A Jigsaw Puzzle Metaphor for Representing Linked Data Mappings

A thesis submitted to the

University of Dublin, Trinity College

in fulfilment of the requirements of the degree of

Doctor of Philosophy

Ademar Crotti Junior

Knowledge and Data Engineering Group (KDEG)

School of Computer Science and Statistics

Trinity College, Dublin

Ireland

Supervised by Prof. Declan O'Sullivan

Co-supervised by Dr. Christophe Debruyne

2019

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed: _____

Permission to lend or copy

I, the undersigned, agree that the Trinity College Library may lend or copy this thesis upon request.

Signature _____   Date _____

Ademar Crotti Junior

Acknowledgements

I would like to thank my supervisors, Declan O'Sullivan and Christophe Debruyne, for all of their support and guidance throughout this Ph.D. I could not have done it without them.

Abstract

This thesis presents a visual representation approach for Linked Data mappings known as **J**igsaw **Pu**zzles for Representing **Ma**ppings, or Juma.

The term Linked Data refers to a set of best practices for publishing and interlinking data on the Web. A Linked Data dataset is structured information encoded using the Resource Description Framework (RDF), in which resources are identified by and linked with other datasets using HTTP URIs.

Linked Data datasets cover a wide range of knowledge domains, where often concepts overlap. In such cases, mappings can be created to reduce heterogeneity and facilitate the consumption of information by informing agents which concepts are related, and how. These types of mappings are called semantic mappings. Another area in which we find use for mappings is when transforming data from one representation to another – from non-RDF to RDF for example. We call those mappings uplift mappings. Producing such mappings can be difficult, even for experts in Semantic Web technologies, requiring knowledge on the specifics of the mapping language being used as well as significant amount of human effort for their creation, modification, curation and maintenance. Nonetheless, literature suggests that this user involvement is fundamental for producing quality mappings. Suitable visual representations may be used to support user involvement and alleviate the knowledge required for producing Linked Data mappings.

Through a systematic literature review, a set of requirements for a visual representation for Linked Data mappings were defined. Juma was then proposed as a novel approach, based on the block metaphor, for the representation of mappings in Linked Data. The block – or jigsaw – metaphor was chosen as it takes advantage of the user's familiarity to jigsaw puzzles, fosters users to explore the combinations of blocks, and for being accessible to experts and non-experts alike. Juma leverages the use of the block metaphor in order to facilitate the interpretation of mappings in Linked Data. In Juma, blocks are used to abstract and capture different mapping constructs, where the connection of the different blocks form a mapping. Each block is then translated to an equivalent mapping representation, which can be done for distinct mapping languages.

The Juma approach was evaluated through five experiments categorized in three aspects: creation (and editing), understanding, and expressiveness. The creation of mappings was evaluated through two user experiments, where participants were asked to create a mapping using applications that apply the Juma approach for the representation of mappings. Another user experiment was conducted to evaluate the understanding of mappings represented using

a Juma application. Finally, two experiments were conducted to evaluate the expressiveness of the Juma approach in the representation of uplift and semantic mappings, respectively. These evaluations indicated that the Juma approach is effective in representing uplift and semantic mappings, and that it aids users in the creation, editing and understanding of Linked Data mappings.

The research in this thesis has yielded one major contribution and three minor contributions. The major contribution is the design and development of the **Jigsaw Puzzles for Representing Mappings (Juma) approach**. The first minor contribution is the **Juma R2RML** application. Juma R2RML applies the Juma approach to the R2RML mapping language. The second minor contribution is the **Juma Uplift** application. Juma Uplift has a higher level of abstraction in order to be able to generate mappings using multiple distinct mapping languages. The Juma R2RML and Juma Uplift applications apply the Juma approach in the representation of uplift mappings. The third minor contribution is the **Juma Interlink** application. Juma Interlink applies the Juma approach in the representation of semantic mappings.

Table of Contents

List of Figures

List of Tables

Abbreviations

AL	Alignment Format
CSV	Comma Separated Values
EDOAL	Expressive and Declarative Ontology Alignment Language
JUMA	Jigsaw Puzzles for Representing Mappings
LOD	Linked Data Cloud
MWL	Mental Workload
NASA-TLX	NASA Task Load Index
OAEI	Ontology Alignment Evaluation Initiative
OWL	Web Ontology Language
PSSUQ	Post-Study System Usability Questionnaire
R2RML	RDB to RDF Mapping Language
RDF	Resource Description Frammework
RML	RDF Mapping Language
SML	Sparqlification Mapping Language
SWRL	Semantic Web Rule Language
WP	Workload Profile
XML	Extensible Markup Language

1. Introduction

1.1. Motivation

The term Linked Data refers to a set of best practices for publishing and interlinking data on the Web (Bizer, Heath, & Berners-Lee, 2009). A Linked Data dataset is structured information encoded using the Resource Description Framework (RDF) (Hayes & Patel-Schneider, 2014), in which resources are identified by and linked with other datasets using HTTP URIs, making those resources accessible via the HTTP protocol.

Linked Data best practices, guidelines and underlying technologies can be applied to *open* and *non-open* data. For instance, one can apply these principles within an organization to facilitate data integration with all resources remaining behind a firewall (Debruyne et al., 2016). To demonstrate the capabilities of Linked Data in an open context, the Linking Open Data (Bizer et al., 2009) project was set up. This project has the goal of publishing open datasets as Linked Data. These open datasets are freely accessible and collectively known as the Linked Open Data (LOD) cloud[1]. The publication of open data as Linked Data was key in the uptake of Linked Data technologies, even in industry.

A number of datasets found in the LOD cloud have overlapping concepts (Euzenat & Shvaiko, 2007). Examples include different identifiers for the same entity in different contexts, such as the concept of *County* in Census data and the concept of *County* in maps. The heterogeneity caused by the representation of such overlapping concepts can become problematic for agents wishing to explore and process Linked Data. In such cases, **mappings**[2] can be created to reduce heterogeneity and facilitate the consumption of information from the LOD cloud by informing agents which concepts are related, and how. For instance, one could define a mapping where the concept of *County* in Census data is equivalent to the concept of *County* in maps. These types of mappings relate concepts from different schemas or vocabularies, and are called *semantic mappings* (Euzenat & Shvaiko, 2007).

Oftentimes, data is stored in "silos" in a variety of formats (e.g., CSV, relational databases, etc.) which creates another problem – how do we transform data from those silos to RDF so that it can be published as Linked Data? Transforming data from one representation to another, from non-RDF to RDF for example, is another area in which the use for mappings can be found. In the thesis those mappings are called *uplift mappings* (Bizer

---

[1] `http://lod-cloud.net/`, accessed in August 2018.
[2] A mapping defines a set of relations between elements of different inputs (Sicilia et al., 2017).

& Seaborne, 2004). Uplift mappings are also responsible for generating a significant part of the LOD cloud itself.

Mappings separate mapping definitions from their execution, allowing the mapping process to be reproduced (e.g. when updating a dataset, or creating a new one, where mappings – or parts of – can be reused). The process of creating and executing uplift and semantic mappings are not mutually exclusive. Although semantic mappings require the availability of Linked Data datasets, the creation of those semantic mappings may be preceded by the creation and execution of an uplift mapping to obtain the required dataset.

Two types of mappings can be identified in a Linked Data context: semantic mappings to relate and interlink Linked Data datasets and uplift mappings to generate Linked Data datasets. But how can one represent and execute such mappings? SPARQL (Harris, Seaborne, & Prud'hommeaux, 2013), the W3C Recommendation specifying a query language for RDF data, can be used to represent semantic mappings through SPARQL CONSTRUCT queries (Thiéblin, Amarger, Haemmerlé, Hernandez, & Trojahn, 2016). An example of a semantic mapping represented using a SPARQL CONSTRUCT query is shown in Listing 1. In this example, one might have discovered that the class `dbpedia:Person` of one schema is equivalent to the class `foaf:Person`[3] in another. This mapping encodes a relationship between `dbpedia:Person` and `foaf:Person`: every instance of the class `dbpedia:Person` is an instance of the class `foaf:Person`. Thus, for every resource that plays the role of `rdf:type` on the class `dbpedia:Person` a triple will be generated where that resource will play the role of `rdf:type` on `foaf:Person`. The generated triples may be added to a dataset, creating interlinks with another Linked Data vocabulary.

```
prefix foaf: <http://xmlns.com/foaf/0.1/>
prefix dbpedia: <http://dbpedia.org/ontology/>

CONSTRUCT {?subj a   foaf:Person . }
   WHERE {?subj a   dbpedia:Person . }
```
Listing 1. Semantic mapping represented as a SPARQL CONSTRUCT query

The RDB to RDF mapping language (R2RML) (Das, Sundara, & Cyganiak, 2012) is, as its name implies, an example of an uplift mapping language. R2RML is a W3C Recommendation used to express mappings from relational databases to RDF. R2RML provides an algorithm and a vocabulary, the latter being provided using the Web Ontology

---

[3] We note that equivalence may be context-dependent. The equivalence between two classes might be true for a particular dataset, but not for all datasets.

Language (OWL) (Hitzler, Krötzsch, Parsia, Patel-Schneider, & Rudolph, 2012). An example[4] of an uplift mapping is presented in Listing 2 (using R2RML and stored in RDF TURTLE (Beckett, Berners-Lee, Prud'hommeaux, & Carothers, 2014) syntax[5]). This mapping transforms information contained in records of the table *Person* into instances of the class `foaf:Person`. The subject URI of the triples is defined as `http://example.org/{id}`, where *id* is a column of the table *Person*. The column *name* is mapped using the predicate `foaf:name`.

```
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<#TriplesMap1>
 rr:logicalTable [ rr:tableName "Person"; ];

 rr:subjectMap [
    rr:template "http://example.org/{id}";
    rr:class foaf:Person;
 ];

 rr:predicateObjectMap [
   rr:predicateMap [ rr:constant foaf:name; ];
   rr:objectMap [ rr:column "name"; ];
 ];.
```

Listing 2. Uplift mapping represented in R2RML

One issue with semantic and uplift mapping languages is that their representation, often based on Semantic Web technologies such as RDF and SPARQL are not intuitive, especially for non-expert users (Pietriga, Bizer, Karger, & Lee, 2006; Rietveld & Hoekstra, 2017). Even users familiar with such technologies may find it problematic as they still require knowledge of the specifics of the mapping language being used (e.g., learning the terminology), and significant human effort on manually creating, editing, curating, and maintaining mappings (Sicilia, Nemirovski, & Nolle, 2017). Additionally, *user involvement* has been identified as one of the main challenges when producing quality semantic mappings (Shvaiko & Euzenat, 2013). The lack of evaluations considering user aspects has also been identified as an issue by the Ontology Alignment Evaluation Initiative[6] (OAEI), which was established in 2004, but only introduced a track concerned with user involvement in 2013 –

---

[4] This example is inspired from the R2RML W3C Recommendation (Das et al. 2012).
[5] TURTLE is only one of the many standardized RDF representations. The R2RML W3C Recommendation uses TURTLE for their examples. In this thesis, whenever we refer to the syntax of RDF, we do refer to the TURTLE notation.
[6] `http://oaei.ontologymatching.org/`, accessed in September 2018.

Interactive Matching. Nonetheless, in 2018, this track had only four system participants[7]. Sequeda and Miranker (2017) report on challenges in automating the mapping process applied to uplift mappings, where user involvement was also considered fundamental for the definition of quality mappings. Suitable visual representations may be used to address such challenges, support user involvement and alleviate the knowledge required for producing semantic and uplift mappings (Granitzer, Sabol, Onn, Lukose, & Tochtermann, 2010).

Existing visual representations for mappings in a Linked Data context can be classified at a high level as tree and graph representations (Rahm, 2011). As the state of the art in Chapter 3 will show, these are designed specifically for one mapping language and therefore can only be applied to a particular (uplift or semantic) mapping representation. It is argued in this thesis that a **visual representation** that would allow users to create semantic and uplift mappings, independently of mapping representation, would benefit non-experts and experts alike. It is also argued that the integration of such a representation in an interface will allow users to focus on what is to be mapped rather than how the mapping is to be implemented in a particular mapping language. As mentioned, user involvement is fundamental during the mapping process, however, the state of the art chapter also shows that existing approaches often neglect considering the **different types of users** involved in the mapping process.

In this thesis, a visual representation for mappings in Linked Data, called **J**igsaw **Pu**zzles for Representing **Ma**ppings (Juma), is proposed and evaluated. Unlike the approaches discussed in the state of the art, Juma is based on the block – or jigsaw – metaphor that has become popular with visual programming languages – where it is called the block paradigm. Metaphors make use of familiar concepts in order to aid users in the understanding of another – unknown or complex – concept (Ziemkiewicz & Kosara, 2008). The block metaphor is capable of expressing information using a representation that takes advantage of a user's familiarity to jigsaw puzzles. Given the importance of the user in the mapping process, this metaphor was applied in the representation of Linked Data mappings as it is a concept that is accessible to both experts and non-experts alike. The use of this metaphor also allows users to focus on creating meaningful mappings from source to target elements, and less on a particular mapping language's syntax. Finally, this metaphor has been used successfully in other domains such as querying Linked Data (Ceriani & Bottoni, 2017), programming robots (García-Zubía et al., 2018), and in the data science domain (Bart, Tibau, Kafura, A. Shaffer, & Tilevich, 2017).

---

[7] `http://oaei.ontologymatching.org/2018/results/interactive/index.html`, accessed in October 2018.

1.2. Research Question

The research question investigated in this thesis is:

To what extent can the interpretation of mappings in Linked Data be facilitated through a visual representation?

The definition of the term *interpretation*, as it is used in the research of this thesis, refers to the understanding of a visual mapping representation, such that it allows users to create, edit and interpret mappings for generating and interlinking Linked Data datasets.

1.2.1. Research Objectives

In order to address the research question outlined above, the following research objectives were identified for this research:

- **RO1:** Perform a state-of-the-art review of existing visual representations for uplift and semantic mappings.
- **RO2:** Propose a visual representation applicable to both uplift and semantic mappings.
- **RO3:** Apply, implement and evaluate the visual representation defined as a result of RO2 to uplift mappings.
- **RO4:** Apply, implement and evaluate the visual representation defined as a result of RO2 to semantic mappings.

1.2.2. Thesis Contributions

The proposed Juma approach is the major contribution of this thesis. The minor contributions are three applications that apply the Juma approach in the representation of uplift and semantic mappings. The different contributions are now elaborated in more detail in this subsection.

1.2.2.1. Jigsaw Puzzles for Representing Mappings (Juma) approach

The major contribution of this thesis is the design and development of the **Jigsaw Puzzles for Representing Mappings (Juma)** approach. Juma leverages the use of the block metaphor in order to facilitate the interpretation of mappings in Linked Data. Findings from

the experiments outlined in this thesis indicate that Juma can be used for the representation of both uplift and semantic mappings, and that users unfamiliar with a particular mapping language are guided into creating (at least syntactically) valid mappings. Experiments also indicate that Juma can facilitate the interpretability of mappings in Linked Data to experts and non-expert users. In addition, a benefit of Juma is that it offers a uniform representation for similar mapping constructs that can be found in both uplift and semantic mappings, which may aid users in creating, editing and understanding Linked Data mappings, regardless of its type.

The Juma approach advances the state of the art by offering a visual representation for Linked Data mappings that can be applied to both uplift and semantic mappings, and that is focused on the different types of users (experts and non-experts) involved in the mapping process. The Juma approach has been evaluated considering the performance[8] of participants, perceived usability and perceived mental workload. The evaluations presented in this thesis apply a standard usability questionnaire that considers different characteristics of a system, including a specific aspect for interface quality – which is of particular importance since this thesis proposes a visual representation. Additionally, to the authors knowledge, this thesis advances the state of the art by presenting the first experiments evaluating the perceived mental workload of creating and understanding uplift mappings. Mental workload assessment procedures quantify the cognitive load of performing a task and can also be used to describe user experience (Longo, 2015).

As well as advancing the state of the art, it is envisaged that the Juma approach would have an impact on the adoption of Linked Data by widening the types of users that could get involve in the creation, editing and understanding of Linked Data mappings. The argument is also made that maintainers of Linked Data datasets can use and integrate Juma in their mapping processes.

It is also hoped that the Juma approach would benefit the research community. Researchers can employ Juma in their mapping processes and use the findings presented in this thesis in their research. Researchers may also apply their expertise to contribute to the approach and its implementations.

1.2.2.2. Juma R2RML

The first minor contribution of this thesis is the development of the **Juma R2RML application**. Juma R2RML applies the Juma approach to the R2RML mapping language for

---

[8] Performance, as it will be explained in Chapter 5, is defined as the accuracy achieved by participants in a particular task.

specifying uplift mappings. The representation of mappings in this application reflects the structure of the R2RML vocabulary, where each block has been designed to represent an R2RML construct. Juma R2RML thus abstracts the R2RML vocabulary's syntax, and guides users in the creation of valid R2RML mappings. A user experiment has shown that users with different background knowledge were able to create and edit mappings using this application with high performance and sufficient usability. This application is available for use by researchers and practitioners for uplift mappings.

1.2.2.3. Juma Uplift

The second minor contribution is the development of the **Juma Uplift application**. This application of the Juma approach has a higher level of abstraction in order to have the capability of generating uplift mappings that are not only compliant with R2RML, but also with other mapping languages, such as SML (Stadler, Unbehauen, Westphal, Sherif, & Lehmann, 2015). An experiment has shown that Juma Uplift generates accurate uplift mappings using the aforementioned uplift mapping languages. User experimentation has shown that Juma Uplift facilitates the creation, editing and understanding of uplift mappings when compared to mappings represented in R2RML (in RDF TURTLE). This application is available for use by researchers and practitioners for uplift mappings.

1.2.2.4. Juma Interlink

The third minor contribution is the application of the Juma approach to semantic mappings, which is called **Juma Interlink**. This application is capable of representing simple and complex semantic mappings that automatically generate executable mappings in the form of SPARQL CONSTRUCT queries. Experiment results have shown that Juma Interlink is expressive enough to represent mappings from a real-world scenario. This application is available for use by researchers and practitioners for semantic mapping/interlinking in LD datasets.

1.2.2.5. Publications

The publications associated with the research in this thesis to date are:

- **Crotti Junior, A., Debruyne, C., Longo, L. and O'Sullivan D. "On the Mental Workload Assessment of Uplift Mapping Representations in Linked Data". In**

Proceeding of the 2nd International Symposium on Human Mental Workload: Models and Applications, H-WORKLOAD 2018.

This publication presents initial results on the mental workload assessment of uplift mapping representations. The experiment presented in this paper compares the Juma Uplift mapping representation to the R2RML (RDF TURTLE) representation.

- **Crotti Junior, A., Debruyne, C. and O'Sullivan D. "An Editor that Uses a Block Metaphor for Representing Semantic Mappings in Linked Data". In** *Proceedings of the 15th Extended Semantic Web Conference: Posters and Demos, ESWC 2018.*

  This publication presents the Juma approach applied to semantic mappings. A demonstration of the approach is also presented.

- **Crotti Junior, A., Debruyne, C. and O'Sullivan D. "Juma Uplift: Using a Block Metaphor for Representing Uplift Mappings". In** *Proceedings of the 12th IEEE International Conference on Semantic Computing, ICSC 2018.*

  This publication presents Juma Uplift as the second application applying the Juma approach. A comparison between Juma R2RML and Juma Uplift's mapping representation is also presented together with an experiment on the expressiveness of Juma Uplift.

- **Crotti Junior, A., Debruyne, C. and O'Sullivan D. "Using a Block Metaphor for Representing R2RML Mappings". In** *Proceedings of the 3rd International Workshop on Visualization and Interaction for Ontologies and Linked Data co-located with the 16th International Semantic Web Conference, VOILA@ISWC 2017.*

  This publication describes in more detail the first Juma application, called Juma R2RML. This publication also presents a detailed analysis of a user experiment carried out to evaluate the creation and editing of uplift mappings using Juma R2RML.

- **Crotti Junior, A., Debruyne, C. and O'Sullivan D. "Juma: an Editor that Uses a Block Metaphor to Facilitate the Creation and Editing of R2RML Mappings". In** *Proceedings of the 14th Extended Semantic Web Conference: Posters and Demos, ESWC 2017.*

This publication presents the Juma approach and the first implementation of the approach (Juma R2RML). Preliminary results of a user experiment are also presented.

The following publications, although not the focus of this thesis, have guided the author into the research area of Linked Data mappings. In these publications a number of uplift mapping languages from the state of the art are evaluated, and the **Fun**ctions into **U**plift Mapping **L**anguages (FunUL) method is proposed. FunUL allows data transformations – which are often needed when transforming data from one representation into another – and the uplift to RDF to happen in unified step, thus facilitating the mapping process by making it more traceable and transparent.

- **Crotti Junior, A., Debruyne, C., Brennan, R. and O'Sullivan D. "An Evaluation of Uplift Mapping Languages".** *International Journal of Web Information Systems* **(IJWIS), 13, (4), 2017, pp. 405-424.**

- **Crotti Junior, A., Debruyne, C., Brennan, R. and O'Sullivan D. "FunUL: A Method to Incorporate Functions into Uplift Mapping Languages".** **In** *Proceedings of the 18th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2016).*

- **Crotti Junior, A., Debruyne, C. and O'Sullivan D. "Incorporating Functions in Mappings to Facilitate the Uplift of CSV Files into RDF". In** *Proceedings of the 13th Extended Semantic Web Conference: Posters and Demos, ESWC 2016.*

1.3. Technical Approach

Initially, a state-of-the-art review of existing visual representations of mappings in Linked Data was undertaken. This review was used to identify the capabilities of the state of the art in visual representations of mappings.

Having carried out these studies, requirements for a visual representation were identified:

R1. The visual representation should be expressive enough to support the representation of uplift and semantic mappings.

R2. The creation and editing of mappings should be supported through the visual representation.

R3. The visual representation should guide users in the creation and editing of mappings, only allowing for representations that generate valid mappings.

R4. Users should be able to create and edit mappings without being preoccupied with a particular mapping language, thus the visual representation should be independent of the underlying mapping language.

Juma was then proposed as a novel approach, based on the block metaphor, for the representation of mappings in Linked Data, which fulfils the aforementioned requirements. The block metaphor is capable of expressing information using a representation that takes advantage of the user's familiarity to jigsaw puzzles, which is a concept accessible for expert and non-expert users, which is the key motivation why it has been applied in this thesis in the representation of Linked Data mappings. The use of this metaphor also allows users to focus on creating meaningful mappings from source to target elements, and less on a particular mapping language's syntax. Moreover, studies have shown that the block metaphor has been successfully applied in other fields, such as querying Linked Data (Ceriani & Bottoni, 2017), programming robots (García-Zubía et al., 2018), and in the data science domain (Bart et al., 2017), especially for non-expert users, and that it has not yet been used to represent Linked Data mappings.

Three applications that apply the Juma approach have also been developed. The first application applied the Juma approach to the R2RML mapping language. This application, called Juma R2RML, reflects the R2RML's vocabulary in the representation of uplift mappings. Juma R2RML also guides users in creating mappings without them being preoccupied with R2RML's vocabulary or a particular RDF serialization. Even though, an obvious limitation is that it was tied to a particular uplift mapping language, the goal was to assess whether Juma was a suitable representation for mappings. We then proceeded with investigating a more abstract representation that would be able to support multiple uplift mapping languages. This led to the second application of the Juma approach, called Juma Uplift. Juma Uplift is able of generating mappings that are compliant with the R2RML and SML mapping languages. The third application showed that the Juma approach can also be used to represent semantic mappings that automatically generate mappings from its representation for interlinking datasets using SPARQL CONSTRUCT queries. General principles during the development of the applications Juma Uplift and Juma Interlink have been identified. These principles were used to represent similar concepts uniformly between these applications, which it is argued will facilitate the interpretation of the different types of mappings supported.

The evaluation of the Juma approach was undertaken through the execution of lab-based experiments and user experiments with the developed applications. The Juma approach was evaluated through two user experiments in the creation[9] of uplift mappings (Section 5.3.3 sand 5.3.4). The understanding of uplift mappings is evaluated through another user experiment (Section 5.4.3). Finally, the expressiveness of Juma in the representation of uplift and semantic mappings is evaluated through two lab-based experiments (Section 5.5.2 and 5.5.3).

The next section provides more detail of the evaluation strategy used.

1.4. Evaluation Strategy

The strategy used to evaluate the Juma approach was as follows:

- **Evaluate Juma for the <u>creation</u> of R2RML mappings:** This evaluation tested the Juma approach in the creation of R2RML mappings through a user experiment. In this experiment, the application Juma R2RML was used to evaluate the performance and usability of Juma for the task of creating uplift mappings for participants with different background knowledge. Results have shown that participants were able to use Juma with high performance and sufficient usability results.

- **Evaluate Juma for the <u>creation</u> of uplift mappings**: This evaluation tested the Juma approach in the creation of uplift mappings through another user experiment. In this experiment, the application Juma Uplift was used to compare the performance, usability and perceived mental workload when creating mappings using Juma to "manually" craft the uplift mappings in R2RML using the RDF TURTLE syntax. Participants of this experiment were split into two groups. One group was asked to use Juma Uplift to create one uplift mapping. The other group was asked to create the same mapping manually, in R2RML using TURTLE syntax, using their preferred text editor. It was found that participants using Juma Uplift had almost three times the performance achieved by participants that created the mappings manually. The usability was also higher for participants that used Juma Uplift and the mental workload was slightly smaller.

- **Evaluate Juma for the <u>understanding</u> of uplift mappings:** This experiment evaluated the understanding of uplift mappings representations through an online survey. Upon access to the survey, users were redirected to one of two possible questionnaires. Both questionnaires used the same database as input, mapping and

---

[9] Creation, as it is used in this thesis, also indicates the editing of mappings.

questionnaire questions. The difference was in the mapping representation. One questionnaire represented the mapping using Juma, through the Juma Uplift representation. The other had the same mapping represented in R2RML encoded with the RDF TURTLE notation. Participants were then presented with one multiple-choice question. This question asked participants to select all triples that the mapping representation would generate when executed. This experiment evaluated Juma through the performance of users in understanding the mapping representation, and the user's perceived mental workload. It was found that participants that answered the survey with the Juma representation had better performance and smaller perceived mental workload.

- **Evaluate the <u>expressiveness</u> of Juma in the representation of uplift mappings:** The aim of this evaluation was to show the expressiveness of the Juma approach in the representation of uplift mappings. This experiment shows that Juma, albeit first conceived for R2RML (with the Juma R2RML application), can generate accurate mappings represented in multiple uplift mapping languages (with the Juma Uplift application which generate mappings in R2RML and SML). For this experiment, 10 use cases for uplift mappings were defined based on the work presented in (J. Sequeda, Priyatna, & Villazón-Terrazas, 2012). An expected RDF output was also created for each use case. Juma Uplift was then used to generate a mapping for each use case. The R2RML and SML mappings generated by Juma Uplift were executed against its respective engine and compared to the expected RDF output. It was found that the generated RDF output was identical to its respective expected output.

- **Evaluate the <u>expressiveness</u> of Juma in the representation of semantic mappings:** The aim of this evaluation was to show that the Juma approach can also be applied to semantic mappings. This experiment evaluates the expressiveness of Juma in the representation of semantic mappings. For this experiment, the Juma Interlink application was used to represent 72 real world mappings between Linked Data datasets. These mappings were devised by the R2R Framework research (Bizer & Schultz, 2010). It was found that this application is able of representing the entire set of mappings.

1.5. Thesis Overview

The remainder of this thesis is structured as follows:

Chapter 2: Background

This chapter provides useful preliminary information for readers of this thesis. It begins with information about the mapping process in Linked Data. It then describes the types of mappings used in the Linked Data domain: semantic and uplift mappings. It also describes some of the many mapping languages available used to describe such mappings.

Chapter 3: State of the Art

This chapter provides a review of existing approaches that support the visual representation of mappings in Linked Data. These approaches are divided based on two main visual techniques found in literature: tree and graph representations. Note that some approaches offer additional visual representations, which are also discussed. The review presented in this chapter focuses on characteristics related to the research of this thesis, such as how mappings are expressed through visual representations, mapping types and mapping languages supported, amongst others.

Chapter 4: The Juma Approach

This chapter describes the requirements derived from the state of the art for a visual representation of mappings, and the **J**igsaw **Pu**zzles for Representing **Ma**ppings (Juma) approach. The chapter also presents the three applications of the Juma approach applied to mappings in Linked Data. The first two implementations apply the Juma approach to uplift mappings. The third application applies the Juma approach to semantic mappings.

Chapter 5: Evaluation

This chapter describes the experiments as outlined in Section 5.

Chapter 6: Conclusion

This chapter presents the key findings of the research described in this thesis. It discusses to what extent the research question of this thesis has been answered and the extent to which the research objectives have been met. Possible directions for further work related to the research in this thesis are also outlined.

