interlinking via graphs

creating links between data in different data sources

certain use cases missed interlinking when using RML directly

20% less when using RMLEditor


File
Edit
Mapping
View
EDITOR
spots.xml
Detail
Lowest
Low
Moderate
High
Highest
spots.xml
spots
spot []
id
title
<?xml version="1.0" encoding="UTF-8"?>
<spots>
<spot>
<id>6598</id>
<title>Play Beach</title>
</spot>
<spot>
<id>6599</id>
<title>Paintball Gent</title>
</spot>
...oc/{id}
event:place
ex:{_id}
schema:Event
dcterms:title
dcterms:title
title
title

Roundup

data owners and Semantic Web have different areas of expertise

existing tools help the mapping process, but have their limitations

RMLEditor brings editing of mappings to non-Semantic Web experts

See it in action at demo #16

Generate your Linked Data with the RMLEditor

and help us improve our tool

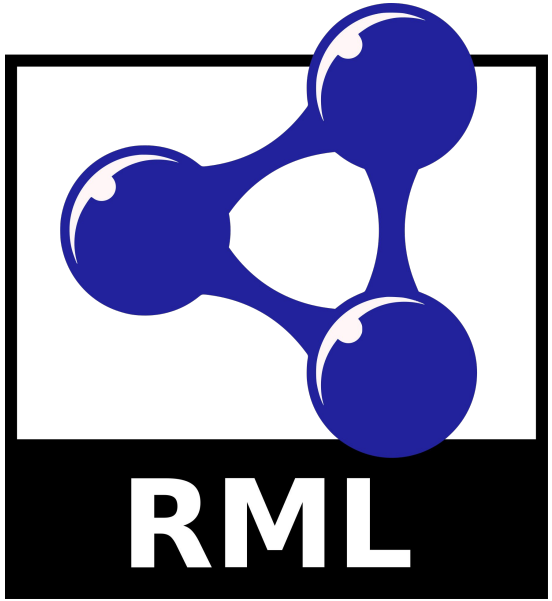
alpha version

free academic license

contact us for access

Pieter Heyvaert

pheyvaer.heyvaert@ugent.be
visit rml.io



Comparison tools

	fluidsOps editor	DataOps	Sheet2RDF	ontopPro	Karma	RDF123	RMLEditor
non-linaer workflow	✗	✗	✓	✗	✗	✗	✓
data-driven approach	✓	✓	✓	✓	✓	✓	✓
schema-driven approach	✓	✓	✗	✗	✗	✗	✓
multiple data sources	✓	✓	✓	✓	✓	✗	✓
multiple data formats	✗	✓	✗	✗	✓	✓	✓
no knowledge language	✗	✗				✓	✓

Details levels to improve editing

highest: all information

high: less details nodes

moderate: no predicates

low: no literals

lowest: no blank nodes