2. Background

This chapter presents background information related to the research of this thesis to aid readers that are unfamiliar with the domain of visual representations and mappings in Linked Data. There is an assumption that the reader is familiar with RDF (Brickley & Guha, 2014), OWL (Hitzler et al., 2012) and SPARQL (Harris et al., 2013). Section 2.1 presents an introduction to visual representations. An overview of the mapping process in Linked Data is presented in Section 2.2. Section 2.3 describes uplift mappings. Section 2.4 presents semantic mappings. Section 2.5 ends the chapter with a summary.

2.1. Introduction

Visual representations support user involvement by representing data through visual elements (or objects) and text in order to convey information in an interpretable form (Ware, 2012).

Visual representations have been widely applied in the Linked Data domain. Editors such as Protégé[10] and WebVOWL (Lohmann, Link, Marbach, & Negru, 2015) apply visual representations to support the development of ontologies. Others, such as SparqlBlocks (Ceriani & Bottoni, 2017) and FedViz (Sana E Zainab et al., 2015), support the exploration of Linked Data datasets through the generation of SPARQL queries. unSCHACLed (De Meester, Heyvaert, Dimou, & Verborgh, 2018) is an editor that applies a visual representation for the definition of constraints used to validate RDF datasets, defined using the SCHACL Shapes Constraint Language[11]. Visual representations are also used to create and edit Linked Data mappings during a **mapping process**.

2.2. Mapping Process

The mapping process is concerned with all the activities executed by stakeholders in order to produce a set of relations, or mappings, between inputs (Sicilia et al., 2017). **Fig. 1** shows a high level view of such a mapping process (Debruyne, Walshe, & O'Sullivan, 2015).

---

[10] https://protege.stanford.edu/, accessed in August 2018.
[11] https://www.w3.org/TR/shacl/, accessed in August 2018.
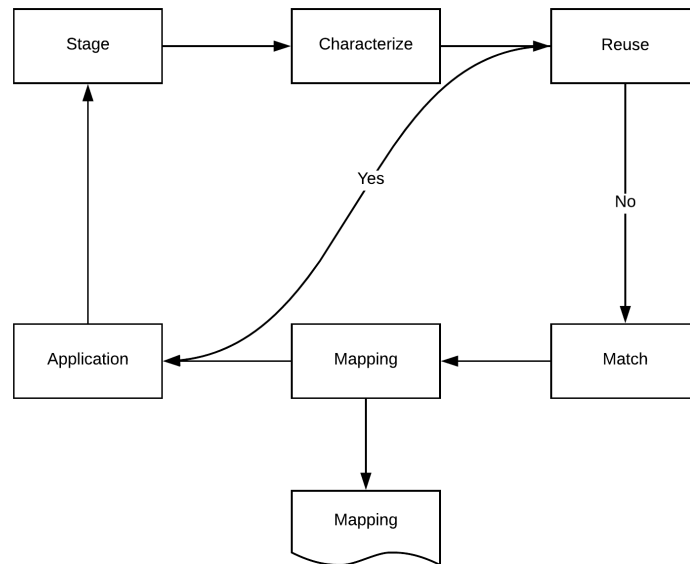
Fig. 1. Overview of a mapping process (Adapted from Debruyne et al., 2015).

In this mapping process, the **Stage** phase is concerned with identifying a community of stakeholders as well as agreeing on the scope of the project and its requirements. The **Characterize** phase is responsible for capturing the discovery and analysis of the inputs that are going to be mapped. **Reuse** utilizes the information described in the Stage and Characterize phases to discover and select possible existing mappings. The **Match** phase is concerned with the use of systems to find correspondences[12] that may be used in the mapping phase. **Mapping** is concerned with creating and storing a mapping. The **Application** phase uses the mapping created in the previous phase, as required by the stakeholders of the process. This phase may also identify issues with a mapping or the need for new mapping, triggering a new iteration of the process (Debruyne et al., 2015).

The mapping process presented in **Fig. 1** shows an overview of such a process that may be applied to both uplift and semantic mappings. We note that some activities in each phase may be specific to a certain type of mapping. **Fig. 2** presents a specialized mapping process that has been proposed in literature to capture the activities involved in the creation of semantic mappings.

---

[12] A correspondence express a semantic relationship between entities (Euzenat & Shvaiko, 2007).
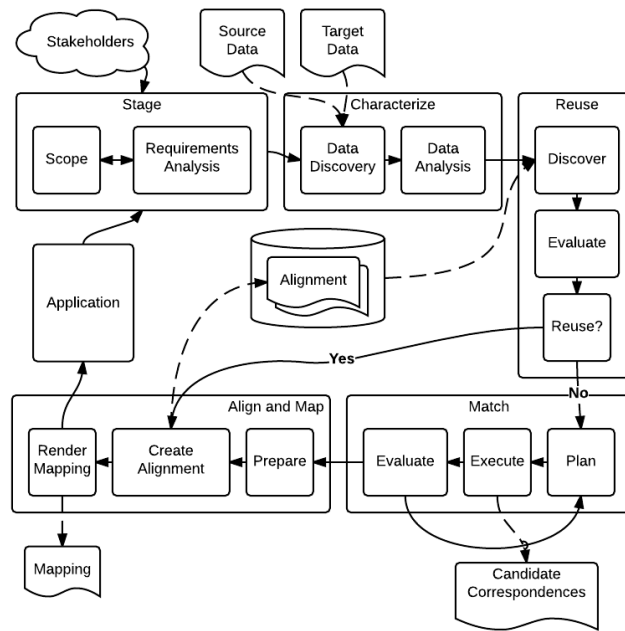
Fig. 2. Semantic mapping process (Debruyne et al., 2015).

In this mapping process, *Stage* identifies a need for semantic mappings, the stakeholders, scope and requirements. The *Characterize* phase is concerned with the discovery and analysis of the source and target ontologies that are going to be mapped. *Reuse* is concerned with activities related to discovering and reusing existing mappings. Often, ontology matching systems are used to discover correspondences – this occurs in the *Match* phase. These correspondences are used in the *Align and Map* phases to create alignments[13] – from which which semantic mappings can be distilled. Finally, in the *Application* phase, the semantic mapping created can be used and monitored – where a new iteration of the process may be triggered when necessary.

Fig. 3 presents a similar mapping process concerned with the creation of uplift mappings. The *Stage* activity is similar for both mapping processes, describing the need for mappings, stakeholders and requirements. *Characterize* is concerned with analysing and identifying the inputs that are going to be mapped as well as the ontologies or vocabularies that are going to be used. *Characterize* will also analyse tools that may be needed during the uplift process, for cleaning or transforming the data, for example. *Reuse* is concerned with finding, analysing and reusing mappings or parts of – called components in **Fig. 3**. The *Mapping* phase creates and assesses a mapping. Finally, the *Execution* phase generates, assesses and publishes a Linked Data dataset.

---

[13] Euzenat and Shvaiko (2007) describe alignments as sets of correspondences. A correspondence, as mentioned, express a semantic relationship between entities. A semantic mapping is the directed version of an alignment.

16
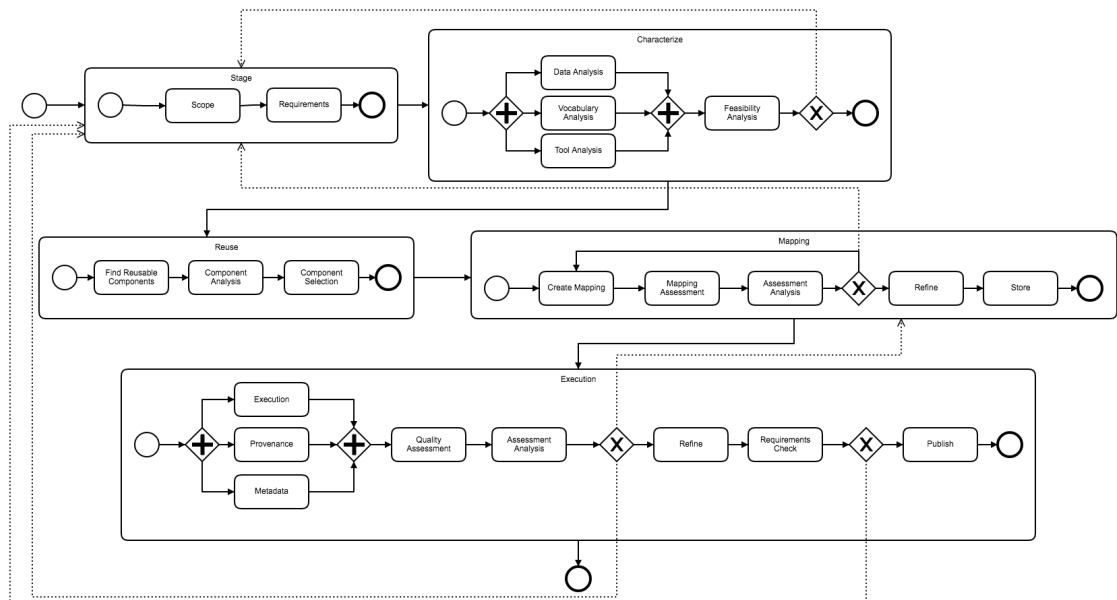
Fig. 3. Uplift mapping process.

The mapping processes described show the importance of user involvement, which is fundamental to a number of the activities outlined. For instance, users are responsible identifying and analysing the inputs to be mapped. Users must also analyse and validate existing mappings so that these may be reused. The creation and assessment of new mappings – when necessary – is also conducted by users, and so on. As stated in Chapter 1, such user involvement has been identified as one of the main challenges for producing quality semantic mappings (Shvaiko & Euzenat, 2013) and uplift mappings (J. F. Sequeda & Miranker, 2017). It is also noted that often the same stakeholders may be necessary for producing semantic and uplift mappings (Marshall et al., 2012). As an example, consider a mapping process concerned with publishing historical data, where uplift mappings must be created in order to transform non-RDF data to RDF. The stakeholders of such a process may have different background knowledge depending on their area of expertise, being web developers, Semantic Web experts and non-expert users, such as historians. These same stakeholders may want to consume and integrate data from historical Linked Data datasets, where another mapping process would be conducted to produce semantic mappings.

This thesis is concerned with the use of visual representations to support user involvement and alleviate the knowledge required for producing Linked Data mappings. The remainder of this chapter provides a brief description of the mappings that can be produced in the Linked Data domain, namely uplift and semantic mappings, and a number of the mapping languages available to formally express them. Chapter 3 will present the state-of-the-art approaches for the visual representation of mappings in Linked Data.

2.3. Uplift Mappings

Uplift mappings are concerned with expressing the transformations needed in order to represent non-RDF data as RDF (Bizer & Seaborne, 2004). The mapping process that produces uplift mappings has non-RDF data and ontologies as their input, with the output being mappings that express relations between such inputs (Crotti Junior, Debruyne, Brennan, & O'Sullivan, 2017).

This section briefly describes a number of uplift mapping languages that are popular in the research community and relevant to the research of this thesis.

2.3.1. D2R

D2R server (Bizer & Cyganiak, 2006)[14] is described as a tool for publishing relational databases as Linked Data. The D2R server supports R2RML (which will be described in Section 2.3.3) and the D2RQ mapping language. Listing 3[15] presents the mapping shown in Listing 2 in R2RML (Chapter 1) in D2RQ. The main components of a D2RQ mapping are the *class map*, which express URI pattern of instances, their types, and a set of *property bridges*, which relates a column to predicate. Property bridges may also define language tags, datatypes, SQL joins, amongst others.

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix d2rq: <http://www.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/0.1#> .

:personMap a d2rq:ClassMap;
    d2rq:dataStorage :database;
    d2rq:class foaf:Person;
    d2rq:uriPattern "http://example.org/@@person.id@@";
.
:namePropertyMap a d2rq:PropertyBridge;
    d2rq:belongsToClassMap :personMap;
    d2rq:property foaf:name;
    d2rq:column "person.name";
.
```

Listing 3. Uplift mapping represented in D2RQ

2.3.2. Direct Mapping

The Direct Mapping (DM) W3C Recommendation (Arenas, Bertails, Prud'hommeaux, & Sequeda, 2012) defines how relational databases can be represented in RDF. The DM approach does not consider existing vocabularies. Instead, the process generates a

---

[14] The D2R server is a popular tool with over 400 citations on Google Scholar.
[15] This example omits database connection properties, which are expressed as part of the mapping in D2RQ.

vocabulary based on the relational database schema. The DM approach can be summarized as follows:

- Table to Class: each table from a relational database is translated into a Class.
- Column to Property: columns of each table are translated to predicates named after the column name.
- Row to Resource: each row is translated to a resource. These are also defined as instances of the Class that represents the table.
- Column value to Literal: each column value (that is not a foreign key) is translated to a Literal object, with the predicate being the property represented by the column name.
- Foreign key to Resource: each foreign key is translated to a Resource. The predicate is also represented by the column name.

2.3.3. R2RML and its extensions

The RDB to RDF Mapping Language (R2RML) (Das et al., 2012) is also a W3C Recommendation for the transformation of relational databases to RDF. R2RML allows for the expression of customized mappings, enabling the reuse of existing vocabularies. Mappings expressed using R2RML are stored as RDF documents. Therefore, it is possible to query and annotate such mappings with additional information (e.g. provenance). An example of an R2RML mapping was presented in Chapter 1 (Listing 2). R2RML mappings consist of one or more triples maps. Each triples map has one *logical table*, one *subject map* and zero or more *predicate object maps*. *Graph maps* may be used in *subject maps* or *predicate object maps* to assign triples to named graphs.

- **Logical Table**. The table, view, or SQL query from which RDF will be generated.
- **Subject Map**. Define the subjects of the RDF triples. These subjects can be IRIs or blank nodes. You may also subjects to be instances of zero or more class types.
- **Predicate Object Map**. Define the predicates, using predicate maps, and objects, using object maps, of the RDF triples. Each predicate object map must have at least one predicate map and one object map. Predicates must be valid IRIs. Objects can be IRIs, blank nodes or literals. For literal values, it is possible to define a data type or a language tag. You may link triples maps using parent triples map. A parent triples map can have zero or more join conditions.
- **Graph map**. Graph maps are used to assign triples to (named) graphs. These may be used in subject maps or in predicate object maps. Let *X* be the set of graph maps of

a subject map. If X is not empty, then all `rdf:type` assertions will be stored in all graphs in X. Otherwise they are stored in the default graph. Let Y be the set of graph maps of a predicate object map. If the union of X and Y is not empty, then all triples generated by the predicate object map are stored in all graphs of the union. Otherwise they are stored in the default graph.

Many extensions have been proposed to R2RML. RML (Dimou et al., 2014) extends the language's vocabulary in order to support a wider set of input formats, such as CSV, XML, amongst others. xR2RML (Michel, Djimenou, Faron-Zucker, & Montagnat, 2015) has also extended the language to support other input formats, including NoSQL databases. R2RML-F (Debruyne & O'Sullivan, 2016) added the support for the definition of data transformation functions as part of the mapping.

2.3.4. SML

The Sparqlification Mapping Language (SML)[16] (Stadler et al., 2015) is an uplift mapping language to transform relational databases to RDF. SML is based on SQL CREATE VIEWS and SPARQL CONSTRUCT queries. The mapping presented in Listing 2 (Chapter 1) in R2RML (RDF TURTLE notation) is presented in Listing 4 in SML. An SML mapping is composed of the following parts:

- **Construct.** Consists of triple patterns, similar to SPARQL CONSTRUCT queries. These are used as templates for the construction of the RDF triples.
- **With.** This clause is used to specify variables whose values are RDF terms from rows of the logical table. These variables may be used in the construct clause to form RDF triples.
- **From.** Define the logical table. As in R2RML, it may be a table, view, or a SQL query.

---

[16] SML is presented here for differing from the D2R and R2RML approaches, where mappings are expressed as RDF graphs. SML is based on SPARQL and has been used in several projects, such as `linkedgeodata.org`.

```
prefix foaf: <http://xmlns.com/foaf/0.1/>

Create View view1 As
 Construct {
    ?s1  a foaf:Person.
    ?s1   foaf:name   ?o1. }
 With
  ?s1 = uri(concat('http://example.org/', ?id))
  ?o1 = plainLiteral(?name)
 From
  Person
```

Listing 4. Example of an uplift mapping represented in SML

2.4. Semantic Mappings

Semantic mappings express relations between semantically similar concepts found in different ontologies (Euzenat & Shvaiko, 2007). Source and target ontologies are the inputs of the mapping process that produces semantic mappings. Semantic mappings can be categorized as simple and complex. Simple mappings relate one entity to another (one-to-one); complex mappings describe relationships between multiple entities (one-to-many, many-to-one, and many-to-many) (Euzenat & Shvaiko, 2007).

Even though the Alignment Format and EDOAL do not represent mappings, these are described here for being a popular format for expressing alignments[17]. Nonetheless, alignments may be rendered as mappings using SWRL or SPARQL, which are also presented in this section.

2.4.1. Alignment Format and EDOAL

The Alignment Format (AF) (Euzenat, 2004) is designed to represent simple alignments, which can be encoded in XML or RDF. The Expressive and Declarative Ontology Alignment Language (EDOAL) (David, Euzenat, Scharffe, & dos Santos, 2011) was introduced as an extension of the Alignment Format, in order to represent complex alignments. Listing 5 presents the semantic mapping from Listing 1 (Chapter 1) using the Alignment Format.

---

[17] The Alignment Format and EDOAL have over 600 citations combined on Google Scholar.

```
<Alignment>
  <xml>yes</xml>
  <level>0</level>
  <type>**</type>
  <map>
      <Cell>
          <entity1 rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
          <entity2
rdf:resource="http://dbpedia.org/ontology/Person"/>
          <measure rdf:datatype="xsd:float">1.0</measure>
          <relation>=</relation>
      </Cell>
  </map>
</Alignment>
```

Listing 5. Example of a semantic mapping represented using the Alignment Format

2.4.2. SWRL

The Semantic Web Rule Language (SWRL) (Horrocks et al., 2004) is a language that can express rules on RDF data. SWRL is defined as a combination of OWL-Description Logics (OWL-DL), OWL Lite and a subset of the Rule Markup Language[18] (RuleML). One of the possible uses for SWRL is the representation of semantic mappings. For example, Listing 6 shows the semantic mapping from Listing 1 in SWRL.

```
<swrlx:Ontology swrlx:name="generatedAl"
                xmlns:swrlx="http://www.w3.org/2003/11/swrlx#"
                xmlns:owlx="http://www.w3.org/2003/05/owl-xml"
                xmlns:ruleml="http://www.w3.org/2003/11/ruleml#">
    <ruleml:imp>
      <ruleml:_body>
        <swrlx:classAtom>
          <owlx:Class owlx:name="http://xmlns.com/foaf/0.1/Person"/>
          <ruleml:var>p</ruleml:var>
        </swrlx:classAtom>
      </ruleml:_body>
      <ruleml:_head>
        <swrlx:classAtom>
          <owlx:Class owlx:name="http://dbpedia.org/ontology/Person"/>
          <ruleml:var>p</ruleml:var>
        </swrlx:classAtom>
      </ruleml:_head>
    </ruleml:imp>
</swrlx:Ontology>
```

Listing 6. Example of a semantic mapping represented using SWRL

2.4.3. SPARQL CONSTRUCT queries

The SPARQL Protocol and RDF Query Language (SPARQL) (Harris et al., 2013) is the W3C Recommendation for querying RDF data. SPARQL CONSTRUCT queries have been

---

[18] http://ruleml.org, accessed in August 2018.

proposed to represent executable mappings, which means that they are directly executable by any processor that supports the SPARQL query language (Rivero & Hernández, 2011). SPARQL CONSTRUCT queries can convert data represented in one vocabulary to another, and support a wide range of functions, allowing for the representation of complex data transformations (Rivero & Hernández, 2011). An example of a semantic class mapping represented using a SPARQL CONSTRUCT query was presented in Listing 1 (Chapter 1).

2.5.   Chapter Summary

The aim of this chapter was to help readers, who may be unfamiliar with mappings in Linked Data, gain an understanding of the processes involved in their creation and editing for the different types of mappings found in the domain.

The mapping process was briefly described, which is concerned with how and in what cases mappings can be used in the domain of Linked Data. The mappings found in the Linked Data domain were also presented based on the type of transformation that they represent, named uplift and semantic mappings. These types of mappings were also briefly described, together with examples of mapping languages used to formally express them.

3. State of the Art

This chapter presents the state of the art of existing approaches for visually representing mappings in Linked Data. Section 3.1 presents an introduction of the chapter. Section 3.2 briefly describes the field of information visualization, which is the area concern with interactive visual representations. Section 3.3 presents 12 approaches that primarily apply graph-based visual representations. Section 3.4 presents 9 approaches that primarily apply tree-based visual representations. The analysis of the state of the art is presented in Section 3.5. Section 3.6 finishes this chapter with a summary.

3.1.   Introduction

The purpose of this chapter is to provide a state-of-the-art review of approaches that support the visual representation of mappings in Linked Data. As stated previously, visual representations may be used during the mapping process to support user involvement, alleviating the knowledge required for producing uplift and semantic mappings, thus facilitating its creation, editing and understanding (Granitzer et al., 2010).

The research in this thesis is focused on to what extent visual representations may facilitate the interpretation of mappings for generating and interlinking Linked Data. Therefore, the key characteristics being reviewed in this chapter are related to how mappings are expressed through visual representations (visual technique), what can they express (mapping type and mapping languages), and how users can interact with it (editing approaches). User experiments conducted to evaluate existing approaches are also reported upon.

The review presented in this chapter is based on publications, and – when available – working implementations. If a working implementation was available at the time of writing, then the section has a reference or pointer to the implementation used.

3.2.   Information Visualization

Information visualization (InfoVis) is defined as the use of interactive visual representations to amplify cognition (Ware, 2012). InfoVis combines several disciplines, such as cognitive psychology, human-computer interaction, and computer science, with the goal of improved understanding (Bederson & Shneiderman, 2003). InfoVis leverages the human's most powerful perception channel – the visual system – in order to expand working memory

capacity and "offload work from cognitive to perceptual system" (Card, Mackinlay, & Shneiderman, 1999).

In InfoVis , visual representations are used to make higher order relations more accessible to human intuition (Tufte, 1990). Visual representations serve as external memory, supporting the exploration of unknown or complex information (Munzner, 2014). For instance, data presented in a plot may make its characteristics more explicit when compared to a table of values, even though the same data is being presented.

Visual representations combine graphic visual elements and text with different characteristics (colour, shape, amongst others) to convey information in a interpretable form (Ware, 2012). Interacting with visual representations is also essential to facilitate understanding and overcome limitations of dealing with this unknown or complex information (Liu & Stasko, 2010).

Many aspects can be considered when developing visual representations. For instance, the visibility of relevant objects and actions may aid users in performing certain tasks (Preece, Rogers, & Sharp, 2001). Consistency within and across applications, such as in interaction style and uniform representation of similar concepts, may facilitate understanding and improve learnability (Shneiderman et al., 2016). Certain properties of graphic visual elements may be used to indicate its possible actions (Norman, 2013). Visual representations should also guide user's actions, prevent mistakes, and maintain the user's focus while performing certain tasks (Norman, 2013). These characteristics make interfaces easy to use or intuitive. An intuitive interface works as the user expects it to (Nielsen, 1994).

Visual representations may apply metaphors in order to exploit specific knowledge that users already have (Ware, 2012). Metaphors make use of familiar concepts in order to aid users in the understanding of another – unknown or complex – concept (Ziemkiewicz & Kosara, 2008). For example, a number of approaches presented in this chapter apply a tree metaphor in the representation of ontologies, data and mappings. Visual representations that rely on the tree metaphor start on an element (root) that is branched to related lowest level elements (leaves) (Schulz, 2011). Users of such visual representation are already familiar with how this representation works, as it is commonly used for navigation and representing the folder hierarchy of file systems (Ware, 2012).

This chapter reviews the state of the art approaches that apply visual representations to support the mapping process, being for editing, viewing or both. Existing visual representations for mappings in a Linked Data context can be classified at a high level as tree and graph representations, which will be presented in Section 3.3 and 3.4, respectively.

3.3. Graph-based visual representations

This section reviews 12 existing systems from the state of the art that apply primarily graph-based visualization representations. Note that some approaches offer additional visual representations for ontologies, input data, and/or mappings.

Graph-based visual representations are based on node-link diagrams. These diagrams represent concepts using nodes, while relations between these nodes are represented using edges (Ware, 2012). In the Linked Data domain, these graph representations often have labeled directed edges, which illustrate the type and direction of the relation. (Ware, 2012). Graph-based visual representations support the declaration of any type of relationship between nodes. Nonetheless, this visualization technique is more influenced by size and density (number of nodes and edges) than other visual representations (such as tree-based visual representations which will be presented in Section 3.4) (Ghoniem, Fekete, & Castagliola, 1997).

The remainder of this section analyses existing approaches which are relevant to the research of this thesis – in the view of the author – that apply primarily graph-based visual representations to Linked Data mappings.

3.3.1. AlViz

AlViz (Lanzenberger & Sampson, 2006) is a Protégé plugin developed to support the visualization of semantic mappings. AlViz provides a graph visual representation of mappings showing source and target ontologies in different panels. The tool also shows the ontologies being mapped through Protégé's class browser, which is pre-installed. In the graph representation, nodes are clustered together according to a selected relationship, and the size of the nodes depends on the number of entities in the cluster. Mappings are represented using different colour nodes in the visualization panels. For example, red nodes represent equal concepts, green represents similar concepts, syntactically equal entities are orange, yellow nodes represent entities that are unique in one of the ontologies, broader concepts are blue and narrower purple. **Fig. 4** shows AlViz's visual representation of mappings.
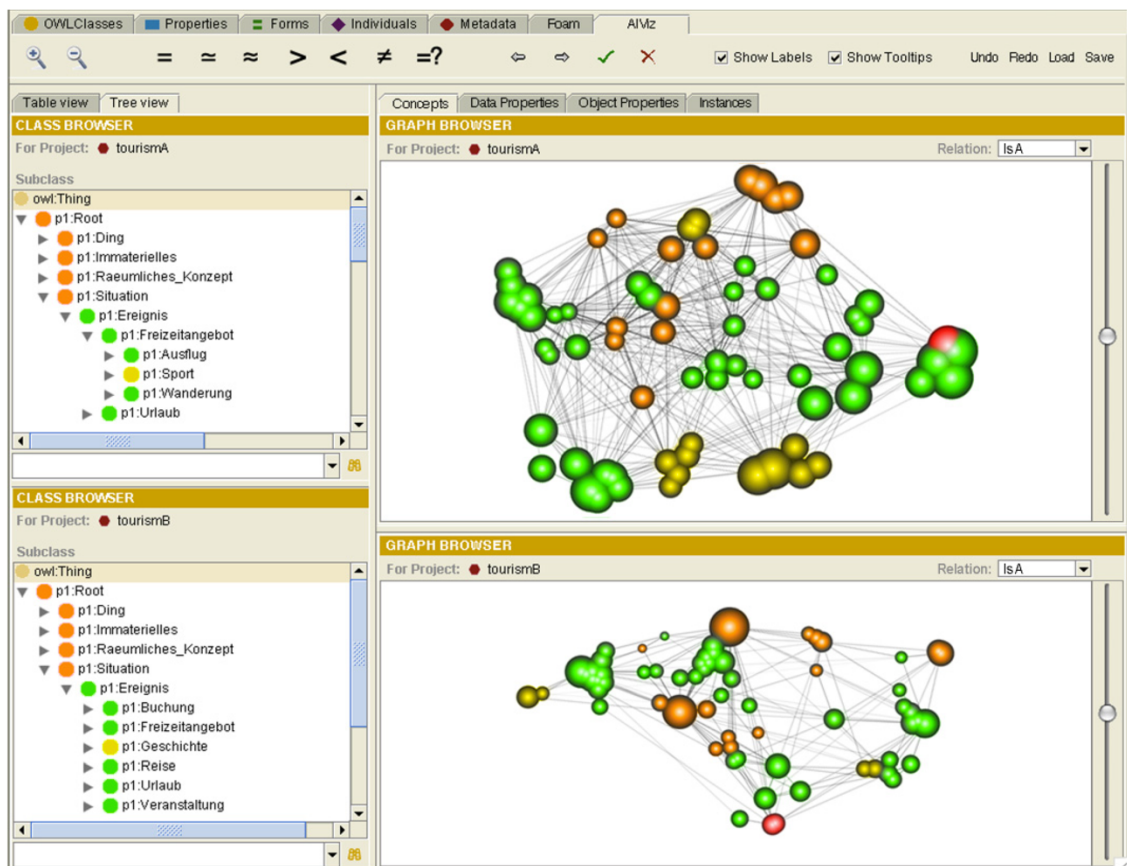
Fig. 4. AlVIz's visual representation of mappings (Lanzenberger & Sampson, 2006)

The tool uses a technique called linking and brushing which, as implemented in AlViz, highlights the related concepts in the tree visualization once a user selects elements from the graph representation. The tool supports the creation and editing of mappings by selecting elements from the tree or graph visualizations, and a relation which is presented in a separate menu. AlViz only supports simple semantic mappings, which can be exported using its own XML schema representation. To the best of our knowledge, there are no reports of user evaluations of the tool or its visual representation. The system was not available at the time of writing.

3.3.2. Agreement Maker Light

The Agreement Maker Light (AML) (Faria et al., 2013) is an ontology matching system that supports semantic mappings. AML is still in active development, participating yearly in the contest organized by the OAEI campaign.

AML was derived from the Agreement Maker framework (Cruz, Antonelli, & Stroe, 2009). The Agreement Maker relies on memory-intensive computations to find the similarities between ontologies. AML, on the other hand, has been designed to handle large ontologies and is based on lexical matching techniques and on the use of external

background knowledge. Moreover, the authors describe the system as flexible and extensible, allowing for the integration of any ontology matching algorithm. **Fig. 5** shows the visual representation for a semantic mapping in the AML system.
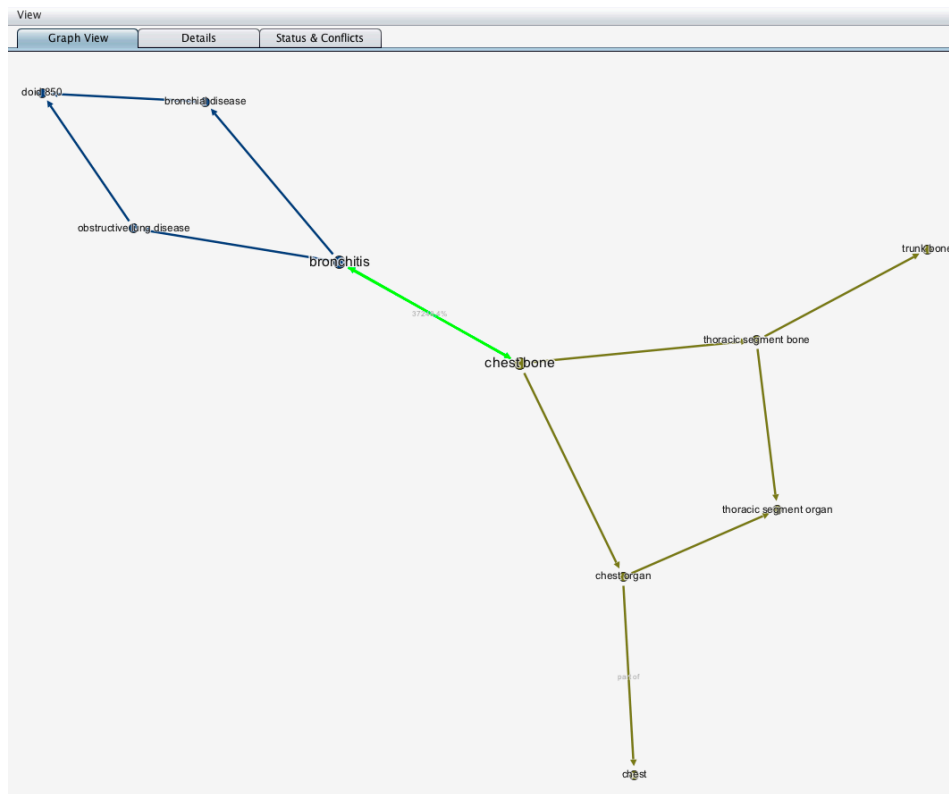
Fig. 5. AML's visual representation of mappings

Colour coding is used to differentiate the ontologies in the graph, while a green edge is used to represent a mapping. Users may also select the number of ancestors and descendants related to the concepts being mapped. The graph representation, however, does not allow users to reshape or move nodes to analyze details of the mappings. Despite representing mappings visually, the system does not allow for the creation or editing of these through its visual representation, and instead relies on a form-based interface for these tasks. The output format supported is the Alignment Format, which means that the system only supports simple semantic mappings. AML, as many other matching systems, focuses on the accuracy of the matching process, where usability evaluations of the approach are often neglected. The system for which this review was based is available[19].

---

[19] `https://github.com/AgreementMakerLight/AML-Project/releases/tag/v3.1`, accessed in August 2018

3.3.3. OWL Lite Alignment

OWL Lite Alignment (OLA) (Euzenat, Loup, Touzani, & Valtchev, 2004) is dedicated to the mapping of ontologies expressed in OWL, with an emphasis on the OWL-Lite sublanguage (one of the tree "profiles" of the first OWL specification). In the OLA system, mappings are created by measuring the similarity of entity pairs. This similarity measure depends on all the similarities of neighbour pairs whose members describe the respective initial entities. The system also applies a lexical similarity mechanism, based on the WordNet database (Miller, 1995), for each term found in the ontologies. **Fig. 6** shows the mapping interface in the OLA system.
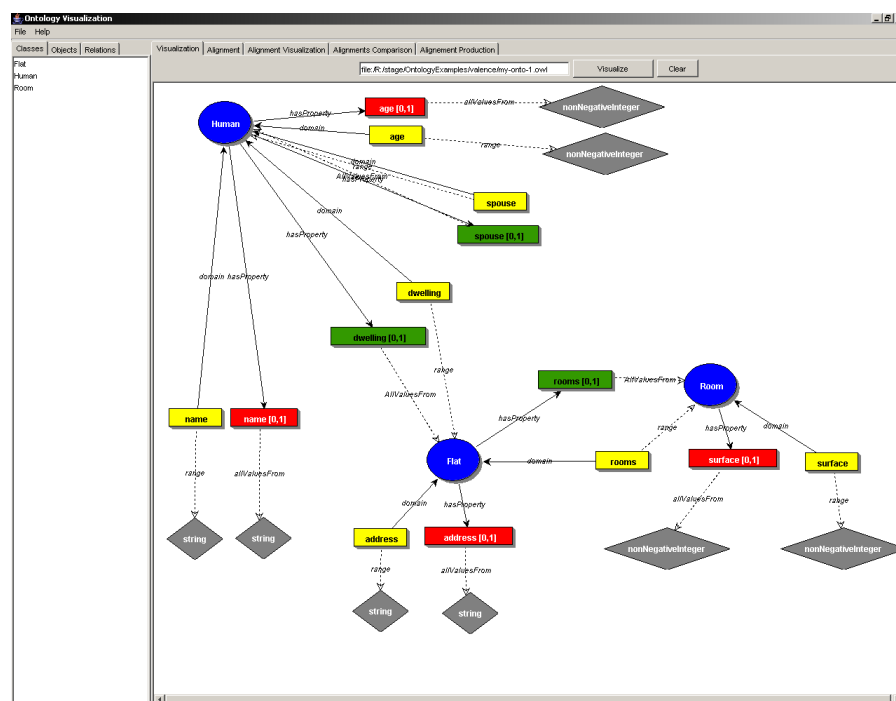
Fig. 6. OLA's visual representation of mappings[20]

The ontologies being mapped and the mappings between these ontologies are presented using a graph representation. Similar to the AML system, modifications of the mappings are done through a separate form-based interface. The system also only supports simple semantic mappings through the Alignment Format. Moreover, there are no reports of user experiments evaluating their interaction with the system or with the graph visual representation. The system was not available at the time of writing.

---

[20] From `https://www.iro.umontreal.ca/~owlola/visualization.html`, accessed in August 2018.

3.3.4. Repair of Ontological Structure Environment

The Repair of Ontological Structure Environment (RepOSE) (Lambrix & Ivanova, 2013) is an ontology alignment system with a debugging component for detecting and repairing semantic mappings. RepOSE supports taxonomy ontologies, which are ontologies containing classes and subsumption axioms only. This means that RepOSE only supports a specific case of semantic mappings. RepOSE works in three steps: generating mapping suggestions, validation and repairing. **Fig. 7** shows the mapping interface in RepOSE.
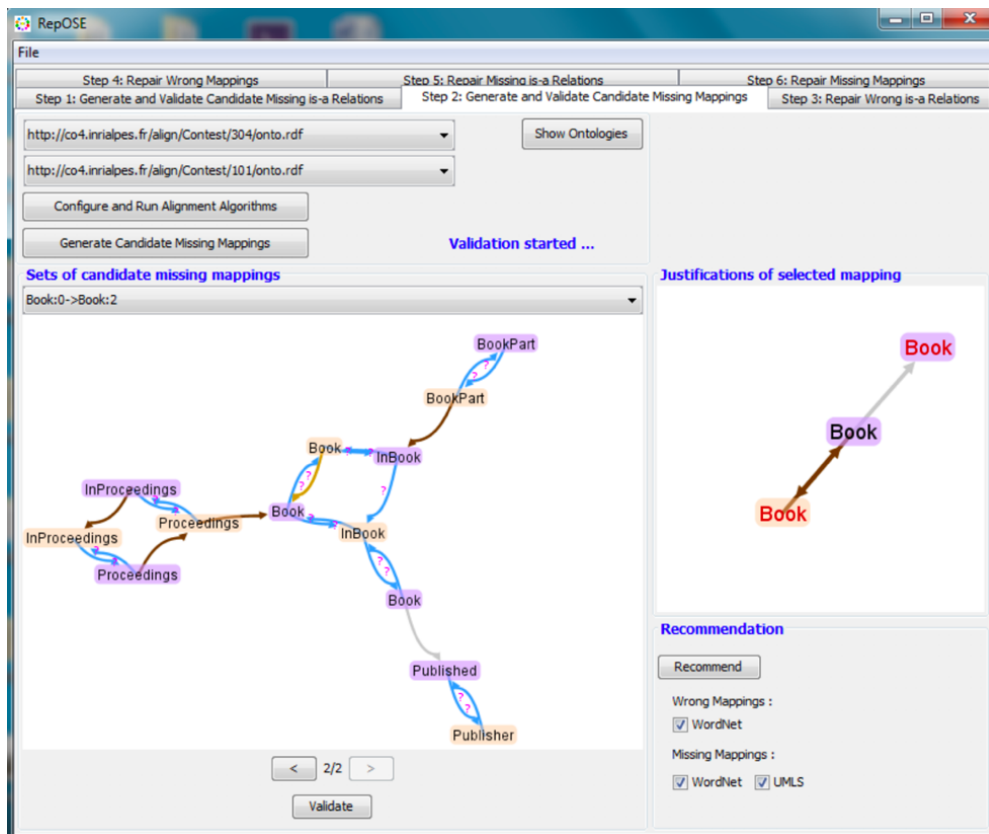
Fig. 7. RepOSE's visual representation of mappings (Lambrix & Ivanova, 2013)

During the validation phase, mapping suggestions are presented as graphs. Nodes are colour-coded according to the ontologies being mapped. Edges are also represented using different colours, where grey edges represent asserted relations, mappings are brown, missing relations are blue, and already repaired relations are black. The graph visualization only presents a fragment of the ontologies being mapped, which contain the mapping being validated or repaired. Modifications in the mapping are done by selecting nodes in the representation, and mapping options which are available in a separate menu. The system works with simple semantic mappings. There were two user experiments reported in (Lambrix & Ivanova, 2013). A domain expert used the system to repair ontologies in different domains in each experiment, where the system was found useful. Nonetheless, even

though the representation only shows part of the mapping, it was reported by users that the graph representation had too many elements in some cases. The paper does not mention how mappings are serialized, and the system was not available at the time of writing.

3.3.5. DataLift

DataLift (Scharffe et al., 2012) is a platform that supports the conversion of non-RDF data to RDF. The tool supports CSV, XML and relational databases. The uplift process in DataLift is done through a step-by-step process. In the first step, the non-RDF data is imported and converted to RDF. This first conversion does not take into account the re-use of existing vocabularies, and is described as a "raw" RDF conversion. A raw RDF conversion, as the authors call it, more or less corresponds to a direct mapping (Chapter 2) for relational data. The second step requests vocabularies that will be used to map the data, and suggests others, based on the Linked Open Vocabularies[21] project. Also in this step, users must define the desired RDF output, by relating classes and properties to the source data, through a form-based interface. These relations generate SPARQL CONSTRUCT queries that are used to modify the raw RDF data generated in the first step. The next step shows a graph visualization of the mapping, which is shown in **Fig. 8**.
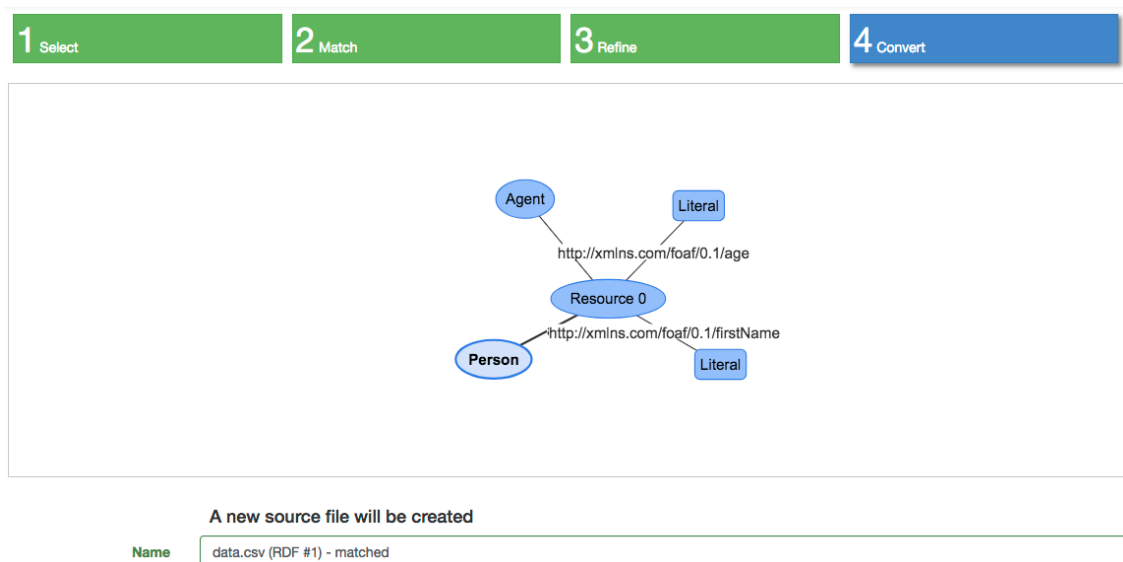
Fig. 8. DataLift's visual representation of mappings

The visual representation of mappings is shown as a graph and does not allow for the creation and editing of mappings. Similar to other tools presented in this chapter, mapping adjustments must be done through a form-based interface. The system also does not support

---

[21] https://lov.linkeddata.es/, accessed in August 2018.

the serialization of the mappings, hence, sharing and reuse are not supported. Data transformation functions are supported through SPARQL CONSTRUCT queries, but these transformations are not represented visually. To the best of our knowledge, there are no user evaluations of the system or its visual representation. The system for which this review was based is available[22].

3.3.6. Karma

Karma (Knoblock et al., 2012) is a web-based application for mapping heterogeneous data to RDF. Karma uses a data centric approach to support the mapping process. In this approach, data needs to be loaded into the tool before mappings can be created. The data then is represented in a tabular format, and the ontologies being mapped are presented using a tree structure. The data centric approached used by Karma makes defining relations between the different inputs complex. The mapping is represented as a graph above each table, which connects elements from the ontologies to elements in the input data. As the representation expands, it quickly becomes cluttered, since nodes are not clustered. **Fig. 9** shows the mapping interface of the Karma approach.
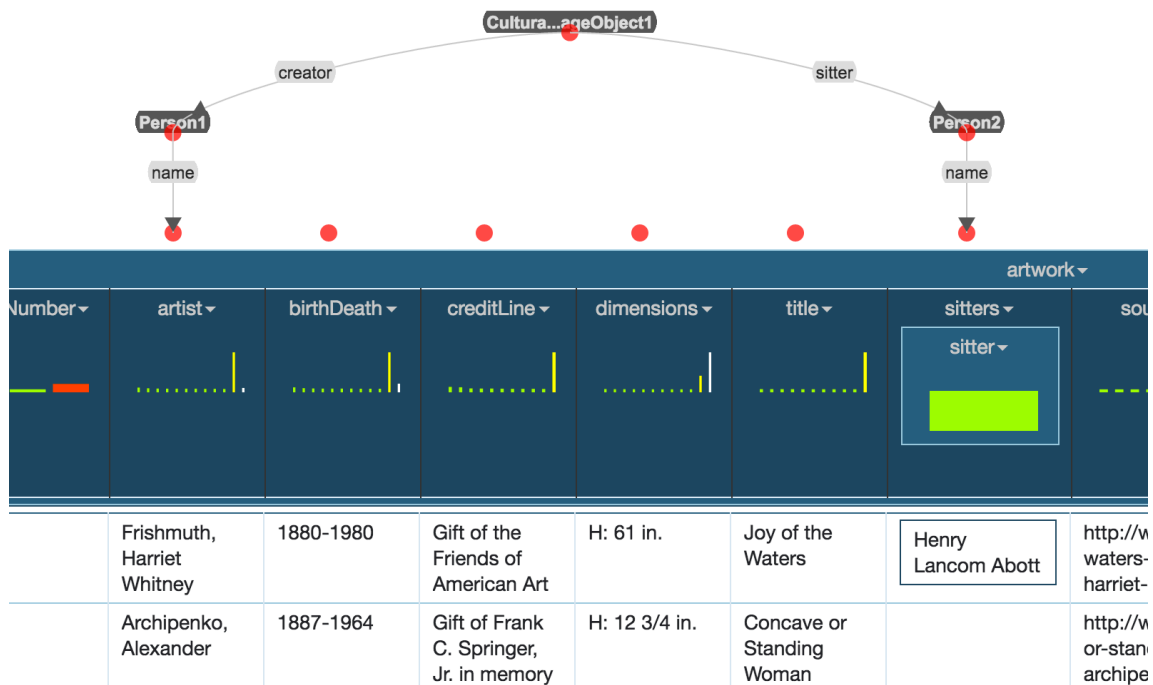
Fig. 9. Karma's mapping interface[23]

An extension of the Karma application has been presented in (Slepicka, Yin, Szekely, & Knoblock, 2015). This extension added support for data transformation functions through a form-based interface, without changing the visual representation of mappings. These functions are defined using the Python programming language. Thus, users need knowledge of Python to be able to use this feature. The graph visualization, however, does not support the visualization of these transformation functions. Karma supports the serialization of mappings using the R2RML mapping language. The Karma extension, which supports data transformation functions, extended the R2RML mapping language to support the serialization of transformation functions. This extension was called KR2RML, for which the authors provide a KR2RML processor. No user evaluations have been reported on the use of Karma's approach or its visual representation of mappings. The system for which this review was based is available[24].

3.3.7. Lembo et. al.

The research of Lembo et al. (Lembo, Rosati, Ruzzi, Savo, & Tocci, 2014) proposed an editor for the creation of uplift mappings. The tool supports uplifting data from relational databases to RDF. The tool also checks the semantics of the mapping, i.e. mappings are logically validated based on the ontologies and vocabularies used to annotate the data. This semantic validation is only available for OWL DL-Lite or OWL 2 QL ontologies. The tool uses a graph visual representation with three different levels of details: mapping-centered (**Fig. 10**), ontology-centered (**Fig. 11**), or source-centered (**Fig. 12**). The mapping-centered view shows the mapping assertions related to a user selected mapping from the list of mappings. The ontology-centered view shows the mapping assertions related to a user selected element from the ontologies list. Finally, the source-centered view shows the mapping assertions related to a user selected table from the input database schema.

---

[24] `https://github.com/usc-isi-i2/Web-Karma`, accessed in August 2018.
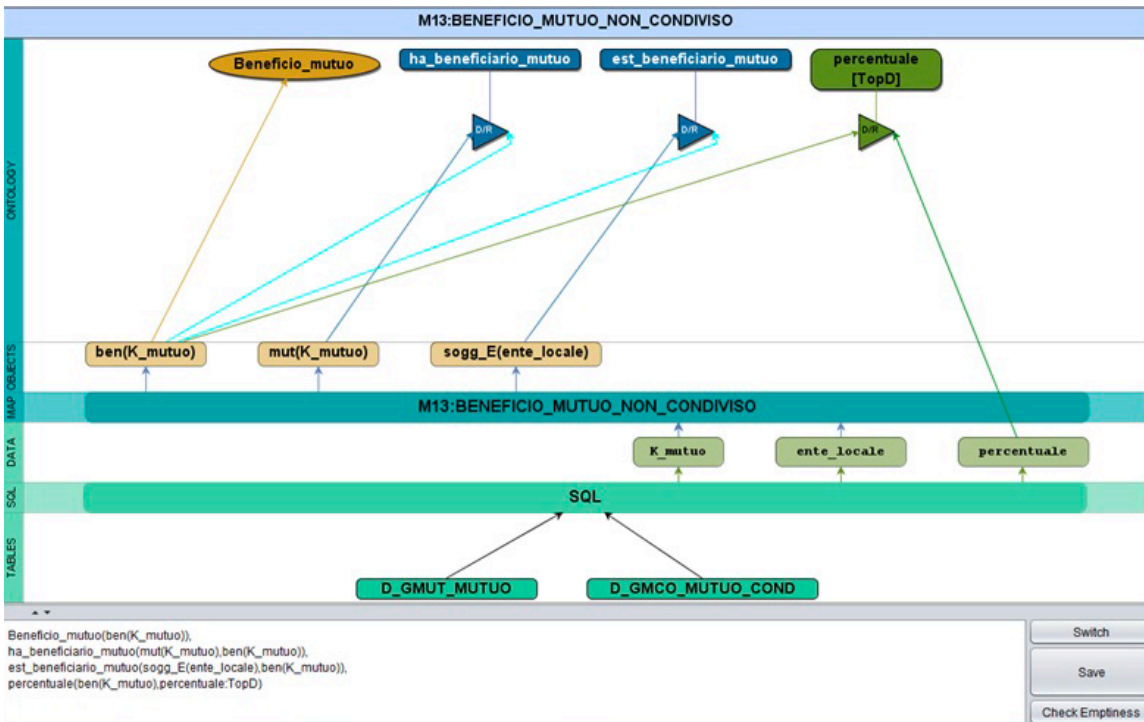
Fig. 10. Lembo et. al.'s mapping-centred visual representation (Lembo et al., 2014)
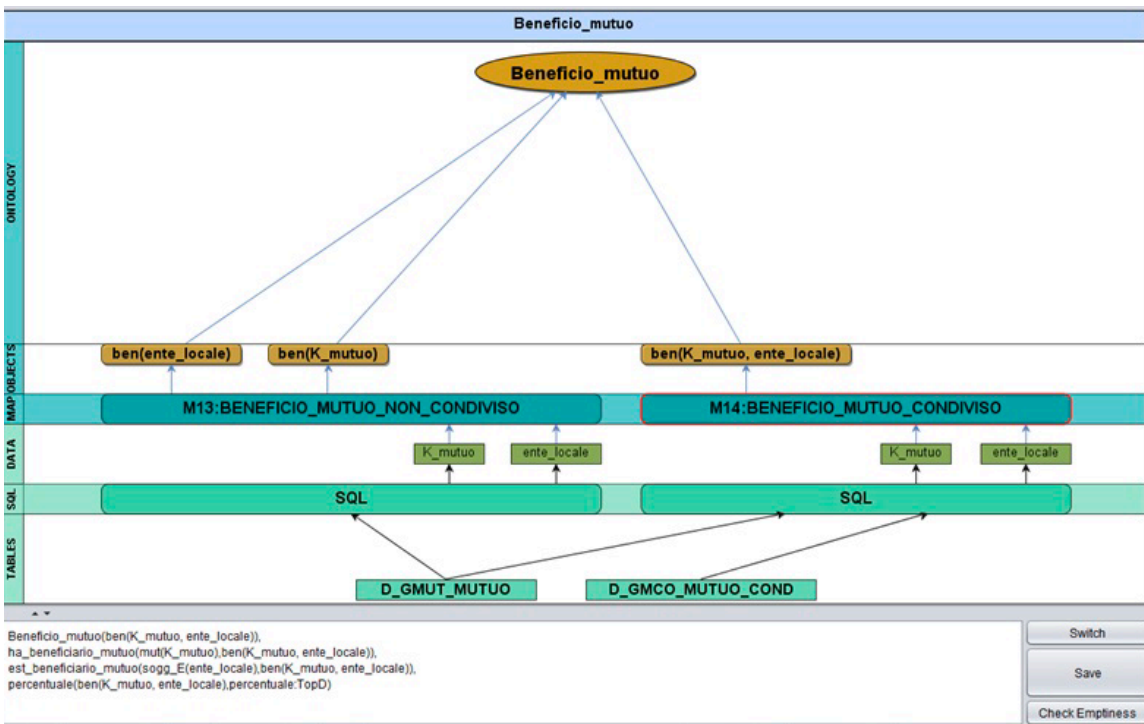
Fig. 11. Lembo et. al.'s ontology-centred visual representation (Lembo et al., 2014)
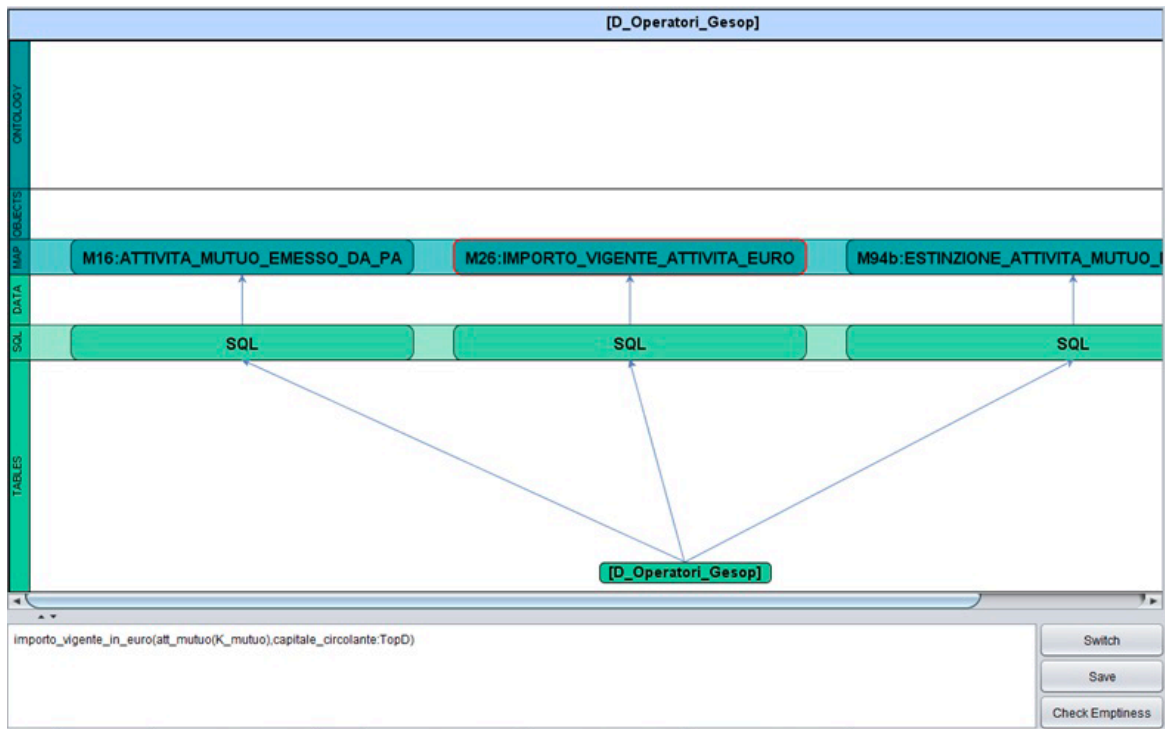
Fig. 12. Lembo et. al.'s source-centred visual representation (Lembo et al., 2014)

The creation and editing of mappings can be done through a form-based interface. These modifications must be done through the use its proprietary language, which is described as a subset of the R2RML mapping language. The manual editing of the mapping rules may make the process prone to error and more complex than one would like. Nonetheless, these modifications can be validated in the visual representation. Despite using their own mapping language, the tool allows for the serialization of mappings using the R2RML mapping language. The tool does not support data transformation functions, and no user evaluations have been reported. The system was not available at the time of writing.

3.3.8. SQuaRE

SPARQL Queries and R2RML mapping Environment (SQuaRE) (Blinkiewicz & Bak, 2016) is a tool that provides a visual environment for the creation of R2RML mappings. The tool also offers the possibility of querying these mapping (which are stored as RDF documents) using SPARQL. The creation of SPARQL queries is supported through a text-based interface. The results of the execution of ASK and SPARQL CONSTRUCT queries are represented visually, while other query results are represented as text. SQuaRE's interface to map relational databases to RDF shows the tables of the input database on the left of the screen. The ontologies are represented on the right as trees. Users can drag the classes and properties from the ontology visualization to the center view. In this view, there is a graph visualization of the mapping, which connects the dragged classes and properties

to the input database schema shown on the left (**Fig. 13**). In (Bąk, Blinkiewicz, & Ławrynowicz, 2017), the authors presented a new version of the tool, where a wizard assisting mapping creation was added to the tool. This wizard has 5 steps: (1) creating a project and defining an URI template for the subjects of the triples, (2) defining the ontologies that are going to be used in the mapping, (3) defining the input database schema, (4) mapping creation using a visual interface, and (5) which offers the possibility of querying the mapping using SPARQL.
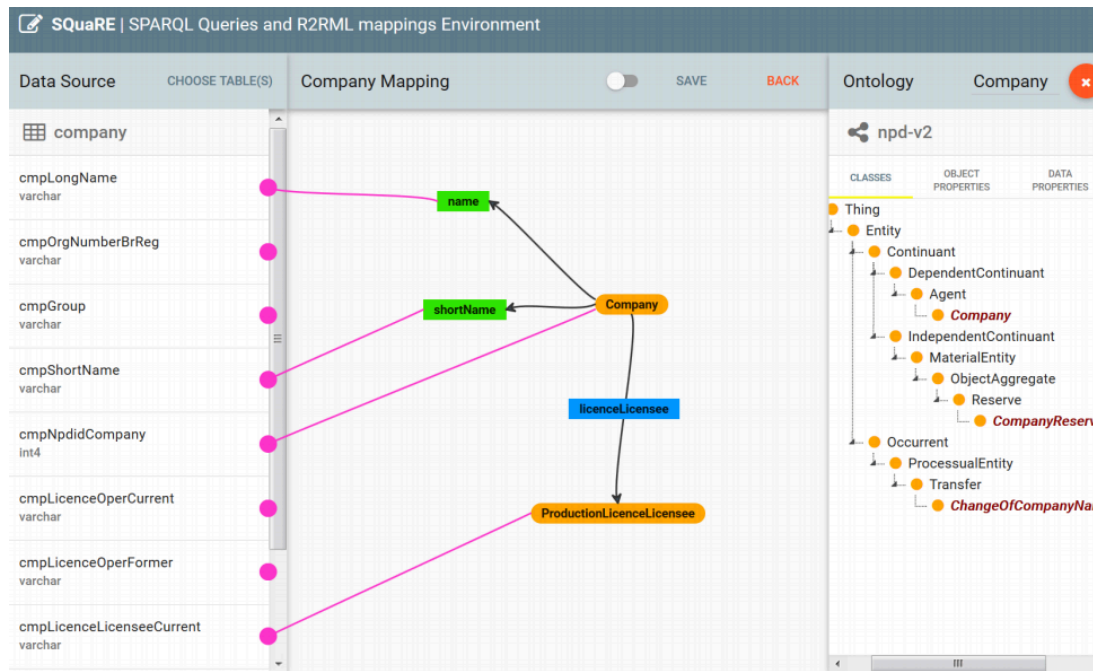
Fig. 13. SQuaRE's mapping interface (Blinkiewicz & Bak, 2016)

The mappings created using SQuaRE are serialized using the R2RML mapping language. The tool offers the possibility for filtering data through the visual interface (which are transformed to SQL queries when the mapping is exported to R2RML). Nonetheless, other data transformation functions are not supported by SQuaRE. A user evaluation with 6 participants was undertaken to evaluate the approach. Out of 6 participants, 5 were considered experts in Semantic Web technologies, and 1 was considered to have general knowledge of Semantic Web technologies. The paper reporting this evaluation only provides a brief description of the experiment, which says that participants used the IMDB Movie Ontology to create 10 mappings. At the end of the experiment, participants answered a qualitative questionnaire on the use of the tool. This questionnaire was created by the authors and was not based on standard usability questionnaires. The authors described users' feedback as positive. Users mentioned that they liked the drag and drop approach and the

colour scheme, while suggesting improvements in the workflow and navigation of the tool. The system was not available at the time of writing.

3.3.9. Map-On

Map-On (Sicilia et al., 2017) is a tool for mapping relational databases to RDF. The mapping process in Map-On starts by loading the input database schema (as an SQL file with DDL statements) and the ontologies (in OWL) that are going to be used in the mapping into the tool. The tool then shows both the input database and target ontology as two separate graphs. The graphs are represented using different colours in order to distinguish between the input database and target ontology elements. Map-On does support the creation and editing of mappings through its graph visual representation, where the definition of relations between elements in the graph visualization are used to define the mapping. Mappings in Map-On are created per input source and ontology. For example, if two ontologies are needed to map one data source, one needs to load the same input data twice, one for each ontology. **Fig. 14** shows the interface for the creation of mappings in Map-On.
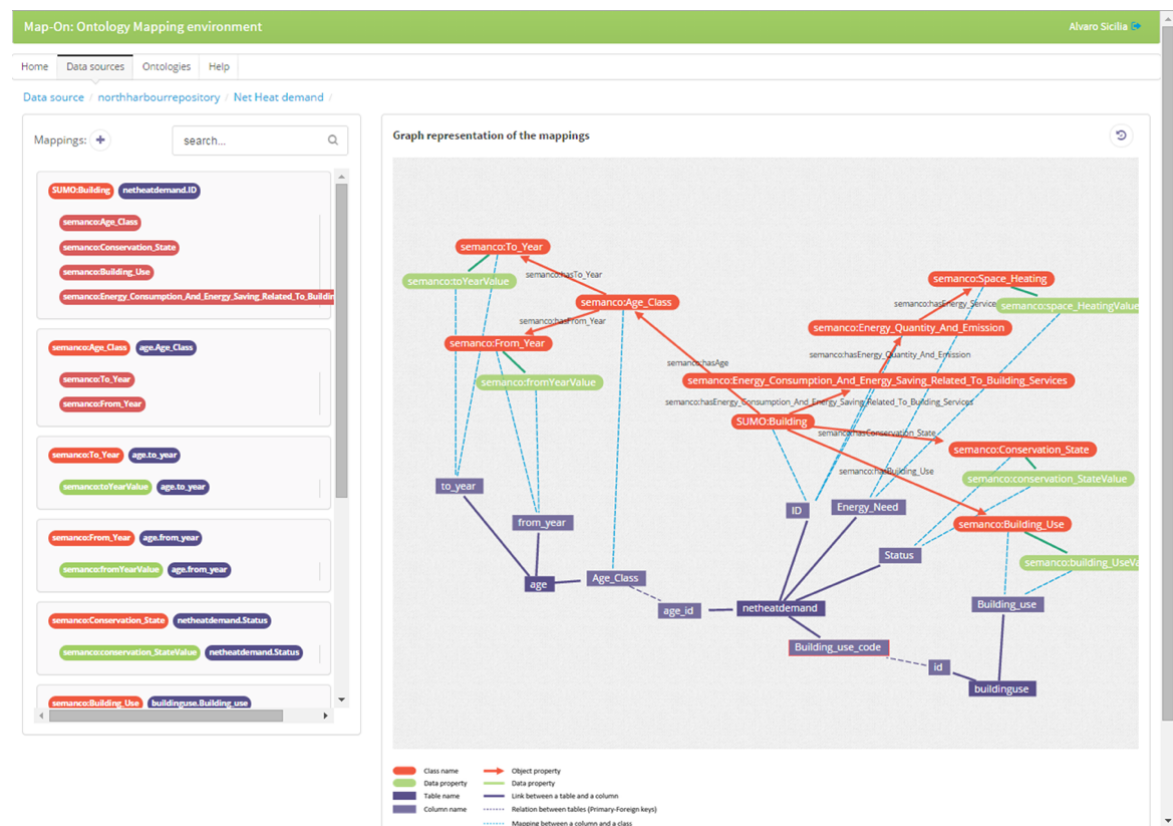
Fig. 14. Map-On's visual representation of mappings (Sicilia et al., 2017)

Map-On allows for the serialization of mappings using the R2RML mapping language, and the tool does not support data transformation functions. A user experiment with 5

participants was conducted to evaluate Map-On. The authors describe the participants as experts in database theory, but not in Semantic Web technologies. The paper does not provide details on the mapping tasks, describing them as three tasks in the domain of research conferences. Map-On was evaluated considering the performance of participants in the creation of mappings, and their perceived usability through the System Usability Scale (SUS) questionnaire. Results showed that the system was considered useful, but the graph representation was found confusing for some users. The system for which this review was based is available[25].

3.3.10. RMLx

RMLx Visual Editor (Aryan, Ekaputra, Kiesling, & Tjoa, 2017) is a web-based tool developed to support the creation of uplift mappings. RMLx is based on the RML mapping language. As stated in Chapter 2, RML extends the R2RML mapping language to support a wide range of input formats, including CSV, JSON, amongst others. RMLx also extends RML's vocabulary by adding support for data transformation functions within the mapping. The RMLx Visual Editor abstracts the RMLx language through the use of a form-based interface. This form-based interface allows users to create and edit mappings. The tool also provides a static graph visual representation of mappings. **Fig. 15** shows RMLx's mapping interface.
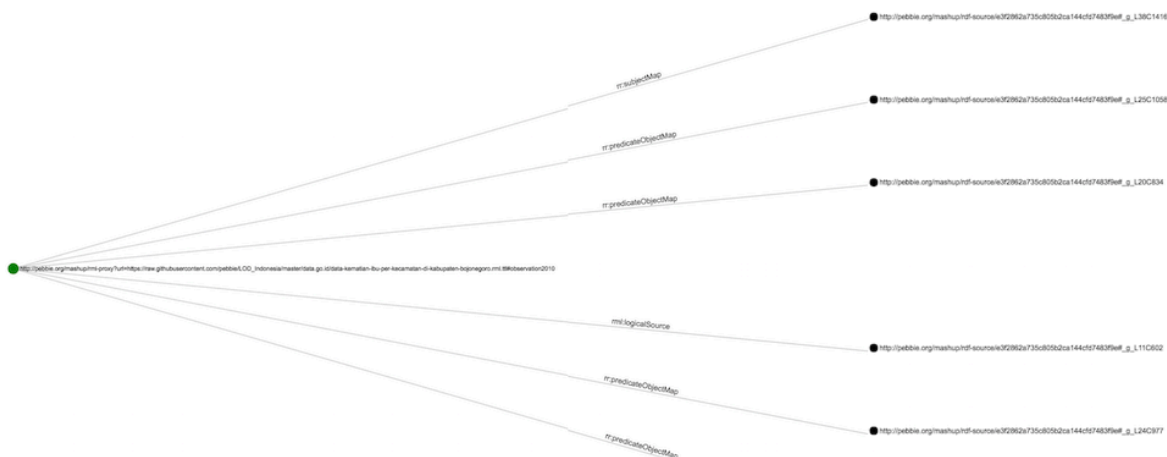
Fig. 15. RMLx's visual representation of mappings[26]

The visual representation does not allow for mapping adjustments, for which users need to define manually though a form-based interface. The data transformation functions are also represented as nodes in the graph. The author does not report on user evaluations. However, in (Heyvaert et al., 2018), a user experiment was conducted to compare RMLx to the RMLEditor, which is discussed in Section 3.3.11. The system for which this review was based is available[27].

3.3.11.RMLEditor

RMLEditor (Heyvaert et al., 2016) is a graph-based tool for the creation of uplift mappings which was built on top of the RML mapping language. RMLEditor's interface is separated in three panels. On the left panel, the input data is shown using a tree structure. The middle panel offers a graph-based visual representation for uplift mappings called MapVOWL. MapVOWL is a visualization for uplift mappings based on VOWL (Lohmann, Negru, Haag, & Ertl, 2016). On the right panel, the results of executing the current version of the mapping are presented. **Fig. 16** presents the interface of the RMLEditor.
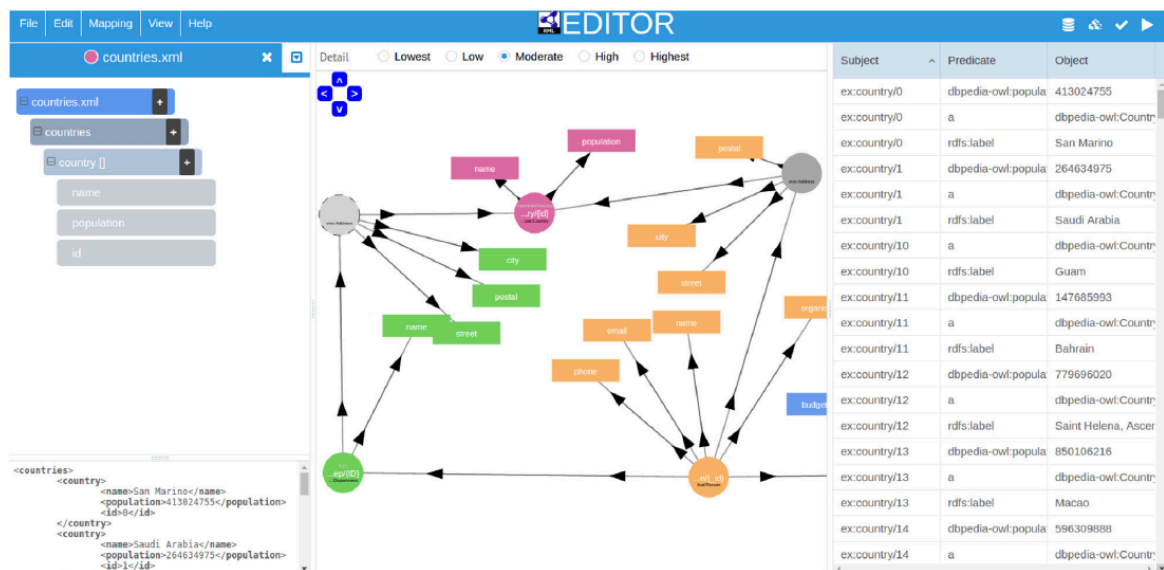
Fig. 16. RMLEditor's mapping interface (Heyvaert et al., 2018)

The RMLEditor is described as independent of the mapping language. However, the editor only supports the RML mapping language. Data transformation functions were added to the editor through the Function Ontology. The Function Ontology (De Meester, Dimou, Verborgh, & Mannens, 2016) allows for the semantically declaration and execution of functions.

---

[27] `http://mashup.pebbie.org/rml`, accessed in August 2018.

User experiments evaluating RMLEditor were reported in (Heyvaert et al., 2016) and (Heyvaert et al., 2018). The experiment presented in (Heyvaert et al., 2016) had 15 participants, of which 10 were considered Semantic Web experts and 5 were considered non-Semantic Web experts. There were two tasks, one involving employees and projects, another involving movies and directors. The RMLEditor was evaluated considering the performance of participants creating the mappings, and the perceived usability of participants using the SUS questionnaire, with generally good results. In (Heyvaert et al., 2018), another 2 user experiments were reported. One related to the interpretability of mapping representations, which had 9 participants, 8 Semantic Web experts and 1 considered to have basic knowledge in the Linked Data domain. In this experiment, participants were asked to answer questions about a mapping that has been presented to them. Mappings were presented either using RML in RDF TURTLE notation, or using the MapVOWL (RMLEditor's visual representation). Example of questions are: "*determine the number of literals per entity"*, and "*determine the number of entities without a class"*, amongst others. Results showed that users preferred MapVOWL, even though users achieved better results when interpreting the RML representation. The second experiment was undertaken to evaluate the creation and editing of mappings. This experiment compared the use of the RMLEditor to RMLx's editor, with 10 participants, all considered Semantic Web experts. The same two tasks used in (Heyvaert et al., 2016) were used in this experiment. This study evaluated the editor in relation to the performance and usability of participants. Results of this experiment showed that RMLEditor had better usability, and better performance, when compared to the RMLx's editor. The system was not available at the time of writing.

3.3.12. Rdf2rdb

Rdf2rdb (Alexiev, 2016) is an approach that combines a graph like visualization and UML class diagrams to represent uplift mappings. Each node in the visual representation renders an RDF resource with its class type and property definitions, while edges represent relations between these nodes. The tool supports the transformation of relational databases to RDF through the R2RML mapping language. Thus, SQL queries must be embedded to the nodes so that R2RML mappings can be generated from the representation. **Fig. 17** shows an example of mapping represented in rdf2rdb.
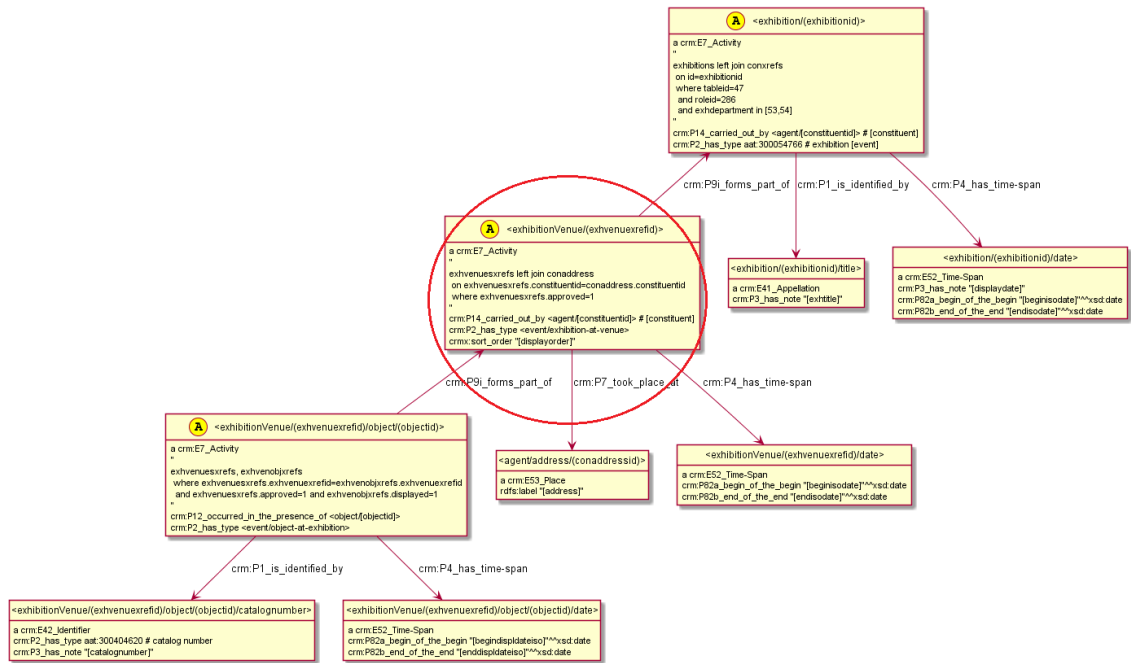
Fig. 17. Rdf2rdb's mapping interface[28]

The visual representation proposed by rdb2rdf is mainly concerned with the readability of mappings, which might aid users with knowledge of UML diagrams. Nonetheless, it is still necessary to understand that the visual representation represents a graph. Furthermore, the creation and editing of mappings is done through text using the PlantUML[29] language. Transformation functions are not supported, and no user evaluations have been reported on the tool or its visual representation. Although the system was available[30], the author of this thesis was unable to successfully get the system to execute it at the time of this review.

In section 3.3, 12 systems from the state of the art that use primarily graph-based visualization to support different tasks have been review. In the next section, systems that use tree-based representations will be reviewed. Section 3.5 will then analyse the state of the art for both visualization representation types.

3.4. Tree-based visual representations

This section reviews 9 approaches from the state of the art that apply primarily tree-based visual representations. Most existing approaches presented in this section show ontologies and/or data using a tree representation, with lines connecting its elements representing

---

mappings. As mentioned before, some of the approaches offer additional visual representations for ontologies, input data, or mappings.

Tree-based visual representations represent concepts using a tree structure, where indentation is used to illustrate the relations between these concepts (Bederson & Shneiderman, 2003). Tree representations are ideal for representing and navigating hierarchies (Ware, 2012). Usually these representations allow for collapsing and expanding elements into sub-groups, which can be used to provide different levels of detail in the tree representation, enabling large structures to be interactively explored (Ware, 2012). In literature, tree visualizations are also called *indented lists* or *indented trees*.

Existing approaches that apply primarily tree-based visual representations are presented in the remaining of this section. In the view of the author, this section presents a comprehensive list of approaches that are relevant to the research of this thesis.

3.4.1. COMA++

COMA++ (Aumueller, Do, Massmann, & Rahm, 2005) is an ontology matching system, hence, the tool supports candidate correspondences from which semantic mappings can be distilled. COMA++ supports different matching strategies, and the reuse of previous results as input to new executions of the matching algorithms. The tool also has a repository for storing mappings created by the tool. This characteristic allows users to compare mappings between common source and target ontologies. This feature also allows for mappings to be merged, or edited. **Fig. 18** shows the visual representation of mappings used by COMA++.
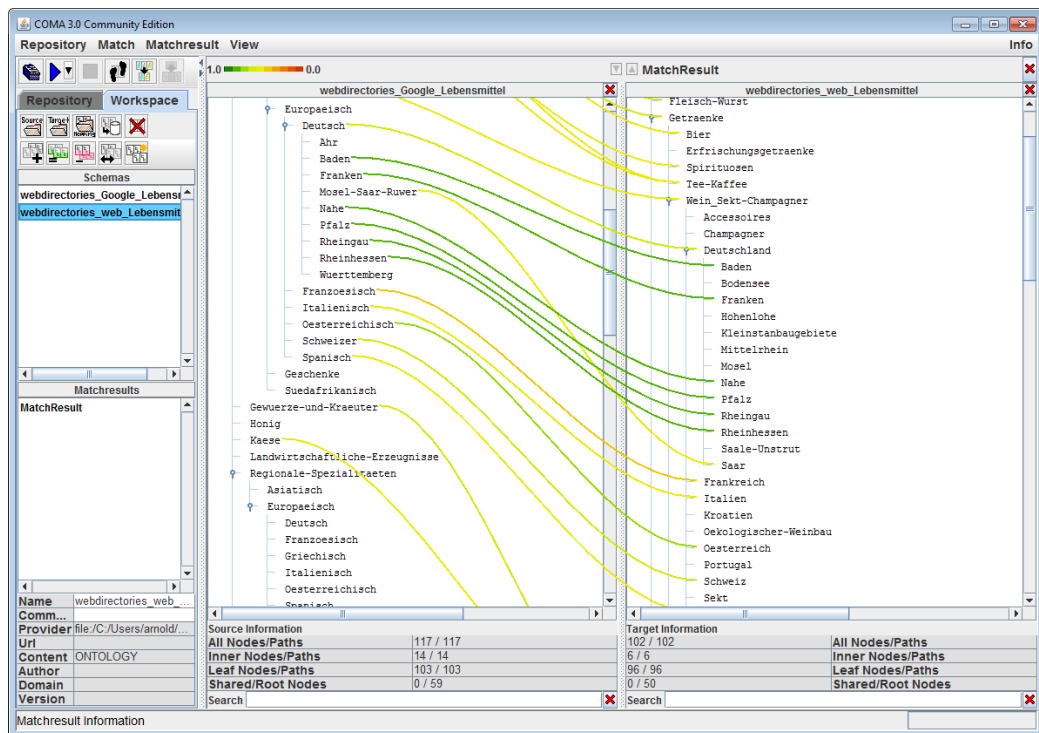
Fig. 18. COMA++'s mapping interface[31]

Source and target ontologies are represented side-by-side as trees. Mappings are represented as lines connecting the elements of the ontologies. Users can interact with the mapping by clicking on the elements of the ontologies or the lines that represent a mapping definition. Modifications of the mapping can be done through the visual representation by selecting elements in the ontologies. COMA++ allows for the serialization of mappings using a proprietary format, which limits the sharing and reuse of mappings with only other COMA++ tools. The tool supports the representation of complex mappings. No user evaluations of the tool or its visual representation of mappings have been reported. The system for which this review was based is available[32].

3.4.2. System for Aligning and Merging Biomedical Ontologies

The System for Aligning and Merging Biomedical Ontologies (SAMBO) (Lambrix & Tan, 2006) is an ontology alignment system. The SAMBO's architecture is domain independent, but the matching algorithms implemented in the system were ones commonly found in the biomedical domain at the time. **Fig. 19** shows the main mapping interface of the SAMBO system.

---

[31] From `https://dbs.uni-leipzig.de/en/Research/coma.html`, accessed in August 2018.
[32] `https://sourceforge.net/projects/coma-ce/`, version 3.0, accessed in August 2018.
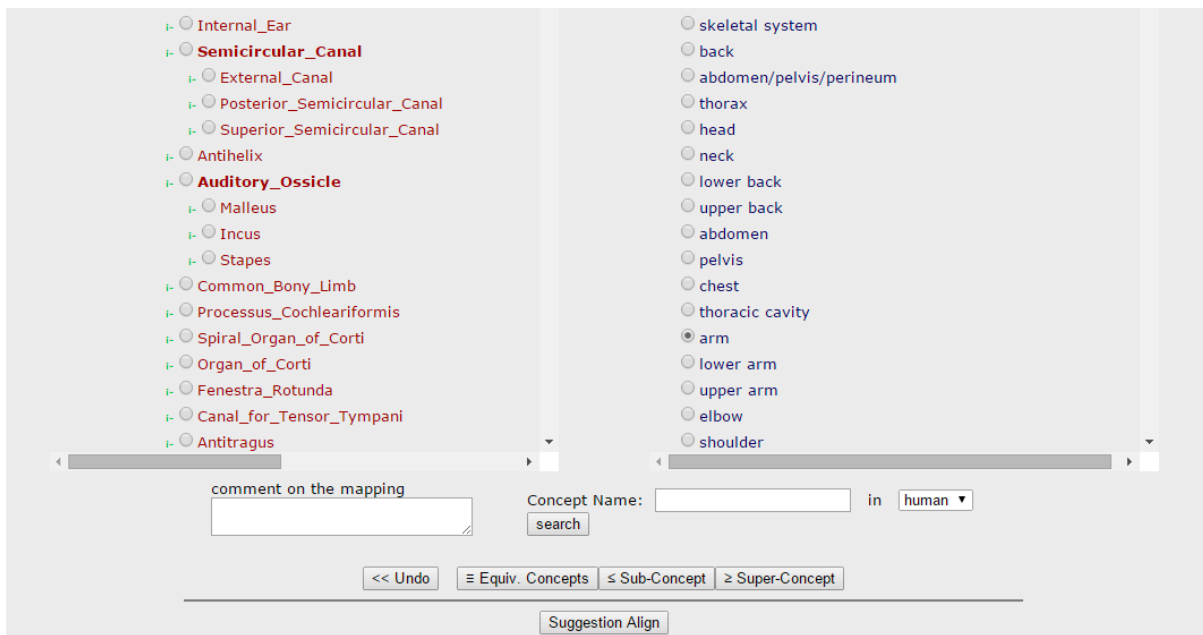
Fig. 19. SAMBO's visual representation of mappings (Lambrix & Tan, 2006)

The system supports simple semantic mappings. The mappings identified automatically by the system are shown using a form-based interface, where users can validate them. The manual creation of mappings in the system presents source and target ontologies as trees, with a search option. Mappings are defined by selecting elements in the trees and clicking on the buttons referring to equivalent concept, sub-concept, or super-concept as shown in **Fig. 19**. Mappings can be serialized using the Alignment format. There was no report of user evaluation. The evaluation presented was concerned with mapping quality and processing time. The system was not available at the time of writing.

3.4.3. Not Yet Another Matcher++

Not Yet Another Matcher (YAM++) (Ngo & Bellahsene, 2012) is an ontology matching system that uses machine learning techniques to discover mappings between ontologies. Machine learning is used to combine different similarity metrics when instance data is available. If no instance data is available, then the system uses metrics based on information retrieval techniques. **Fig. 20** shows the mapping interface in YAM++.
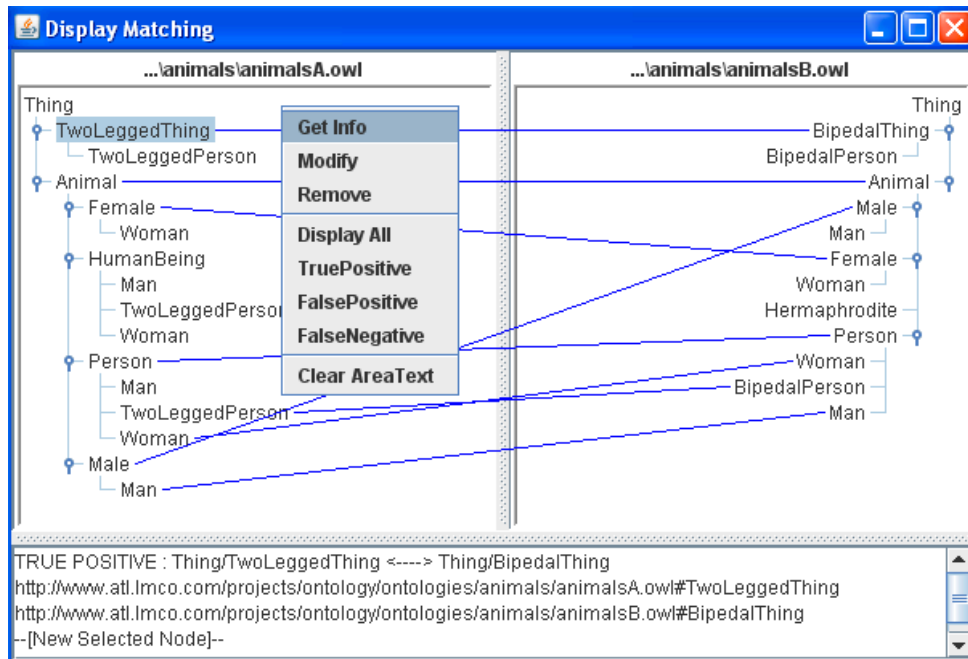
Fig. 20. YAM's visual representation of mappings (Ngo & Bellahsene, 2012)

The visual representation of mappings in YAM++ is similar to the tree representation used in COMA++. Ontologies are shown as trees side-by-side, and lines connecting the elements of the ontology represent mapping definitions. As in COMA++, the visual representation allows for the creation and editing of mappings. In contrast to COMA++, YAM++ allows mappings to be exported using the Alignment Format. No user studies of the tool or its visual representation have been reported. The system was not available at the time of writing.

3.4.4. OntoMap

OntoMap[33] is a plugin for the NeOn Toolkit (Haase, Lewen, Studer, & Tran, 2008) that supports the creation and management of semantic mappings. The NeOn Toolkit platform is described as an ontology engineering environment. **Fig. 21** shows the mapping view in OntoMap.

---

[33] http://neon-toolkit.org/wiki/1.x/Ontology_Mapping.html, accessed in August 2018.
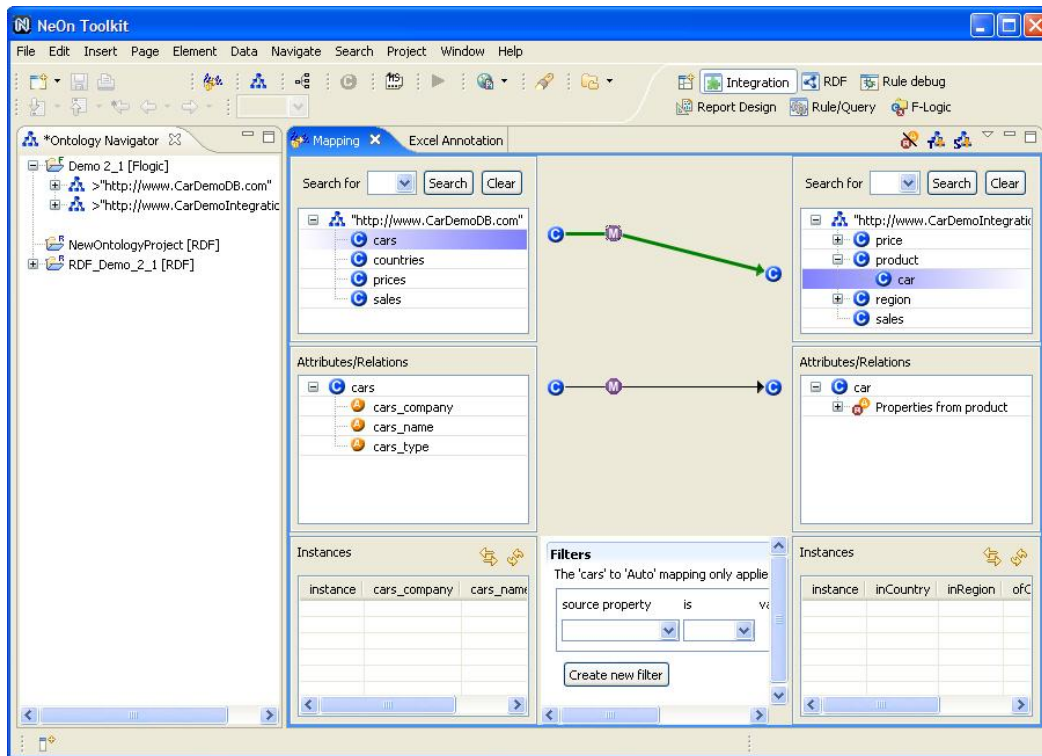
Fig. 21. Neon's OntoMap mapping interface [34]

The tool shows the ontologies using a tree structure. Mappings are created by dragging concepts from the ontology trees into the other panels, which are also represented as trees. There is one panel responsible for showing the classes of the ontologies, another responsible for showing the properties, and another which presents instance data. Mappings are represented with lines that connect the elements from these three panels. The tool only supports simple mappings, and these are serialized using a proprietary format. There is no report of user evaluations of its visual representation of mappings. The system was not available at the time of writing.

3.4.5. Visual Ontology Alignment Environment

The Visual Ontology Alignment Environment (VOAR) (Severo, Trojahn, & Vieira, 2014) is a web-based system that provides primarily a tree-based visual representation of semantic mappings. Unlike other systems presented in this chapter which focus on matching techniques, VOAR is concerned with the visualization, editing, and evaluation of mappings.

As mappings are loaded into the tool, users can select which colour they will be presented in the visualization panel. The main visual representation in VOAR shows source and target ontologies as indented trees, with lines connecting its elements representing mappings. The

---

[34] From `http://neon-toolkit.org/wiki/1.x/Ontology_Mapping.html`, accessed in August 2018.

support for visualizing mappings as graphs and in a tabular format were added in another release of the tool presented in (Severo, Trojahn, & Vieira, 2017). **Fig. 22** shows the tree visualization of ontologies with mappings as lines and the graph visualization.
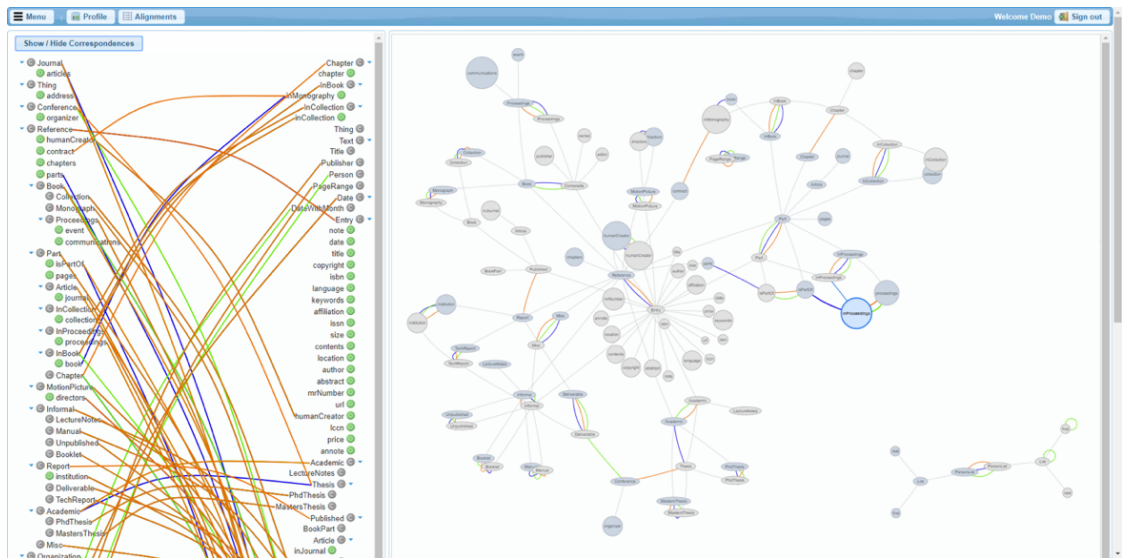
Fig. 22. Visualization of mappings in the VOAR system (Severo et al., 2017)

The creation and editing of mapping rules are supported through a form-based interface. The Alignment Format is used to import and export mappings, which means that complex mappings are not supported. No user evaluations of the visual representation have been reported. The system was not available at the time of writing. However, in (Severo et al., 2017) the authors refer to a demo video[35].

3.4.6. Cognitive Support and Visualization for Human-Guided Mapping Systems

Cognitive Support and Visualization for Human-Guided Mapping Systems (COGZ) (Falconer & Storey, 2007) is a plugin built as an alternative UI for the Prompt (Noy & Musen, 2003) system. Prompt is also a plugin of the Protégé ontology editor. Prompt begins the mapping of ontologies with users selecting source and target ontologies. The tool then applies an ontology matching algorithm which generates the semantic mapping. In a next step, users can edit the existing mappings created by the tool or add new ones. **Fig. 23** shows the mapping interface in COGZ.

---

[35] `https://youtu.be/wq-yPBOFN_I`, accessed in August 2018.
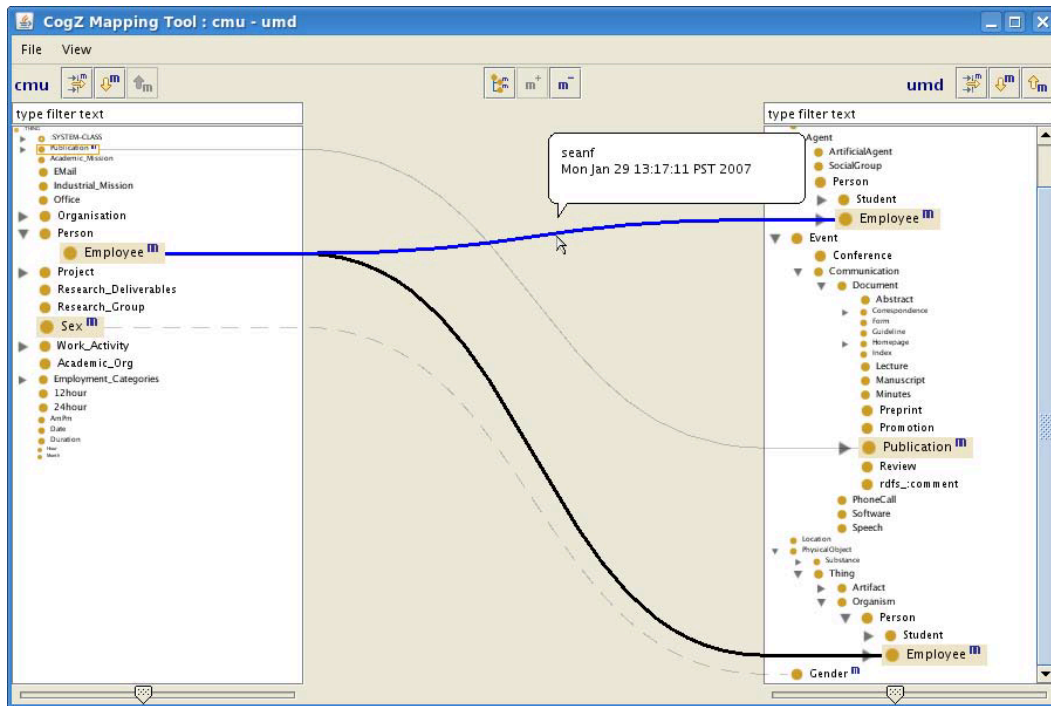
Fig. 23. COGZ's visual representation of mappings (Falconer & Storey, 2007)

The representation of mappings in COGZ is similar to COMA++'s. Source and target ontologies are shows as trees, being their primary visual representation, while mappings between source and target elements are represented using lines. When a mapping is selected, its details are presented in a different panel. COGZ adds interface functionalities to Prompt, such as search and filtering options for ontologies and mappings. The tool also supports the creation and editing of mappings through its visual representation. The tool supports simple mappings, which can be exported using the Alignment Format. No user evaluations have been reported that validate the approach, and the system was not available at the time of writing.

3.4.7. ODEMapster

ODEMapster (Rodriguez & Gómez-Pérez, 2006) supports the transformation of relational databases to RDF. ODEMapster is described as a framework processor that supports the creation of mappings visually. **Fig. 24** shows the mapping interface in ODEMapster.

48
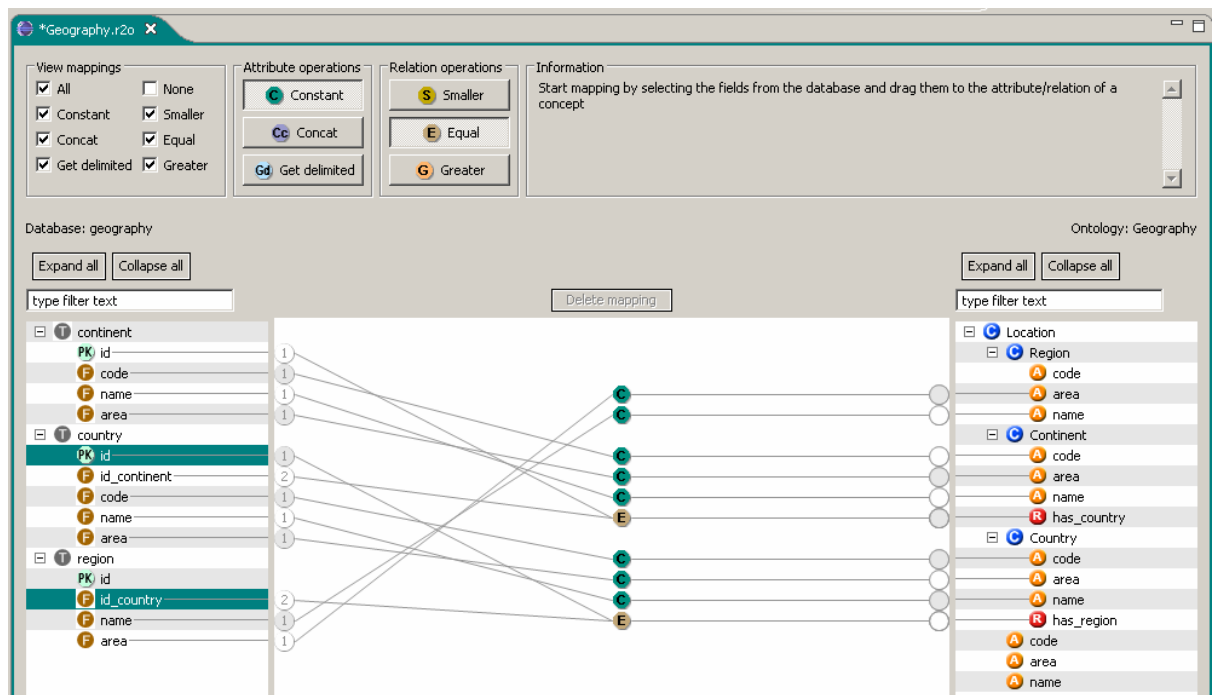
Fig. 24. ODEMapster's visual representation of mappings[36]

The mapping process within ODEMapster is done through a tree-based visual representation. The input database schema and target ontology are represented as trees, and the mappings are created from relations between the tree elements. ODEMapster serializes mappings using the R²O (Relational to Ontology) mapping language. R²O (Barrasa, Corcho, & Gómez-pérez, 2004) is outlined as an extensible declarable XML-based language to express mappings between relational databases and ontologies. ODEMapster supports conditionals and some operations when generating RDF datasets, since these can be expressed using the R²O mapping language. Therefore, the tool partially supports data transformation functions. No user experiments have been reported on the use of the tool. The system was not available at the time of writing.

3.4.8. R2RML By Assertion

R2RML By Assertion (RBA) (Neto, Vidal, Casanova, & Monteiro, 2013) is a tool developed for the creation of uplift mappings using the R2RML mapping language. The mapping creation process is defined in three steps. In the first step, users need to load the input database schema and target ontology into the system. In the next step, RBA shows the input and target ontologies side-by-side using in a tree representation. In this step, users are required to define assertions between the input database and target ontology, as it is shown

---

in **Fig. 25**. Each assertion will be used to create a mapping rule. Finally, in the last step, RBA generates the R2RML mapping based on these assertions.
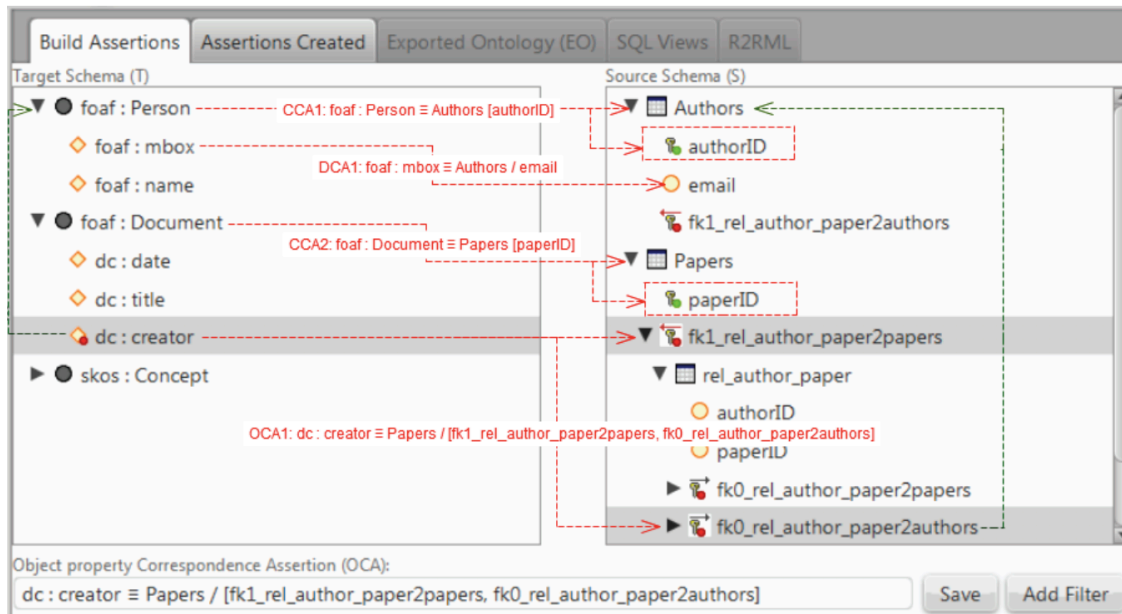
Fig. 25. R2RML by Assertion mapping interface (Neto et al., 2013)

As previously stated, RBA serializes mappings using the R2RML mapping language. The tool does not support data transformation functions, and no user evaluations of the tool have been reported. The system was not available at the time of writing.

3.4.9. MIPMap

MIPMap (Stoilos, Trivela, Vassalos, Venetis, & Xarchakos, 2017) is a data integration tool based on the ++Spicy application (Marnette, Mecca, Papotti, Raunich, & Santoro, 2011). The tool was developed to support data integration in the Medical Informatics Platform (MIP), which is a part of the Human Brain Project (HBP). Originally, the tool only had support for mappings between CSV files and relational databases. At a later stage, support for transforming CSV files and relational databases to RDF was added. Despite being developed for the HBP project, the tool is generic and can be used with any relational database or CSV files. **Fig. 26** shows the interface of MIPMap for the mapping of a CSV file to a relational database.
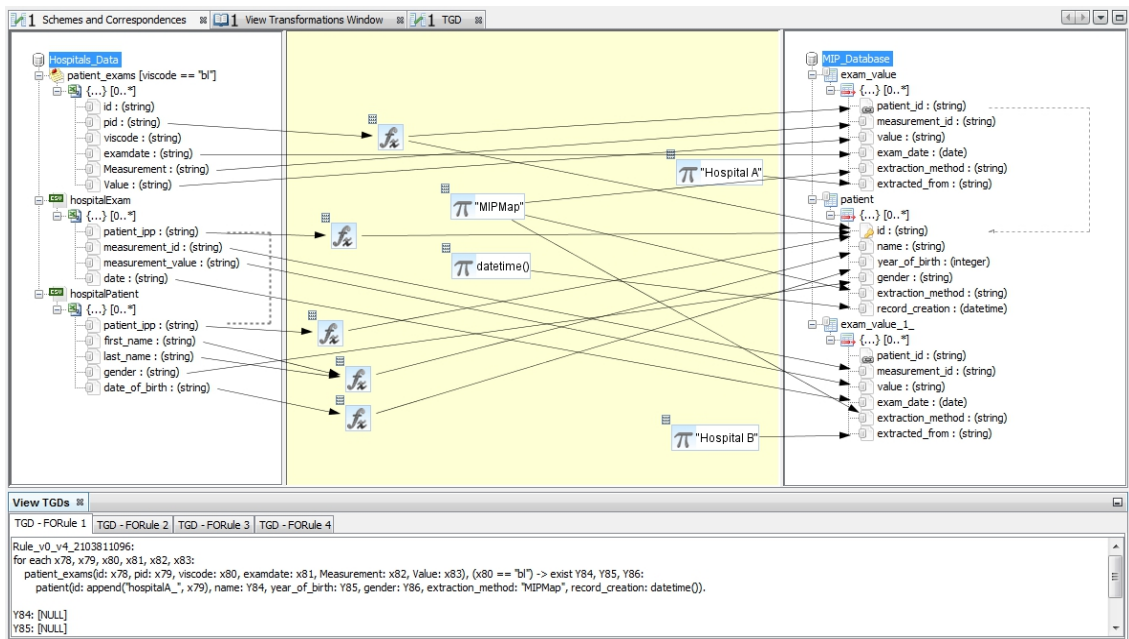
Fig. 26. MIPMap's mapping interface (Stoilos et al., 2017)

The tool represents source data and target ontologies using a tree representation. Source data is presented on the left, and target ontologies are presented on the right. Mappings are defined using the central panel, represented by lines between the elements of the trees. These mapping definitions are used to generate R2RML mappings. The tool supports data transformation functions partially, which are transformed to SQL queries when the mappings are serialized in R2RML. No user evaluations have been reported on the tool or its visual representation of mappings. The system was not available at the time of writing.

In Section 3.4, 9 systems from the state of the art that use tree-based visualizations primarily have been reviewed. In the next section, these are analysed alongside the graph-based visualization systems reviewed in Section 3.3.

3.5.   Analysis of Existing Approaches for Representing Mappings

In this section, an analysis of the approaches in the state of the art that support the visualization of mappings in Linked Data are presented.

Table 1 shows the approaches reviewed in this chapter, their primary visual representation, mapping type and mapping languages supported, editing approach and if the system have been evaluated through user experiments. The publication year (Pub. Year) of the existing approaches, together with the publication year of the last paper discussing the approach (Last. Pub. Year) are also presented – this is to show that even though some approaches are quite old they are still relevant to the field. Approaches marked with (*) offer additional visual techniques. For editing approach, a *check mark* – ✓ – means that it is

possible to create and edit mappings through the visual representation, which is argued in this thesis to be more intuitive, as users do not need to access a separate interface to perform modifications in the mapping; a *check mark in parenthesis* – (✓) – means that users can select elements in the visual representation, but the creation and editing options are available through separate menus or forms; the ✖ sign means that modifications are only allowed through separate form-based interfaces. For user evaluations, approaches with a *check mark* have evaluated their approaches considering performance and usability, a *check mark in parenthesis* means that only one aspect was evaluated (RepOSE evaluated performance and SQuaRE used a qualitative survey to evaluate the approach), an ✖ means that no user evaluated has been reported.

Table 1. Summary of key features in existing visual representations for mappings in Linked Data.

Approach			Pub. Year	Last Pub. Year	Mapping Language	Editing approach	User evaluation
Graph	Semantic	AlViz (*)	2006	2017	Proprietary	(✓)	✖
		AML	2013	2017	AF	✖	✖
		OLA	2004	2018	AF	✖	✖
		RepOSE	2013	2016	N/A	(✓)	(✓)
	Uplift	DataLift	2012	2017	Proprietary	✖	✖
		Karma (*)	2012	2015	(K) R2RML	(✓)	✖
		Lembo et. al.	2014	2017	R2RML	✖	✖
		SQuaRE (*)	2016	2017	R2RML	✓	(✓)
		Map-On	2015	2017	R2RML	✓	✓
		RMLx	2017	2017	RML	✖	✖
		RMLEditor (*)	2016	2018	RML	✓	✓
		Rdf2rdb	2016	2016	R2RML	✖	✖
Tree	Semantic	COMA++	2005	2016	Proprietary	✓	✖
		SAMBO	2006	2018	AF	(✓)	✖
		YAM++ (*)	2012	2017	AF	✓	✖
		OntoMap	2008	2016	Proprietary	✓	✖
		VOAR (*)	2014	2017	AF	✖	✖
		COGZ	2007	2015	AF	✓	✖
	Uplift	ODEMapster	2006	2017	R^2O	(✓)	✖
		RBA	2013	2017	R2RML	(✓)	✖
		MIPMap	2017	2017	R2RML	✓	✖

Table 1 shows that existing approaches have being developed for one type of mapping, which is supported by one specific mapping language. This means that users need to learn completely new interfaces when working with different types of mappings or with a different mapping language – which may better suit a specific use case (Heyvaert et al., 2018). Moreover, as stated in Chapter 2, the mapping processes for producing uplift and semantic mappings are similar, and often have the same stakeholders. The table also indicates that

graph representations are more common for tools with support for uplift mappings, while tree representations have a slightly more even distribution between the approaches' visual technique and the type of mapping they support. It is also evident that the R2RML mapping language and its extensions (RML and KR2RML) are predominant in approaches that support uplift mappings, while the Alignment Format (AF) is the language that has been most used for the representation of semantic mappings. The Alignment Format, however, can only express simple mappings. It should also be noted that user experiments have only been reported for 4 of the 21 reviewed approaches, and only 2 of these evaluated both performance and usability.

The remainder of this section elaborates on the key characteristics presented in **Table 1**.

3.5.1. Visual technique

The existing approaches were classified based on their primary visual technique, as graph-based visualizations or tree-based visualizations. Some of the approaches have additional visual representations to support user involvement, which were also discussed.

As mentioned before, in graph-based representations, concepts are represented as nodes and their relations are represented with edges. Tree-based visual representations have their concepts organized in a tree structure, where indentation is used to illustrate the relations between concepts. The main difference between these visualization techniques is that in tree-based representations there is only one single path between any pair of elements, while graph representations may have multiple (Ware, 2012). In other words, when elements in a tree are accessible from multiple paths, they need to be "duplicated" in different places in the tree. In graph representations this is supported by having multiple edges between a set of nodes. **Fig. 27** shows such case using a fragment of the FOAF[37] ontology, where the class `foaf:Person` is a subclass of `foaf:Agent` and `geo:SpatialThing`.
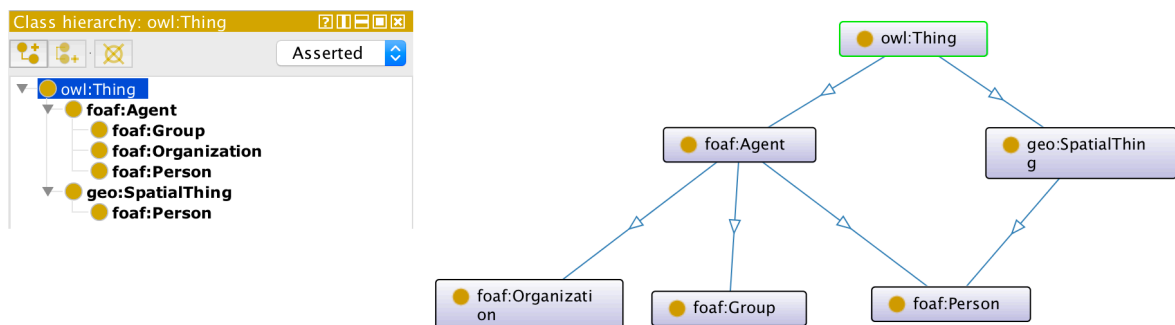
Fig. 27. FOAF ontology fragment as a tree (in Protégé), and as a graph (using the Protégé plugin ontoGraf).

---

[37] `http://xmlns.com/foaf/0.1/`, accessed in August 2018.

Two different studies have compared tree-based and graph-based visual representation for ontologies (Fu, Noy, & Storey, 2013; Katifori, Torou, Halatsis, Lepouras, & Vassilakis, 2006). These studies reported that trees were found to be more organized, familiar (especially for non-experts, as most users have had used similar representations) and predictable, while graph ones were found to have less redundancies, particularly for ontologies with multiple inheritance.

3.5.2. Editing approach

Three types of editing approaches were found in the state-of-the-art with respect to creating or editing mappings: (i) through the visual representation; (ii) without interacting with the visual representation, where modifications are done through separate form-based interface; (iii) or by selecting elements from the visual representation and using mapping options that can be accessed through separate menus.

It is argued in this thesis that creating or editing mappings through separate interfaces or interface components is not as intuitive as one would like. Accessing separate interfaces may increase the complexity of performing such a task – in the number of steps and learning how to use this extra interface – thus affecting performance and usability (Norman, 2013). Moreover, this may also break the flow of action as users shift their attention from the visual representation to this separate component or interface (Shneiderman et al., 2016).

3.5.3. Mapping type

Existing approaches support one mapping type: semantic or uplift. Which means that users need to learn completely new interfaces when working with different types of mappings.

In this characteristic, note that for semantic mappings, it appears that COMA++ is the only tool that supports the representation of complex mappings. The support for data transformation functions for uplift mappings is only fully supported by the RMLEditor and Karma approaches. Nonetheless, only the RMLEditor represents these visually. Other tools offer partial support for data transformation functions, such as DataLift, ODEMapster, MIPMAP, SQuaRE and RMLx.

3.5.4. Mapping language

The existing approaches have support for only one mapping language. If users need to use a different engine to execute their mappings, which might be supported by a different mapping language, it is necessary to learn a new visual representation, or work directly with the new mapping language itself. It is important to note that the approaches AlViz, COMA++, DataLift and OntoMap use their own proprietary mapping language for the serialization of mappings. While RepOSE does not report on how mappings can be serialized.

3.5.5. User evaluation

The existing approaches support user involvement through the representation of mappings visually. However, only four of them have been evaluated by their authors through user experiments. These are the RepOSE system, which supports semantic mappings, and the uplift mapping approaches SQuaRE, Map-On and the RMLEditor. Note that the RMLx tool was not evaluated by their authors, but used in the RMLEditor's evaluation.

RepOSE had two user experiments with domain experts, where the system was found useful. However, it was noted by users that the graph visual representation had too many elements for some use cases. SQuaRE's user experiment evaluated the tool based on a qualitative survey developed by the authors, with generally positive feedback. The Map-On and the RMLEditor approaches were evaluated considering the performance of users interacting with the tool, and their perceived usability. These approaches also had positive feedback, while some participants in Map-On's evaluation described its graph representation as confusing. It is worth noting that these evaluations typically involved a small number of participants. Moreover, participants were mainly experts in Semantic Web technologies, or in the domain that the systems were designed to support (e.g medicine).

3.6. Chapter Summary

This chapter has presented a state-of-the-art review of existing visual representations for mappings in Linked Data. As mentioned before, visual representations support user involvement and alleviate the knowledge required to produce such mappings. The existing approaches were classified based on their primary visual technique as trees or graphs. As stated in this chapter, graph visualizations represent concepts as nodes, and the relations between these concepts, in the Linked Data domain, as labelled directed edges. Tree visualizations represent concepts as elements of a tree structure, and their relations are represented with indentation.

The analysis indicates that existing approaches have been developed based on one type of mapping, and one mapping language. This means that users need to learn completely different interfaces when working with different types of mappings, or when an engine that supports a different mapping language is needed. This review has considered the different editing approaches available in the state-of-the-art as well, where a number of them do not support the creation and editing of mappings through the visual representation. It is argued in this thesis that editing mappings through a different interface may not be as intuitive as one would like. The analysis also indicates that support for complex mappings and transformation functions, which are often necessary to produce the desired output, are often neglected. A final aspect analyzed was related to user evaluations, which are, surprisingly, rarely performed to evaluate visual representation of mappings in Linked Data. Even though user involvement is considered fundamental during the mapping process.

The next chapter presents the requirements derived from the analysis of the state of the art. Our approach, which fulfils the derived requirements, is also presented.

4. The Juma Approach

This chapter describes the **J**igsaw p**u**zzles for representing **ma**ppings (Juma) approach. Juma has been designed taking into account requirements derived from the state of the art for a visual representation of mappings in Linked Data. These requirements are presented in Section 4.1. Section 4.2 presents the Juma approach for representing mappings in Linked Data. Section 4.3 presents three applications that apply the Juma approach. Section 4.4 presents the limitations of the approach. To conclude the chapter, Section 4.5 presents a summary.

4.1. Requirements

Through a combination of the research question and the state of the art analysis the following requirements for a new approach for representing mappings in Linked Data were defined.

R1. Expressive

The visual representation should be expressive enough to represent semantic and uplift mappings. As the state of the art has shown, existing approaches target only one type of mapping. This means that even though the mapping process undertaken to generate and interlink Linked Data is similar, and often has the same stakeholders (as discussed in Chapter 2), users still need to learn completely new interfaces when working with a different type of mapping. Moreover, literature suggests that a uniform representation of similar constructs and consistent interaction style within and across applications improve learnability and recognizability (Preece et al., 2001).

Even though many approaches from the state-of-the-art analysis do not support complex mappings, being for uplift or semantic mappings, these are often necessary during the mapping process (Euzenat & Shvaiko, 2007). Thus, this requirement is also concerned with a visual representation that is expressive enough to support both simple and complex mappings.

R2. Intuitive editing

To facilitate the creation and editing of mappings, these should be carried out through the visual representation. It was found in the state-of-the-art review that many approaches use separate interfaces to create and edit mappings. As stated in Chapter 3, accessing separate interfaces may break the flow as the user's attention is shifted away from the visual representation, thus affecting performance and usability (Shneiderman et al., 2016).

R3. Syntactic validation and guidance

The visual representation should guide users during the creation and editing of mappings, only allowing the representation of valid mappings. In order to guide users and bootstrap the mapping process, the visual representation should also provide common mapping constructs.

All existing approaches reviewed in this chapter have support for this requirement. Nonetheless, as stated in Chapter 3, validation and guidance are important aspects to be considered when developing interactive visual representations.

R4. Mapping language independent

Users should be able to create and edit mappings without being preoccupied with a particular mapping language, i.e. the visual representation should be independent of the underlying mapping language. As mentioned before, if another mapping language is needed, then users need to learn a new editor (which generates these mappings for this specific mapping language) or the mapping language itself. Such cases might happen since different mapping languages might better suit a particular use case (Heyvaert et al., 2018). The visual representation should also be independent of the engine that executes the mapping, improving its interoperability and reusability.

Table 2 shows the approaches reviewed in the state-of-the-art and their support to the requirements presented.

Existing state of the art approaches partially support **R1**, as only one type of mapping – uplift (U) or semantic (S) – can be expressed. Some existing approaches fully support **R2**. Full support, in this case, means that the creation and editing is supported through the visual representation; partial support means that users can select elements in the visual representation, but mapping modifications are done through separate menu options; no support means that modifications are supported through form-based interfaces that are disconnected from the visual representation, which is argued not to be intuitive. Syntactic validation and guidance (**R3**) is supported by all existing approaches. RMLEditor is the only approach described as mapping language independent (Heyvaert et al., 2016). Thus, it is the only existing approach with support for **R4**. Nonetheless, it is noted that the RMLEditor as yet only allows the serialization of mappings through the RML mapping language.

Table 2. Requirements support provided by existing approaches.

Approach	R1	R2	R3	R4
AlViz	S	(✓)	✓	✘
AML	S	✘	✓	✘

OLA	S	✖	✓	✖
RepOSE	S	(✓)	✓	✖
DataLift	U	✖	✓	✖
Karma	U	(✓)	✓	✖
Lembo et. al.	U	✖	✓	✖
SQuaRE	U	✓	✓	✖
Map-On	U	✓	✓	✖
RMLx	U	✖	✓	✖
RMLEditor	U	✓	✓	✓
Rdb2rdf	U	✖	✓	✖
COMA++	S	✓	✓	✖
SAMBO	S	(✓)	✓	✖
YAM++	S	✓	✓	✖
OntoMap	S	✓	✓	✖
VOAR	S	✖	✓	✖
COGZ	S	✓	✓	✖
ODEMapster	U	(✓)	✓	✖
RBA	U	(✓)	✓	✖
MIPMap	U	✓	✓	✖

✓ = full support   (✓) = partial support   ✖ = no support

4.2. Jigsaw Puzzles for Representing Mappings

This section presents the **J**igsaw p**u**zzles for representing **ma**ppings (Juma) approach for representing mappings in Linked Data, which was developed to fulfil the requirements presented in Section 4.1.

As stated in Chapter 3, metaphors can be used to help users understand or interpret a certain concept by using another, possibly more familiar one (Ware, 2012). In other words, a metaphor exploits knowledge that users already have in order to facilitate the understanding of a possibly new or more complex domain. Metaphors can be found in many different applications, such as Skype's interface for dialling numbers (patterned after physical phones), icons, such as an exclamation mark (representing a warning), amongst others. Given that Linked Data is being generated and consumed by different types of users at significantly high rates, finding a proper metaphor that can facilitate the mapping process is of significant importance. Among several metaphors being used in visual representations, the block – or jigsaw – metaphor was chosen as it takes advantage of the user's familiarity to jigsaw puzzles, fosters users to explore the combinations of blocks, and for being accessible to experts and non-experts alike. Moreover, as stated in Chapter 1, this metaphor has been successfully applied in other fields, such as querying Linked Data (Ceriani & Bottoni, 2017), programming robots (García-Zubía et al., 2018), and in the data science domain (Bart et al., 2017).

Juma leverages the use of the block metaphor in order to facilitate the interpretation of mappings in Linked Data, which is the focus of the research in this thesis. Juma expresses a set of constructs as blocks, or pieces of a jigsaw puzzle. These blocks are represented using different colours, shapes and sizes. Just like in a jigsaw puzzle, each block may be connected to others, in order to form a mapping. The mapping is then represented as a tree, from the main mapping block (root), others may be connected in order to represent a complete mapping. Recall from the state-of-the-art two studies (Fu et al., 2013; Katifori et al., 2006) that compared tree-based and graph-based visual representations for ontologies, where it was suggested that trees are more intuitive and familiar for different types of users, especially non-expert ones.

Juma fulfils the requirements presented in this chapter. Requirement **R1** states that the visual representation should be expressive enough to represent (simple and complex) semantic and uplift mappings. Juma allows the definition of abstract constructs, which are represented as blocks. Each block is designed to capture a distinct mapping construct, which can be connected to others in order to represent a mapping. The support for complex mappings is also achieved by such abstract blocks. Another characteristic that will benefit users is that similar constructs may have a uniform representation in potentially different applications. It was postulated that these may aid in users in creating different types of mappings without having to learn completely new interfaces.

Requirement **R2** stated that the visual representation should support the creation and editing of mappings through its visual representation. The Juma approach fulfils through the use of the block metaphor, which allow users to create and edit mappings through mapping constructs that are represented as jigsaw puzzle pieces or blocks.

Requirement **R3** states that the visual representation should prevent syntactic errors and guide users in the creation and editing of mappings. Juma fulfils this requirement by only allowing the connection of blocks that produce syntactically valid mappings. During the mapping process, visual cues are used to show users how elements of the visual representation can be used in order to yield valid mappings. These characteristics guide users in the creation and editing of mappings. Juma applications can also have example mappings, that can also be used to guide users and bootstrap the mapping process.

The Juma approach represents mappings independently of how mappings are serialized, which can be done for different mapping languages, fulfilling requirement **R4**. As mentioned before, each block abstracts and captures a different mapping construct, and the connection of the different blocks represent a mapping. Juma then translates each block to an equivalent mapping representation, which can be done for distinct mapping languages.

4.2.1. Juma Features

The main features of the Juma approach are presented below:

- **Elements as blocks**: Each element in the visual representation is represented as a jigsaw puzzle piece like block. Blocks have different visual properties, such as shape, colour and labels. These characteristics are used to differentiate between the elements of the visual representation. Blocks may also have user inputs that can be free text, dropdown menus, check boxes, or open text areas outside the visual representation for the input of longer texts.

- **Block composition**. Block composition allows for users to connect blocks to each other through jigsaw-like connectors. The visual representation defines how the available blocks may be connected to each other, only allowing the definition of constructs that yields a syntactically valid mapping. Possible connections are hinted to users by the block's connectors, and by visually dragging blocks to each other, which shows a visual cue when this a valid connection in the representation. This characteristic also reduces the chances of users making syntactic errors, especially when compared to creating mappings by hand using a text editor.

- **Workspace**. The workspace presents the environment where blocks can be dragged to in order to create and edit mappings. In this context, the workspace represents the current version of the mapping. Blocks in this environment can be disabled and enabled. Disabled blocks are still visible in the workspace but are not used in the serialization of the mapping. Blocks in the workspace can also be collapsed or expanded in order to offer different levels of detail of the current mapping. The workspace also supports scrolling and zooming options, adding comments to specific blocks (that can be used to explain decisions made during the mapping process), and the deletion of blocks. The workspace may start empty or with an initial configuration. This initial configuration can be used by users as a starting point for the creation of mappings.

- **Inventory of blocks**. A menu presenting the blocks organized in different categories is presented on the left of the workspace. In order to create or edit mappings, users can drag and drop blocks from the inventory into the workspace. The inventory may also have default block constructs, created through block composition, that can be used to bootstrap mapping creation and editing. For example, for uplift mappings, a common construct that can be represented in the inventory is a mapping from a table

– of a relational database, for example – to a subject with one class and one predicate object pair definition.

- **Generation of mappings**. The visual representation of mappings is independent of how mappings are generated using mapping languages. As mentioned before, the current mapping is presented in the workspace, which is then automatically translated to mappings in different mapping languages. This characteristic also allows for the execution of mappings to happen independently of mapping creation and editing. Therefore, users can select any available engine, that supports the mapping language chosen, to execute their mappings.

4.3. Juma Applications

This section presents three applications of the Juma approach. A webpage presenting the approach and its applications is available[38].

The first application, **Juma R2RML** (Section 4.3.1), was developed to reflect the R2RML's vocabulary. The second application, **Juma Uplift** (Section 4.3.2), was subsequently developed to support multiple uplift mappings languages, in this case R2RML and SML. This involved identifying and abstracting uplift mapping constructs that would be used to generate mappings in distinct mapping languages. **Juma Interlink** (4.3.3) was developed to support semantic mappings that generate executable mappings in the form of SPARQL CONSTRUCT queries. Juma Uplift and Juma Interlink were developed following general principles found between semantic and uplift mappings, that were used to define a uniform representation for equivalent constructs between these types of mappings (Section 4.3.4). As stated in Chapter 3, consistency in representation and interaction styles within and across applications aids readability and promotes recognizability (Preece et al., 2001).

The Juma applications that are presented in this section were developed using the Blockly (Fraser, 2018) library. Blockly[39] is a JavaScript library maintained by Google. The Blockly library works on the client side, and is compatible with all major browsers, including Chrome, Firefox, Safari, Opera and Internet Explorer. Blockly is also customizable and extensible, allowing for the creation of new block environments. Moreover, several applications have been developed on top of this library, such as the MIT App Inventor 2 (Wolber, Abelson, Spertus, & Looney, 2014), SparqlBlocks (Ceriani & Bottoni, 2017), amongst others.

---

[38] `https://www.scss.tcd.ie/~crottija/juma/,` accessed in August 2018.
[39] `https://github.com/google/blockly,` accessed in August 2018.

4.3.1. Juma R2RML

This section presents the first application of the Juma approach that was developed by the author of this thesis, called Juma R2RML. This application supports uplift mappings by applying the Juma approach to the R2RML mapping language. As stated in Section 2.3.3, R2RML is a W3C Recommendation and provides a vocabulary and an algorithm to transform relational databases to RDF. Juma R2RML abstracts the R2RML vocabulary, while being capable of generating mappings that are fully compliant with its specification. A tutorial video of Juma R2RML is available[40].

In Juma R2RML, each block has been designed to represent an R2RML statement that automatically generates a correspondent R2RML construct. The categories in the inventory of blocks in Juma R2RML is structured as a tree, which shows the hierarchy of the R2RML vocabulary. There are also two extra categories: templates and prefixes. The templates option shows a complete triples map, and a complete predicate object map, with their default values (as stated in the R2RML specification). There are two blocks in the prefixes category, one where users can define any vocabulary, and one with common predefined vocabularies (such as FOAF, RDFS and others). A diagram showing an overview of the Juma R2RML application is presented in **Fig. 28**.

---

[40] `https://www.youtube.com/watch?v=5X2ZHWjDOQ8`, accessed in August 2018.
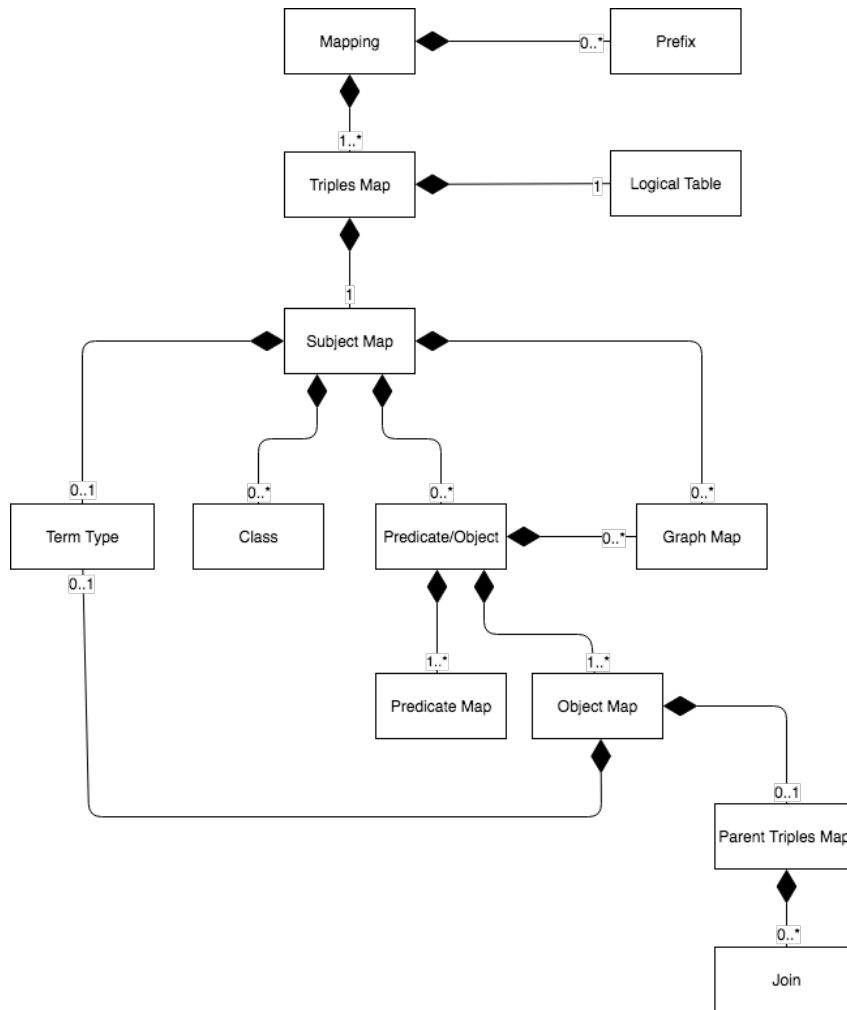
Fig. 28. Overview of the Juma R2RML application.

In relation to term types, objects maps may be literals, blank nodes or IRIs. Subject maps may be IRIs or blank nodes. The R2RML specification defines that predicate and graph maps may have term types, but these must be IRIs. For this reason, Juma R2RML does not provide blocks to define term types for predicate and graph maps. Instead, Juma R2RML sets the term type as IRI automatically, in the generated R2RML mapping, when these constructs are used in the workspace. Juma R2RML allows for objects to have either a language tag or a datatype. The R2RML specification defines that if an object map has one of these constructs, then its type is a literal. In Juma R2RML, users must use a term type block to define the object map to be a literal. This block has a connector that allows for the optional definition of a language tag or a datatype (**Fig. 29**).

Fig. 29. Language tag and datatype definition in Juma R2RML.

Fig. 30 shows Juma R2RML's interface with an example mapping. The inventory of blocks is presented on the left, while the workspace is presented on the right. This mapping defines a prefix, using a block with predefined prefixes, with the FOAF vocabulary, and a logical table referring to the input source *students*. A triples map defines subjects to have the IRI `http://example.org/student/{id}`. Subjects are also declared to be instances of the class `foaf:Person`. A predicate object map relates the subjects with the predicate `foaf:name` to values from the column *name* of the logical table.
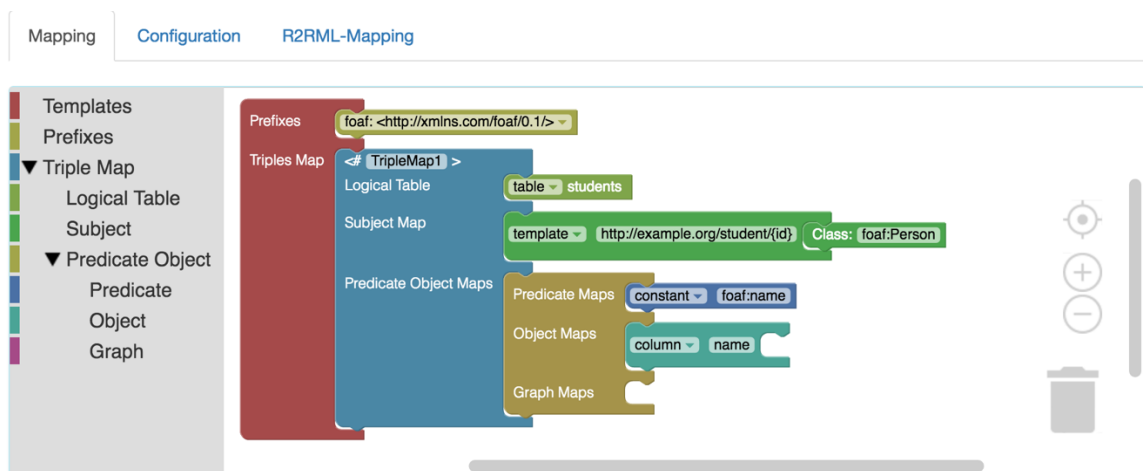
Fig. 30. Juma R2RML's mapping interface.

Juma R2RML's mapping interface is separated in three tabs. In the first tab, called *Mapping*, the inventory of blocks and the workspace where these can be created and edited are shown. In *Configuration*, one can define a title and a description for the current mapping, and the properties of the configuration file. The configuration file is used as input to an R2RML processor together with the R2RML mapping file. This file contains information about how to connect to the input database and options for how the RDF output should be serialized. In the *R2RML-Mapping* tab, the R2RML mapping generated by Juma R2RML is presented.

The R2RML mapping generated by Juma R2RML for the mapping presented in **Fig. 30** is presented in Listing 7.

```
# Mapping created using Juma R2RML.
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<#TriplesMap1>
  rr:logicalTable [ rr:tableName "students"; ];

  rr:subjectMap [
    rr:template "http://example.org/students/{id}";
    rr:class foaf:Person;
  ];

  rr:predicateObjectMap [
    rr:predicateMap [
      rr:constant foaf:name;
      rr:termType rr:IRI;
    ];

    rr:objectMap [
      rr:column "name";
    ];
  ];.
```

Listing 7. R2RML mapping generated by Juma R2RML.

4.3.2. Juma Uplift

The second application of the Juma approach developed by the author of this thesis is called Juma Uplift. Like Juma R2RML, this application also supports the creation and editing of uplift mappings. Juma Uplift has been developed with a higher level of abstraction in order to have the capability to generate mappings using distinct mapping languages, as per user choice. The mapping languages R2RML and SML are supported by the Juma Uplift application. A tutorial video of Juma Uplift is also available[41]. **Fig. 31** presents a diagram as an overview of the Juma Uplift application.
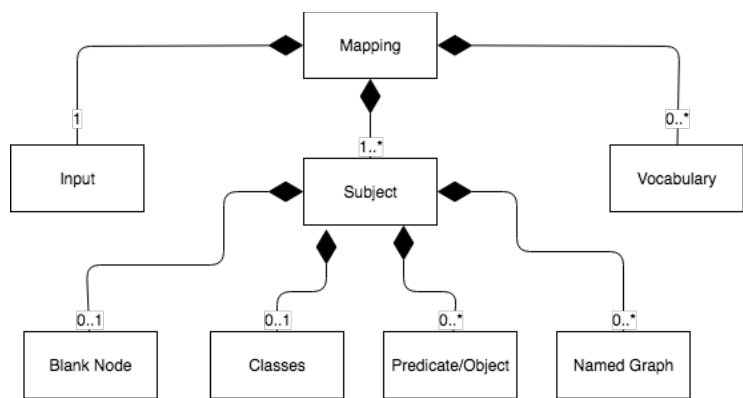
Fig. 31. Overview of the Juma Uplift application.

---

[41] https://www.youtube.com/watch?v=Q97YeZtu_tA, accessed in August 2018.

The inventory of blocks organised as categories in Juma Uplift is presented below:

- **Mapping.** This category provides the main block which relates an input source (this being a table, a view or a SQL query) to zero or more vocabularies and one or more subject definitions. In Juma Uplift, one can relate multiple subject definitions to one input source.

- **Vocabularies.** The vocabularies that are going to be used can be found under this option. There are two types of vocabulary blocks. One with common predefined vocabularies and another customizable one.

- **Subjects.** The category subjects provides blocks for generating the subject, declaring it as a blank node, and as instances of classes. Subject blocks have an *id* (which is used when linking subjects) and are associated with the mapping block. As mentioned before, a mapping block may have many subject definitions. In contrast, Juma R2RML, which follows the R2RML specification, mappings are created per logical table and subject.

- **Predicate/Object.** The blocks to define predicates and objects are defined under this category. Objects can be defined as IRI's, blank nodes, literals, to have a datatype or a language tag. In contrast with Juma R2RML, this block has a checkbox that shows users these optional values (**Fig. 33**). These blocks are associated with a subject definition block. Each predicate/object block defines a new triple for the associated subject. In Juma R2RML one can define many predicate maps and object maps in a predicate object map. In experiment 1, users reported that this was not intuitive, as it is necessary to understand how the R2RML algorithm processes these constructs. This is the rationale for allowing for the definition of these constructs in pairs in Juma Uplift.

- **Linking.** This category provides one with a predicate object pair block where the object is a link to another subject definition, which is identified by its *id*. The linking block also allows users to define how the subjects being linked are related. This relation translates to SQL joins in the output mapping. This block is strongly related to parent triples maps in Juma R2RML. It was found during experiment 1 (Section 5.3.3) that users find the naming convention used in R2RML (parent and child columns) difficult. For this reason, Juma Uplift use the labels *from this table* and *from selected table,* in order to be more intuitive.

- **Functions.** Under this category, data transformation functions have been built-in. A few examples of functions available are concatenation, replace and summation.

Others can be added in future implementations. Section 4.3.2.1 describes how functions are represented in Juma Uplift.

- **Graph.** This category defines an optional block to generate the RDF triples in a specific named graph. Graph blocks are associated to subjects, defining in which named graph the triples associated to the subject will be generated. In Juma R2RML, graph maps may be defined in subject or predicate object maps. Juma Uplift only allows for the definition of named graphs to subjects, which the author of this thesis also believes to be more intuitive.

In relation to assigning triples to named graphs, in R2RML, this can be done in two ways. One where all related triples are generated in specific named graphs (when graph maps are defined in subject maps) or to specific triples (when graph maps are defined in predicate object maps). As mentioned before, Juma Uplift only allows for the definition of named graphs to subjects. Nonetheless, it is still possible to express these mappings in Juma Uplift. For example, Listing 8 shows an R2RML mapping with one triples map where there is one graph map in a subject map and another in a predicate object map. **Fig. 32** shows a mapping in Juma Uplift that would generate the same output when executed. Note that this mapping has two subject definitions related to the same input source.

```
# Mapping created using Juma R2RML.
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<#TriplesMap1>
  rr:logicalTable [ rr:tableName "person"; ];

  rr:subjectMap [
    rr:template "http://example.org/{id}";
    rr:class foaf:Person;
    rr:graphMap [ rr:constant ex:graph  ];
  ];

  rr:predicateObjectMap [
    rr:predicateMap [ rr:constant foaf:name ] ;
    rr:objectMap  [ rr:column "name"; ];
    rr:graphMap [ rr:constant ex:person ];
  ];.
```

Listing 8. Assigning triples to named graphs in R2RML.
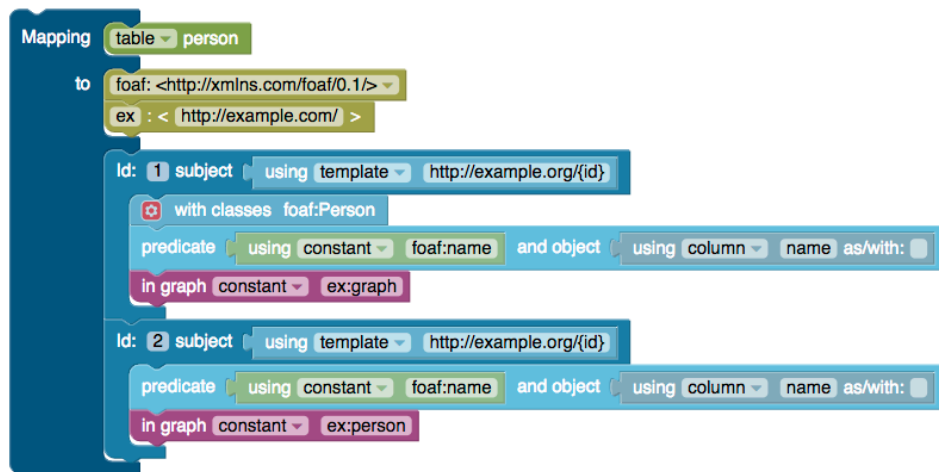
Fig. 32. Assigning triples to named graphs in Juma Uplift.

To be compatible with the R2RML mapping language, it was decided to design the blocks for generating subjects, predicates, objects and graphs using the R2RML constructs types: templates, constants and columns. The mapping presented in **Fig. 30** (in Juma R2RML) is presented in **Fig. 33** using the Juma Uplift representation.
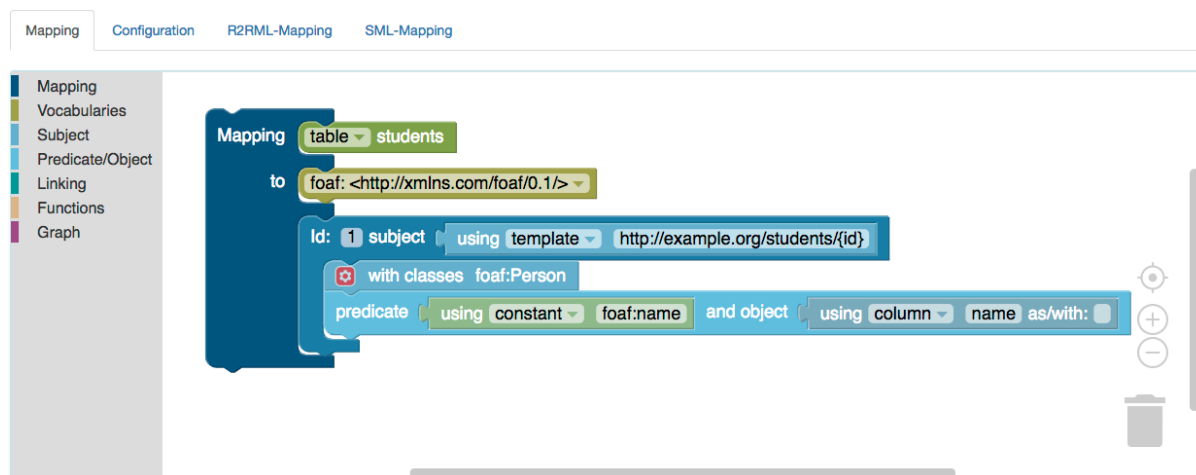
Fig. 33. Juma Uplift's mapping interface.

Juma Uplift has the same three tabs presented in Juma R2RML, plus the tab *SML-Mapping,* which presents the mapping using the SML mapping language. The SML mapping generated by Juma Uplift from the mapping in **Fig. 33** is presented in Listing 9.

```
// Mapping created using Juma editor.
Prefix foaf: <http://xmlns.com/foaf/0.1/>

Create View view1 As
 Construct {
    ?s1  a foaf:Person.
    ?s1    foaf:name    ?o1. }
 With
```

69

```
  ?s1 = uri(concat('http://example.org/students/', ?id))
  ?o1 = plainLiteral(?name)
From
  students
```

Listing 9. SML mapping generated by Juma Uplift.

4.3.2.1. Functions in Juma Uplift

As mentioned before, some built-in functions are available in Juma Uplift, while being extensible to others being added. **Fig. 34** shows a mapping with an object being generated from a concatenation function. The parameters of this function are the columns *last_name* and *first_name* separated by a comma. The definition of parameters is done by clicking on the red toggle, and by dragging elements from the left panel to the parameters connector on the right (**Fig. 35**). It is also possible to define a term type for function definitions.
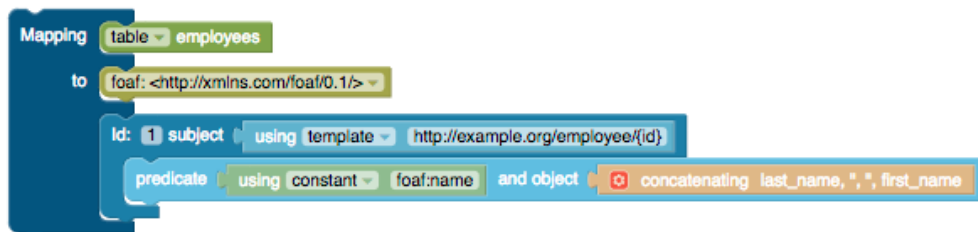
Fig. 34. Example of a Juma Uplift function representation.
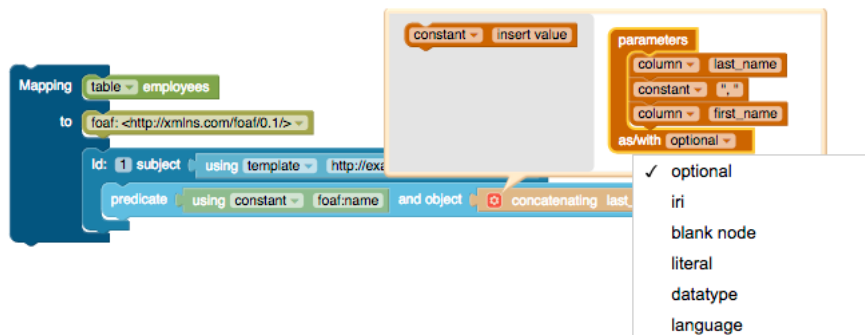
Fig. 35. Example of the definition of parameters in a Juma Uplift function.

The execution of functions is made available through an R2RML-F processor (Debruyne & O'Sullivan, 2016). R2RML-F extends R2RML to add support for the definition of functions. In R2RML-F, *functions* have a *name* and a *body*. Each function declaration must have one function name and one function body. Function names are unique. Function bodies define a function using a standardized programming language. A function body has a signature and a set of parameters. The definition of parameters is optional. Every function defined in a function body must have a return statement. R2RML-F's vocabulary also

70

includes notions for calling and passing parameters to functions. A *function call* refers to a *function*. *Parameters* are optional and can be passed as references to values from the input data or as fixed values. A function call generates an RDF Term based on the function return statement defined in the function body.

The R2RML-F engine integrated to Juma Uplift supports functions in JavaScript. Listing 10 shows the mapping that is generated from **Fig. 34**, which includes a function represented using R2RML-F's vocabulary.

```
# Mapping created using Juma Uplift.
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix rrf: <http://kdeg.scss.tcd.ie/ns/rrf#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<#TriplesMap1>
  rr:logicalTable [ rr:tableName "employees"; ];
  rr:subjectMap [ rr:template "http://example.org/employee/{id}"; ];

  rr:predicateObjectMap [
    rr:predicateMap [ rr:constant foaf:name; rr:termType rr:IRI; ];
    rr:objectMap [
      rrf:functionCall [
        rrf:function <#Concat3> ;
        rrf:parameterBindings (
          [ rr:column "last_name"; ]
          [ rr:constant ", "; ]
          [ rr:column "first_name"; ]
        );
      ]; ]; ]; .

<#Concat3>
  rrf:functionName "Concat3";
  rrf:functionBody """
    function Concat3(args0,args1,args2){
      return String(args0) + String(args1) + String(args2);
    }
  """;
```
Listing 10. Juma Uplift mapping with a concatenation function in R2RML-F.

SML has some built-in functions, such as concatenation, while being extensible to the definition of more functions. Listing 11 shows the mapping that is generated from **Fig. 34** in SML.

```
// Mapping created using Juma Uplift.
Prefix foaf: <http://xmlns.com/foaf/0.1/>

Create View view1 As
 Construct {
    ?s1    foaf:name    ?o1. }
 With
  ?s1 = uri(concat('http://example.org/employee/', ?id))
  ?o1 = plainLiteral(concat(?last_name,", ",?first_name))
 From
```

71

```
employees
```

4.3.3. Juma Interlink

The third application of the Juma approach that has been developed by the author of this thesis, called Juma Interlink, is concerned with the representation of semantic mappings. Juma Interlink generates executable mappings in the form of SPARQL CONSTRUCT queries. As mentioned before, SPARQL is a W3C Recommendation, widely used, and supported by many applications. Moreover, literature has shown that SPARQL can be used to represent a wide variety of semantic mappings (Meehan, Brennan, Lewis, & O'Sullivan, 2014a; Rivero & Hernández, 2011). A tutorial video of Juma Interlink is available[42].

Juma Interlink makes a distinction between simple and complex semantic mappings. As stated in Section 2.4, simple mappings relate one entity to another (one-to-one); complex mappings describe relationships between multiple entities (one-to-many, many-to-one, and many-to-many) (Euzenat & Shvaiko, 2007). An overview diagram of the Juma Interlink application is presented in **Fig. 36**.
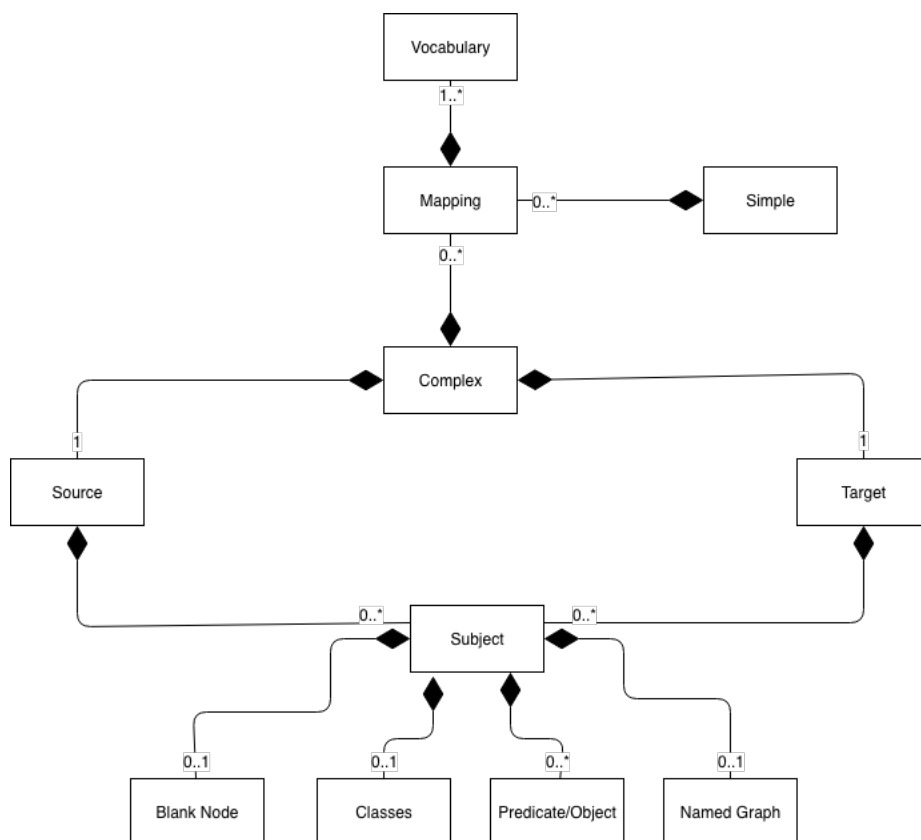
Fig. 36. Overview of the Juma Interlink application.

---

[42] `https://www.youtube.com/watch?v=23RhrKbeM50`, accessed in August 2018.

The inventory of blocks organised as categories in Juma Interlink is presented below:

- **Mapping.** This category provides the main block where vocabularies are related to simple or complex mapping definitions.
- **Vocabularies.** This category provides blocks to define the vocabularies that are going to be mapped. There are two types of vocabulary blocks. One with common predefined vocabularies and another customizable one.
- **Simple.** This category has a set of blocks where one can define a relation between two concepts (source and target). This relation can be used to define a class equivalence, a property equivalence, amongst others.
- **Complex.** This category presents users with constructs related to the definition of complex mappings.
  - **Definition.** This category presents users with a complex mapping block, where users must define source and target definitions. The source definition is used to select the elements that are going to be used in the target definition.
  - **Subject.** Blocks that define subjects are found under this category. These blocks can be related to source or target definitions. One can also find blocks to define subjects as instances of classes, or that these are blank nodes.
  - **Predicate/Object.** The blocks to define predicates and objects is defined under this category. Objects can be defined as IRI's, blank nodes, literals, to have a datatype or a language tag. These blocks are associated with a subject definition block.
  - **Functions.** Some built-in functions that can be used in the source and target definitions are available under this category. It is also possible to add more functions to Juma Interlink.
  - **Variables.** This category allows users to define variables that can be used to refer to subjects, predicates, and objects in source and target definitions. For example, the *subject* and *name* in **Fig. 37** are defined using variable blocks. The *subject* block is used to represent the same resource in the source and target definitions. The *name* variable is used as a parameter to the string before and string after functions, which are defined when clicking on the red toggle.

It is worth noting that the representation of complex mappings can be used to represent simple mappings. In (Thiéblin et al., 2016), the authors argued that simple mappings are

73

common in many applications, which is why it was decided to make this distinction in Juma Interlink.

Fig. 37 shows a complex mapping[43] where the property `foaf:name` is mapped to `foaf:givenName` and `foaf:familyName`. The property `foaf:givenName` uses the function string before, and the property `foaf:familyName` uses the function string after. Listing 12 shows the mapping as a SPARQL construct query generated by Juma Interlink.
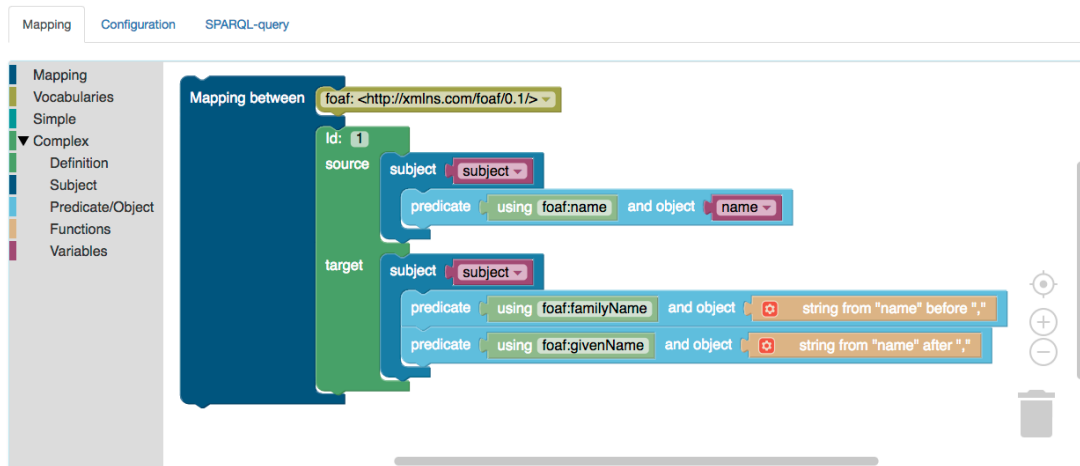
Fig. 37. Visual representation of a semantic complex mapping

```
prefix foaf: <http://xmlns.com/foaf/0.1/>
#Id1
CONSTRUCT {
      ?subject  foaf:givenName   ?result1 .
      ?subject  foaf:familyName   ?result2 .    }
WHERE {
      ?subject  foaf:name   ?name .
      BIND (STRBEFORE(?name, ",") AS ?result1)
      BIND (STRAFTER(?name, ",") AS ?result2)    }
```

Listing 12. SPARQL CONSTRUCT query generated from a complex mapping

Juma Interlink's mapping interface is separated in three tabs. In the first tab, called *Mapping*, the inventory of blocks and the workspace is shown. In *Configuration*, one can define a title and a description for the current mapping. In the *SPARQL-query* tab, the current mapping as a SPARQL construct query is presented.

As mentioned before, Juma Interlink makes a distinction between the representation of simple and complex mappings. Fig. 38 shows an example[44] of a simple mapping where the

---

[43] Based on the mappings published in (Bizer & Schultz, 2010).
[44] Based on mappings published in (Bizer & Schultz, 2010).

class `dbpedia:MusicalArtist` is mapped to `mo:MusicArtist` from the Music
Ontology[45].

Fig. 38. Visual representation of a semantic simple mapping

4.3.4. General Principles

A uniform visual representation aids readability and promotes recognizability (Preece et al., 2001). This thesis argues that using a uniform representation for equivalent constructs may aid users in the creation and editing of both uplift and semantic mappings, where becoming familiar with one representation may facilitate the understanding of the other. Moreover, uniform visual representations are beneficial and have great impact to all users but particularly on less experienced ones (Dzbor et al., 2006). General principles aimed at achieving such uniform visual representation that were adopted during the development of the applications Juma Uplift and Juma Interlink are presented in this section.

The main difference between uplift mappings and semantic mappings is their inputs, and how these are referred to in the mapping. Regardless of being an uplift or a semantic mapping, when executed, RDF is generated. As an example, see Fig. 39, which shows an uplift mapping, and Fig. 40, which shows a semantic mapping. These examples show the similarities in the definition of triples.

The main block in each type of mapping is distinct, one referencing non-RDF data and vocabularies, and the other referencing vocabularies, that are going to be used to define source and target elements. The definition of subjects, predicates and objects have a similar representation. The difference being that Juma Uplift relies on the R2RML constructs template, column and constant to express relations between input and the RDF representation. Juma Interlink uses two different blocks to define the same constructs, one where free text can be informed (`ex:name` and `foaf:name` blocks in the example) and one referring to variables (*subject* and *name* blocks in the example).

---

[45] `http://musicontology.com/specification/`, accessed in August 2018.
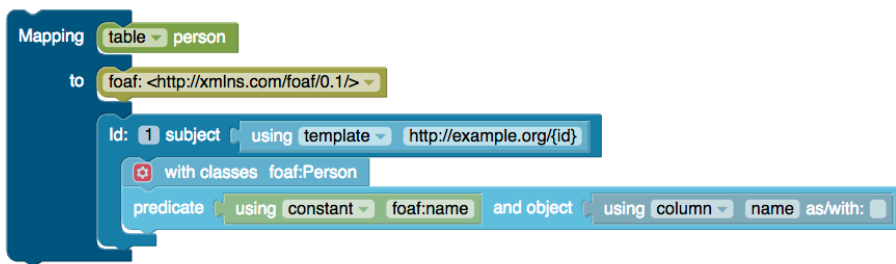
Fig. 39. Juma Uplift's general principle example.
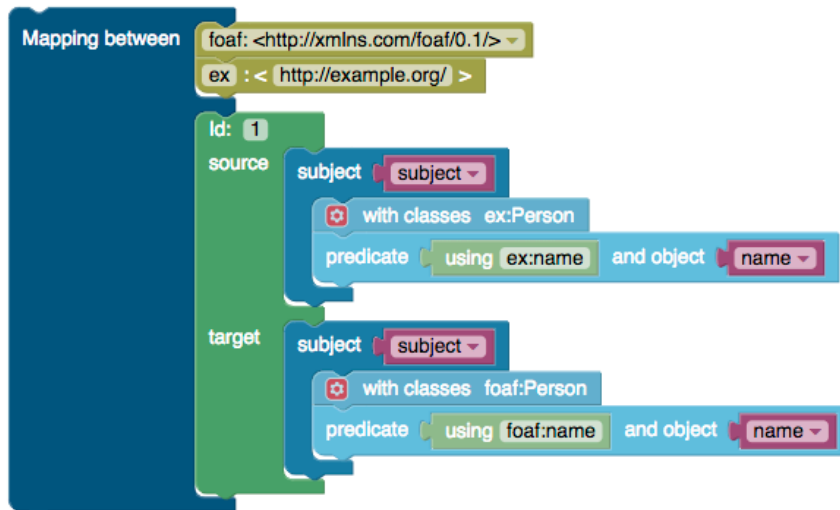
Fig. 40. Juma Interlink's general principle example.

Functions also have a uniform representation, as it has been shown for Juma Uplift (**Fig. 34**) and Juma Interlink (**Fig. 37**). Other constructs such as assigning triples to named graphs and the definition of blank nodes also have the same representation in these applications.

4.4. Limitations

Two important limitations should be noted due to Juma's approach in representing mappings using a tree structure.

Listing 13 presents an R2RML mapping in RDF TURTLE notation with two triples maps and different logical tables, referring to the same predicate object map. Juma forces a tree representation, so these elements are represented with two different predicate object pairs, as it is shown in **Fig. 41** using the Juma Uplift application.

```
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<#TriplesMap1>
  rr:logicalTable [ rr:tableName "employees"; ];
   rr:subjectMap [
      rr:template "http://example.org/employee/{id}";
   ];
  rr:predicateObjectMap <#pom_first_name> .

<#TriplesMap2>
  rr:logicalTable [ rr:tableName "person"; ];
  rr:subjectMap [
      rr:template "http://example.org/person/{id}";
  ];
  rr:predicateObjectMap <#pom_first_name> .

<#pom_first_name>
    rr:predicateMap [ rr:constant foaf:givenName; ];
    rr:objectMap [ rr:column "first_name"; ]; .
```
Listing 13. R2RML mapping representing the reuse of resources in RDF TURTLE.
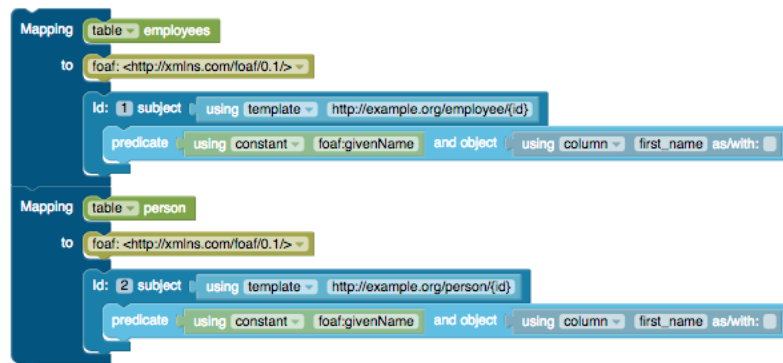
Fig. 41. Mapping represented using the Juma Uplift application with two different predicate object pairs.

The first limitation arises due to RDF being a graph data model, so that it is possible to refer to a node a number of times, such as in the above example. Being a graph, it allows for modifications to be made in one place. Modifications in SML mappings can also be done in the one place, since it is possible to refer to the same variable multiple times. In Juma, such modifications would be done in multiple places due to the tree representation.

The second limitation is also related to RDF being a graph data model. Every representation in Juma yields a valid mapping, but the reverse might not always be true. For example, to load the mapping presented in Listing 13 into Juma R2RML or Juma Uplift, the R2RML document would need to be pre-processed in order to rewrite the graph with the necessary duplication in order to obtain a tree representation. As mentioned previously, SML mappings allow for variables to be referred to in multiple places, thus, a similar transformation would be necessary when loading SML mappings into Juma.

4.5. Chapter Summary

This chapter has presented the requirements derived from the state of the art, which were used to design the Juma approach for the representation of mappings in Linked Data, also presented in this chapter. Juma represents mappings using a block metaphor. The blocks contain one or more placeholders for information (UI elements or blocks) and each mapping is a tree of such blocks.

This chapter has also presented three applications that apply the Juma approach in the representation of mappings. The applications Juma R2RML and Juma Uplift provide support for uplift mappings, while the Juma Interlink application provides support for semantic mappings. Juma Uplift and Juma Interlink apply general principles to the visual representation used. It is argued that this will make it more intuitive for the same user to engage in different uplift and semantic mapping tasks. Juma R2RML was the first application that applied the Juma approach to the R2RML mapping language. The second application, Juma Uplift, has a higher level of abstraction, supporting two distinct uplift mapping languages, these being R2RML and SML. Juma Uplift also supports the representation of data transformation functions. The third application applied the Juma approach to semantic (simple and complex) mappings that automatically generates executable mappings in the form of SPARQL construct queries.

The chapter finishes with reflections on two current limitations of the Juma approach in the representation of mappings.

5. Evaluations

This chapter describes and presents the findings of the five experiments undertaken to evaluate the Juma approach for Linked Data mappings. The findings are grouped together according to three different aspects of Juma that we have evaluated: creation (and editing), understanding, and expressiveness, as is shown in **Fig. 42**. Section 5.1 provides an overview of each of the experiments in this chapter. Section 5.2 presents the user evaluation instruments used in user based experiments. Section 5.3 presents the first two experiments evaluating the Juma approach in the creation of uplift mappings. Section 5.4 presents experiment 3, which evaluated the Juma approach in the understanding of uplift mappings. Section 0 presents the fourth and fifth experiments, which evaluated the expressiveness of the Juma approach in the representation of uplift and of semantic mappings. Section 5.6 presents the overall conclusions. Section 5.7 ends the evaluation chapter with a summary.
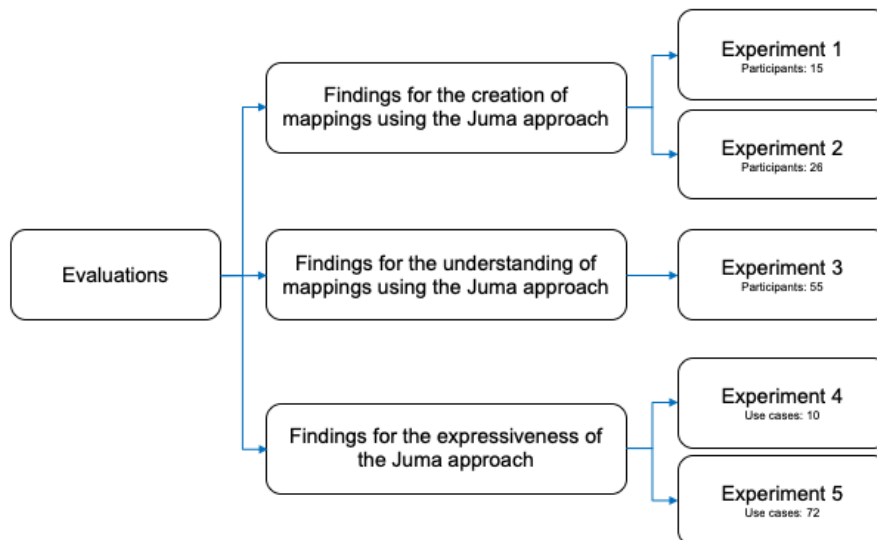
Fig. 42. Evaluation diagram of this thesis

5.1. Experiment summaries

This section provides an overview of the experiments undertaken to evaluate Juma.

Experiments 1, 2, and 3 are user experiments that evaluate the Juma approach in the creation and understanding of mappings. In these experiments, Juma is evaluated considering the **performance** achieved by participants in a specific task. Performance, in this thesis, is defined as the accuracy of the output generated from executing the produced mappings in comparison to a gold standard. For experiments 1 and 2, this task involves participants creating an uplift mapping, and their performance is calculated by counting the number of correct triples that are generated when executing these mappings. In experiment

3, this task involves the participant's understanding of a mapping, which is defined as their ability to deduce the triples that a mapping representation would generate when executed.

Experiments 1 and 2 also evaluated the **usability** of the Juma approach using the PSSUQ (Lewis, 1992) questionnaire, which is presented in Section 5.2.1. Experiments 2 and 3 evaluated the **mental workload** of participants in creating and understanding mappings using the Workload Profile (WP) (Tsang & Velazquez, 1996) and the NASA Task Load Index (NASA-TLX) (Hart, 2006) instruments, which will be presented in Section 5.2.2.

Another aspect assessed in these experiments is how the evaluated aspects – performance, usability, and mental workload – compare between the groups of participants and correlate to each other. Cronbach's alpha, which is presented in Section 5.2.3, is used to show the reliability of the usability and mental workload questionnaires applied in these experiments.

Experiments 4 and 5 evaluated the **expressiveness** of Juma in the representation of uplift and semantic mappings, respectively. It is important to note that these experiments are evaluating the applications Juma R2RML, Juma Uplift, and Juma Interlink, that apply the Juma approach in the representation of mappings in Linked Data. **Fig. 43** presents the aspects evaluated in each experiment. Summaries of each experiment presented in this chapter are presented next.
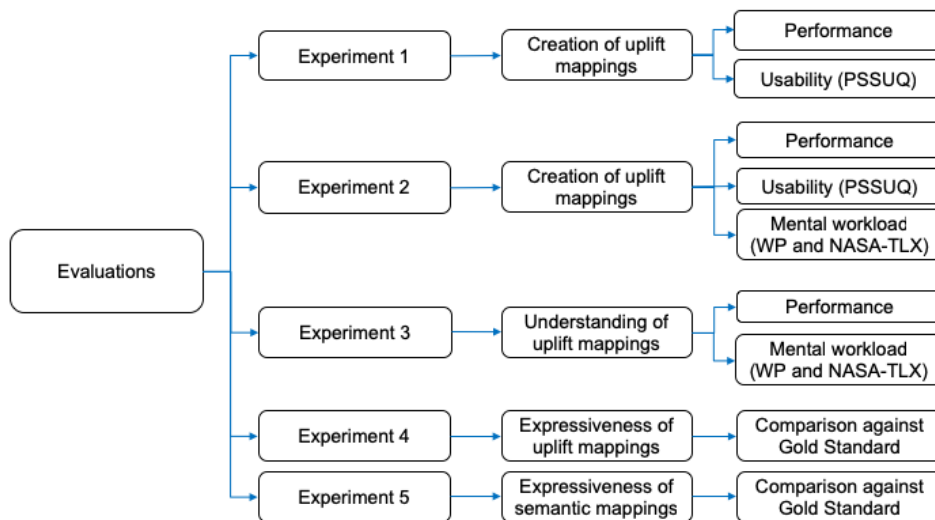
Fig. 43. Aspects evaluated in each experiment

Experiment 1:

The purpose of the first experiment is to evaluate the application of the Juma approach in the creation of R2RML mappings. This experiment evaluated the Juma approach considering users with different background knowledge in the creation of one uplift mapping using the Juma R2RML application. There were 15 participants in this experiment. These users were

categorized based on their expertise in one of these groups: web developers, knowledge engineers, and users familiar with the R2RML mapping language. Our intuition is that these users are more likely to create and publish Linked Data datasets. In this experiment, Juma is evaluated considering the performance achieved by participants in a uplift mapping task, and their perceived usability. Some of the findings of this experiment have been published in (Crotti Junior, Debruyne, & O'Sullivan, 2017, 2018c).

Experiment 2:

Similar to experiment 1, the second experiment also evaluates the Juma approach in the creation of uplift mappings. In contrast with experiment 1, this evaluation compares the results of using the Juma approach to those of creating mappings by means of a text editor – which is common for developers to use. In addition, this evaluation also measures the mental workload of creating uplift mappings, which was not evaluated in experiment 1. There were 26 participants from a third level M.Sc. class taking part in this experiment. Participants were randomly split into two groups. One group used the Juma Uplift application to create an uplift mapping. The second group was asked to create the same mapping "manually" in R2RML RDF TURTLE syntax, using their preferred text editor. In this experiment, Juma is evaluated by comparing the performance, the perceived usability, and the perceived mental workload of participants using the Juma Uplift application to participants who "manually" crafted mappings in R2RML. Part of the findings of this experiment have been published in (Crotti Junior, Debruyne, Longo, & O'Sullivan, 2018).

Experiment 3:

The purpose of the third experiment is to evaluate Juma for the task of understanding uplift mappings. In this experiment, the understanding of a mapping represented in Juma is compared to the understanding of a mapping represented in R2RML. Participants were recruited through email from universities and the research community. This experiment was executed through an online survey by 55 participants. The survey would show either the mapping represented using the Juma Uplift representation or an equivalent mapping represented in R2RML (in RDF TURTLE). Participants were then asked one multiple-choice question. In this question, participants should select all triples that the mapping representation would generate when executed. This experiment evaluated Juma by comparing the performance and mental workload of participants understanding a mapping represented using the Juma Uplift application to participants understanding an equivalent mapping represented in R2RML.

Experiment 4:

The purpose of the fourth experiment is to evaluate the expressiveness of the Juma approach in the representation of uplift mappings. This experiment shows that Juma, although first conceived for R2RML (with the Juma R2RML application), is capable of representing mappings that can be generated in multiple uplift mapping languages (with the Juma Uplift application). For this experiment, 10 use cases with a respective expected RDF output were defined. These use cases were based on the work presented in (J. Sequeda et al., 2012). Juma Uplift was then used to create a mapping for each of the use cases. The mappings generated by Juma Uplift were then executed against its respective engines and compared to the expected RDF output. Part of the findings of this experiment have been published in (Crotti Junior, Debruyne, & O'Sullivan, 2018b).

Experiment 5:

The purpose of the fifth experiment is to evaluate the expressiveness of the Juma approach in the representation of semantic mappings. For this experiment, the Juma Interlink application was used to represent 72 mappings between Linked Data datasets. The mappings used in this experiment were devised by the R2R Framework research (Bizer & Schultz, 2010). Some of the findings of this experiment have been published in (Crotti Junior, Debruyne, & O'Sullivan, 2018a).

5.2. User Evaluation instruments

In this section, the evaluation instruments used in the user experiments (1 to 3) are presented in this chapter. The PSSUQ questionnaire, which is presented below, is used in experiments 1 and 2. The mental workload instruments presented in Section 5.2.2 are used in experiments 2 and 3. The Cronbach's alpha measurement, presented in 5.2.3, is used to show the reliability of the self-reporting questionnaires used in experiments 1, 2, and 3.

5.2.1. PSSUQ questionnaire

The usability test used in this thesis is the Post-Study System Usability Questionnaire (PSSUQ) questionnaire (Lewis, 1992). The PSSUQ questionnaire was designed to assess overall satisfaction with system usability and was chosen over other questionnaires, like the System Usability Scale (SUS) (Brooke, 1996), as it explicitly assesses other aspects of a system beyond usability. For instance, the PSSUQ questionnaire specifically gives scores

for interface quality, which is of particular importance in the evaluation of visual representations. Furthermore, PSSUQ was designed for scenario-based usability studies, where some questions are more targeted, such as "I was able to complete the tasks and scenarios quickly using this system". PSSUQ also allows for more nuanced responses by using a 1-7-point Likert scale, in contrast with the 1-5 Likert scale used by SUS. PSSUQ is a 19-item questionnaire (see Appendix A for the full questionnaire) with a scale from 1 (strongly agree) to 7 (strongly disagree), it also has a not applicable option (N/A) and a comment area in each question. The PSSUQ questionnaire gives scores in four aspects:

- System usefulness (SysUse): average the responses to questions 1 to 8;
- Information quality (InfoQua): average the responses to questions 9 to 15;
- Interface quality (IntQua): average the responses to questions 16 to 18;
- Overall: average the responses to questions 1 to 19. The last dimension thus considers an extra question that inquires the participant about their overall impression of the system.

It is **important to note** that the lower a PSSUQ score is, the better. This will be important when comparing and correlating PSSUQ scores.

5.2.2. Mental workload instruments

Human mental workload (MWL) is a fundamental design concept used to investigate the interaction of human with computers and other technological devices (Longo, 2015). Mental workload instruments quantify the cognitive load associated to performing a task and can be used to describe user experience (Longo, 2015). Literature suggests that both mental overload and underload can affect performance (Cain, 2007). In this thesis, we apply the Workload Profile and the Nasa Task Load Index instruments for the task of creating and understanding mappings in Linked Data. These instruments have been successfully applied in several fields, such as in health care (Longo, 2016) and the Linked Data domain (Hoefler, Granitzer, Veas, & Seifert, 2014; Thakker, Dimitrova, Lau, Yang-Turner, & Despotakis, 2013).

  The **Workload Profile** (WP) assessment procedure (Tsang & Velazquez, 1996) is built upon the Multiple Resource Theory proposed in (Wickens, 1992, 2008). In this theory, individuals are seen as having different capacities or 'resources' related to:

- stage of information processing: perceptual/central processing and response selection/execution;
- code of information processing: spatial/verbal;

- input: visual and auditory processing;
- output: manual and speech output.

Each dimension is quantified through subjective rates (questionnaire available in Appendix B) and subjects, after task completion, are required to rate the proportion of attentional resources used for performing a given task with a value in the range $0..1 \in \Re$. A rating of 0 means that the task placed no demand while 1 indicates that it required maximum attention. The aggregation strategy is a simple sum of the 8 rates d:

$$WP : [0..100] \in \Re$$

$$WP = \frac{1}{8} \sum_{i=1}^{8} d_i * 100$$

The **NASA Task Load Index** (NASA-TLX) instrument (Hart, 2006) also belongs to the category of self-assessment measures. It is a combination of six factors believed to influence MWL (questionnaire available in Appendix C). Each factor is quantified with a subjective judgement coupled with a weight computed via a paired comparison procedure. Subjects are required to decide, for each possible pair (binomial coefficient, $\binom{6}{2} = 15$) of the 6 factors, "Which of the two contributed the most to mental workload during the task?", such as "Mental or Temporal Demand?", and so forth. The weights w are the number of times each dimension was selected. In this case, the range is from 0 (not relevant) to 5 (more important than any other attribute). The final score is computed as a weighted average, considering the subjective rating of each attribute d_i and the correspondent weights w_i:

$$NASATLX : [0..100] \in \Re$$

$$NASATLX = \left(\sum_{i=1}^{6} d_i * w_i \right) \frac{1}{15}$$

5.2.3. Cronbach's alpha

Cronbach's alpha coefficient estimates the internal consistency of a set of items. It is a commonly used measure for the reliability of questionnaires. Alpha values should be above 0.7, as it is suggested in literature (Nunnally, 1979). Even though the questionnaires adopted in this thesis have been widely used in other domains, Cronbach's alpha were nevertheless applied to evaluate the reliability of these instruments for the experiments proposed in this thesis.

5.3. Findings for the creation of mappings using the Juma approach

This section presents two user experiments developed to evaluate Juma in the context of creating and editing mappings. Section 5.3.1 presents the introduction. Section 5.3.2 presents the mapping task used in experiments 1 and 2. Sections 5.3.3 and 5.3.4 present experiments 1 and 2, respectively.

5.3.1. Introduction

This section presents two user experiments conducted to evaluate the Juma approach in the creation of uplift mappings. As mentioned before, uplift mappings are used to express how non-RDF data is transformed into an RDF representation.

Experiment 1 focuses on users with different background knowledge using the application Juma R2RML to create one mapping. Experiment 2 is concerned with comparing the Juma approach to "manually" crafting mappings in R2RML. Experiments 1 and 2 evaluated Juma considering the performance, and perceived usability of participants. Additionally, in contrast with experiment 1, the second experiment also assesses the mental workload of creating uplift mappings using Juma and a text editor. The mapping task for these experiments is presented in the next section.

5.3.2. Mapping Task

The mapping task was built on top of the Microsoft Access 2010 Northwind sample database that has been ported to MySQL[46]. The table being mapped, as shown to participants, is presented in Fig. 44. A summary of the mapping task separated in three parts is presented next.

---

[46] Available at `https://github.com/dalers/mywind`, accessed in August 2018.

Fig. 44. Table employees[47].

- **Part 1**: in this part, participants had to define a mapping with one subject per row of the table *employees*. The subject URI for the triples should be `http://data.example.org/employee/{id}`. These subject should be declared instances of the `foaf:Person` class from the FOAF[48] vocabulary. The mapping definition should also create, for these subjects, the predicate `foaf:givenName` with object from the column *first_name*. The predicate `foaf:familyName` with object from the column *last_name*. Finally, the predicate `foaf:name` should have as object the concatenation of the columns *last_name* and *first_*name separated by a comma;

- **Part 2**: in the same mapping, participants were asked to define another subject from the table *employees*. The subject URI should be `http://data.example.org/city/{city}`. These subjects should be declared instances of the class `foaf:Spatial_Thing`. The mapping should generate the predicate `rdfs:label`, from the RDFS[49] vocabulary, with object from the column *city*;

- **Part 3**: in the last part, participants were asked to link the subject from Part 1 with the subject from Part 2 using the predicate `foaf:based_near`.

---

[47] Adapted from `https://github.com/dalers/mywind/blob/master/northwind-erd.pdf`, accessed in August 2018.
[48] `http://xmlns.com/foaf/0.1/`, accessed in August 2018.
[49] `http://www.w3.org/2000/01/rdf-schema#`, accessed in August 2018.

It is important to note that the instructions given to participants included a sample of the RDF output, and that users could run their mapping and compare it to the sample provided. The performance of participants is calculated by executing the mappings produced during the experiment and comparing it to the expected output (gold standard). A perfect solution would have 53 triples over the 9 records from the table *employees* of the Northwind database. Part 1 should have 36 triples. Part 2 should have 8 triples, and Part 3 should have 9 triples. It is also important to note that the correctness of the solutions and generated triples were examined, as there are multiple correct solutions (some elements of the task could be achieved in different ways). Table 3 shows the challenges associated to the mapping task.

Table 3. Challenges associated to the task

Part	Short description	Challenge/Non-trivial aspects
#1	Map and type entities to a class with three attributes	Map an entity with one attribute being the concatenation of other two attributes.
#2	Map and type another entity with one attribute	Map a second entity from the same input.
#3	Linking	Linking subjects created previously in Part 1 and Part 2.

5.3.3. Experiment 1: Evaluation of Juma in the creation of R2RML mappings

This section presents the first evaluation of the Juma approach for the creation of uplift mappings.

5.3.3.1. Hypothesis

This hypothesis is concerned with users with different background knowledge (i.e. experts and non-experts in Semantic Web technologies) being able to use Juma to create R2RML mappings, and is stated as follows:

- **Hypothesis H1**: the creation of R2RML mappings using the Juma approach (applied to R2RML) yields high task performance with sufficient usability for users with different background knowledge.

The usability was evaluated using the PSSUQ questionnaire, which gives scores using a 7-Likert scale from 1 (strongly agree) to 7 (strongly disagree). Sufficient usability is defined in this experiment as values that are smaller than half the scale used by PSSUQ, i.e. values

that are smaller than 3.5. As stated in Section 5.2.1, small values indicate better usability. We consider task performance above 75% to be high.

5.3.3.2. Methodology

To test the aforementioned hypothesis **H1**, a user experiment targeting three groups of users based on their background knowledge was executed. As mentioned before, participants were grouped as web developers, knowledge engineers, or familiar with the R2RML mapping language. Our intuition is that these are users more likely to create and publish Linked Data datasets. Participants of this experiment were asked to create one R2RML mapping using the Juma R2RML application. The mapping task and how the performance of participants is calculated was presented in Section 5.3.2. Usability was evaluated using the PSSUQ questionnaire. The group's performance and usability are also compared using statistical tests, to check whether the differences between such groups can be considered significant. This experiment was executed individually with each participant. Furthermore, there was no time limit for the duration of the experiment.

5.3.3.3. Procedure

The study was structured as follows:

- **Informed consent**: in this part, we explained the experiment to participants. At the end, only the participants that consented to that information proceeded with the experiment.
- **Pre-questionnaire**: participants were asked to evaluate their knowledge in relevant fields (Semantic Web technologies and more specifically about the R2RML mapping language). Participants evaluated their familiarity using a 7-point Likert scale from 1 (strongly agree) to 7 (strongly disagree). Note that their response to this questionnaire was used to group them based on their background knowledge, as it will be explained in the following section.
- **Technical debriefing**: after filling out the pre-questionnaire, participants had the opportunity to watch videos about RDF, R2RML and Juma. Participants were able to skip the material if they desired to do so, or access this material during the experiment[50].

---

[50] Experiment material, expected mapping and expected output are available at
`https://www.scss.tcd.ie/~crottija/juma/material/` .

- **Mapping task**: in the main part of this experiment, we asked participants to create one R2RML mapping using Juma R2RML. The task was described in Section 5.3.2. Participants could ask questions and any help needed to solve the task was recorded.
- **Post-questionnaire**: after completion of the task, we asked participants to fill out a questionnaire about the usability of Juma R2RML. At this stage, we have also conducted an informal interview with participants.

5.3.3.4. Participants

Participants were recruited through email, and categorized in groups based on their responses to the familiarity questionnaire. For this experiment, we focused on users that we considered more likely to need to create and publish Linked Data datasets. We consider that these users would be one of three types: non-experts in Semantic Web and R2RML, but with background in computer science and/or web development; Semantic Web experts with no knowledge of R2RML; and Semantic Web experts with knowledge of the R2RML mapping language. The groups were defined as follows:

- **Web Developers (WD)**: these participants had background in computer science with experience on web development, but not in Semantic Web technologies or the R2RML mapping language.
- **Knowledge Engineers (KE)**: these participants had knowledge in Semantic Web technologies such as RDF, OWL and so on. Furthermore, these users are not familiar with R2RML.
- **R2RML familiar (RF)**: participants in this group considered themselves familiar with the R2RML mapping language.

This experiment was executed with 15 participants, 5 in each group, thus 10 participants had no knowledge of R2RML. As mentioned before, participants were asked to answer a questionnaire using a 7-Likert scale to evaluate their familiarity with Semantic Web technologies and R2RML. If their answer was in the 1 to 4 range, then they were considered familiar with the technology.

5.3.3.5. Results

This section is broken down into: task performance, usability, correlations, reliability, discussion, and conclusions.

5.3.3.5.1. Task Performance

Table 4 presents the task performance, which is calculated using the total number of correct triples generated from executing the mappings, as explained in Section 5.3.2. The time (in minutes) taken for participants to complete the task is also shown in this table. Participant numbers are shown in the first column. Note that participants 1 to 5 are in the WD group, 6 to 10 are in the KE group, and 11 to 15 are in the R2RML familiar group. **Table 5** shows the same results per group.

Table 4. Task performance (TP) and time per participant.

#	Part 1 (36)			Part 2 (8)			Part 3 (9)			Combined (53)		
	Total	TP %	Time	Total	TP %	Time	Total	TP %	Time	Total	TP %	Time
1	36	100.00	25	8	100.00	5	9	100.00	10	53	100.00	40
2	36	100.00	18	8	100.00	10	0	0.00	10	44	83.10	38
3	36	100.00	16	8	100.00	6	9	100.00	7	53	100.00	29
4	27	75.00	25	8	100.00	7	0	0.00	8	40	75.50	40
5	36	100.00	17	8	100.00	7	9	100.00	10	53	100.00	34
6	36	100.00	13	8	100.00	4	9	100.00	8	53	100.00	25
7	36	100.00	6	8	100.00	3	9	100.00	3	53	100.00	12
8	36	100.00	21	8	100.00	9	0	0.00	11	44	83.10	41
9	36	100.00	18	8	100.00	7	9	100.00	11	53	100.00	36
10	36	100.00	14	8	100.00	7	9	100.00	6	53	100.00	27
11	36	100.00	15	8	100.00	3	9	100.00	5	53	100.00	23
12	36	100.00	9	8	100.00	4	9	100.00	7	53	100.00	20
13	36	100.00	17	8	100.00	6	9	100.00	5	53	100.00	28
14	36	100.00	13	8	100.00	4	9	100.00	10	53	100.00	27
15	36	100.00	5	8	100.00	4	9	100.00	4	53	100.00	13

Table 5. Task performance (TP) and time averages per group.

Group	Part 1 (36)			Part 2 (8)			Part 3 (9)			Combined (53)		
	Total	TP %	Time	Total	TP %	Time	Total	TP %	Time	Total	TP %	Time
WD	34.2	95.00	20.2	8	100.00	7	5.4	60.00	9	47.6	89.80	36.2
KE	36	100.00	14.4	8	100.00	6	7.2	80.00	7.8	51.2	96.60	28.2
RF	36	100.00	11.8	8	100.00	4.2	9	100.00	6.2	53	100.00	22.2
All	35.4	98.30	15.5	8	100.00	5.7	7.2	80.00	7.7	50.6	95.50	28.8

The performance achieved by the RF (R2RML Familiar) group was higher than the one achieved by the KE (Knowledge Engineer) group, which was higher than the one achieved by the WD (Web Developer) group. The time taken to execute each part of the mapping task also follows the same pattern. Furthermore, the task performance achieved by each

individual group is considered high. **Table 6** shows the p-values of the independent two sample Welch T-Test and the Wilcoxon test for the performance between groups. The main difference between these tests is that the Welch T-Test assumes normality of the data, and the Wilcoxon does not. Taking a threshold of 0.05, the table seems to indicate there are no significant differences between the groups with respect to the performance data. It should be noted, however, that the p-values for the comparison between WD and RF are closer to the threshold.

Table 6. Welch and Wilcoxon p-values for the performance between groups.

Performance	Welch	Wilcoxon
WD vs. KE	0.40	0.52
WD vs. RF	0.21	0.18
KE vs. RF	0.36	0.42

The Welch T-test and the Wilcoxon test can only be used to compare two groups at a time. In order to compare the three groups simultaneously we use Analysis of Variance (ANOVA). Applying one way ANOVA, with 2 as the degree of freedom between groups and 12 within groups, results in a p-value of 0.30. Also suggesting that there is no statistically significant difference between the performance of the groups, when considering the same threshold of 0.05.

During the experiment, some participants needed help in the execution of the task. In Part one, 2 Web developers and 1 Knowledge engineer needed help. The only help needed in Part one was on concatenating two columns. In Part two, none of the participants asked for help. In Part three, 4 Web developers and 2 Knowledge engineers asked for help. This help was related to defining parent and child values for the join condition.

5.3.3.5.2. Usability

Table 7 shows the PSSUQ responses and scores per participant. **Table 8** shows the average (AVG) and standard deviation (STD) per group, for the same data. These results show that the RF group has the best usability scores (i.e. smaller values).

Table 7. PSSUQ responses per participant,

PSSUQ	Participant														
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
Q1	4	3	2	2	2	2	3	6	7	4	2	1	1	2	2
Q2	5	4	2	3	1	1	2	2	7	5	2	2	1	5	2
Q3	3	3	2	4	1	2	2	3	2	3	1	2	2	2	1
Q4	6	4	3	3	1	2	4	2	5	4	2	3	2	2	2
Q5	6	2	4	2	2	2	3	3	6	4	2	1	2	3	2
Q6	3	3	1	3	1	2	2	2	7	5	2	1	1	1	1
Q7	3	1	1	4	1	1	1	2	6	4	1	1	2	1	2
Q8	1	1	1	3	1	2	3	2	2	2	1	2	1	2	2
Q9	2	6	3	2	3	5	2	4	7	6	4	1	2	6	4
Q10	3	2	2	4	2	1	6	2	5	5	4	2	1	4	2
Q11	2	1	3	4	1	2	2	3	3	5	3	1	1	6	3
Q12	4	1	4	2	1	3	2	3	6	6	3	2	1	3	2
Q13	4	1	1	3	2	1	2	3	5	5	1	2	1	3	2
Q14	3	1	2	3	1	2	2	3	6	5	1	1	1	3	1
Q15	2	2	1	3	3	1	2	2	3	5	1	1	1	4	2
Q16	1	1	1	2	2	1	2	1	2	3	1	1	1	1	2
Q17	1	1	1	2	2	1	2	2	3	2	1	1	1	1	2
Q18	1	3	3	2	2	3	4	1	5	4	1	2	1	2	5
Q19	3	2	2	2	1	2	3	2	5	3	1	1	1	2	3
SysUse	3.9	2.6	2.0	3.0	1.3	1.8	2.5	2.8	5.3	3.9	1.6	1.6	1.5	2.3	1.8
InfoQual	2.9	2.0	2.3	3.0	1.9	2.1	2.6	2.9	5.0	5.3	2.4	1.4	1.1	4.1	2.3
IntQual	1.0	1.7	1.7	2.0	2.0	1.7	2.7	1.3	3.3	3.0	1.0	1.3	1.0	1.3	3.0
Overall	3.0	2.2	2.1	2.8	1.6	1.9	2.6	2.5	4.8	4.2	1.8	1.5	1.3	2.8	2.2

Table 8. PSSUQ average scores per group.

PSSUQ	WD		KE		RF		All	
	AVG	SD	AVG	SD	AVG	SD	AVG	SD
SysUse	2.6	0.99	3.2	1.36	1.8	0.29	2.5	0.74
InfoQual	2.4	0.51	3.6	1.46	2.3	1.17	2.8	0.71
IntQual	1.7	0.41	2.4	0.86	1.5	0.84	1.9	0.47
Overall	2.3	0.57	3.2	1.25	1.9	0.61	2.5	0.74

We did assess the normality of the data, which is shown in **Table 9**, and checked for outliers, using a boxplot presented in **Fig. 45**. Assuming a threshold of 0.05, we can observe in **Table 9** that the aspect interface quality for the RF group is not normal. It can be observed 2 outliers (participants #9 and #10 from the KE group) in the data (**Fig. 45)**. Nonetheless, it was decided not to exclude any data for the analysis of this experiment due to the sample size – 5 participants per group.

Table 9. Shapiro-Wilk normality test p-values.

PSSUQ	WD	KE	RF	All
SysUse	0.99	0.73	0.10	0.06
InfoQual	0.42	0.16	0.45	0.07
IntQual	0.15	0.50	0.01	0.06
Overall	0.79	0.38	0.78	0.07
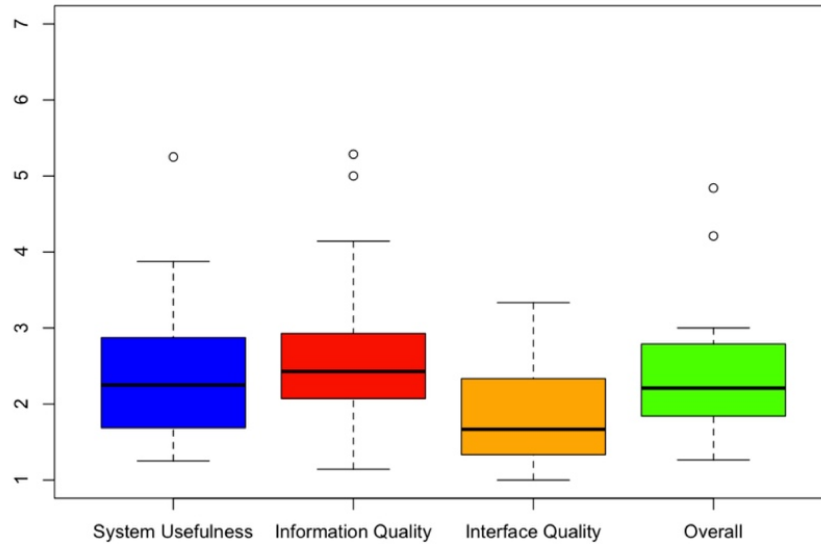
Fig. 45. Boxplot of the PSSUQ scores of all participants (i.e. not separated by groups).

The two sample independent Welch T-Test and the Wilcoxon test between the groups were applied in every aspect of PSSUQ. **Table 10** shows these p-values, where the only significant different was found by the Wilcoxon test between the KE and RF groups.

Table 10. PSSUQ Welch and Wilcoxon p-values between the groups.

PSSUQ	Welch			Wilcoxon		
	WD vs. KE	WD vs. RF	KE vs. RF	WD vs. KE	WD vs. RF	KE vs. RF
SysUse	0.40	0.15	0.07	0.60	0.21	0.03
InfoQual	0.15	0.85	0.17	0.25	0.75	0.15
IntQual	0.14	0.76	0.15	0.29	0.39	0.11
Overall	0.20	0.29	0.08	0.42	0.29	0.10

One way ANOVA was used to compare the usability scores between the three groups, which are presented in **Table 11**. The degree of freedom between the groups is 2 and within groups is 12. These results indicates that, with a threshold of 0.05, the difference between the groups is not statistically significant.

Table 11. ANOVA's p-values between PSSUQ score's groups.

ANOVA	p-value
SysUse	0.10

InfoQual	0.18
IntQual	0.17
Overall	0.09

5.3.3.5.3. Correlations

The Pearson's and Spearman's tests were used to correlate task performance and PSSUQ scores. For the WD group these tests shown weak negative relations. For the KE group, there were positive relations, but weaker than the ones found for the WD group. For the RF group, it was not possible to calculate the correlations, as these are based in the variance of the data, which is non-existent for the performance of this group as all participants achieved the same performance (100%). Nonetheless, none of these correlations were found to be statistically significant, when using a confidence level of 0.05.

5.3.3.5.4. Reliability

As stated in Section 5.2.3, Cronbach's alpha shows reliability and internal consistency. The alpha index for the PSSUQ questionnaire applied in this experiment was 0.94, indicating high reliability and strong consistency in the responses.

5.3.3.6. Discussion

The performance of participants was considered high (i.e. above 75%) individually and within their own groups (**Table 5**). Moreover, the performance comparison suggests that there are no statistically significant differences between the groups (**Table 6**). It should be noted that the RF group achieved the highest task performance. Furthermore, participants from the RF group did not ask for help to complete the mapping task. We believe that this may be explained by the naming conventions used in Juma R2RML, which reflects the R2RML vocabulary.

Considering the time taken to execute the task, the WD group spent significantly more time than the other groups. The biggest difference is in the execution of Part one, which may indicate a higher learning curve for participants that are not familiar with Semantic Web technologies.

Participants needed help the most when interlinking subjects. Participants were able to use the block that interlinks subjects, but had difficulties defining the parent and child values for the join condition, which requires knowledge on SQL joins.

The best average PSSUQ scores can be seen in the R2RML familiar group, as it is shown in **Table 8**. As mentioned before, this may be explained by the naming conventions adopted by Juma R2RML. The differences between each group pair were analysed using the independent two sample Welch T-Test and the Wilcoxon test. The only significant difference was identified by the Wilcoxon test between the KE (3.2) and RF (1.8) group for the system usefulness aspect (**Table 10**). One way ANOVA did not indicate any statistical differences between the groups (**Table 11**).

Participants mentioned in an informal interview conducted after the experiment that R2RML is complex and that abstractions in such technologies help with its adoption, especially for non-expert users. Knowledge Engineers commented that they expected more help from Juma R2RML. Participants in the RF group said that they felt comfortable with Juma R2RML. Participants from the RF group also mentioned that Juma R2RML works as a template, as one does not need to know all classes and properties of the R2RML vocabulary when creating R2RML mappings. Moreover, that the system quickly shows the possible mapping constructs, only allowing the connection of blocks that create a (syntactic) valid mapping.

5.3.3.7. Conclusion

This section presented the first experiment of this thesis which evaluated users with different background knowledge on the task of creating an uplift mapping using Juma R2RML. Evaluation of performance of performance of participants in creating accurate mappings and in the perceived usability of the application was undertaken.

The performance of the participants was high individually and in each group, and the difference between the groups was found not to be statistically significant. Participants in the RF group achieved higher performance than the KE group, and the KE group's performance was higher than the WD group. The time taken to execute the task follows the same pattern. This may suggest a higher learning curve for participants with no knowledge of Semantic Web technologies. It seems that the most difficult part of the task was the interlinking of subjects, where only participants in the RF group were able to create them without any help. The PSSUQ usability scores, within each group and overall, were considered sufficient (**Table 8**). As stated previously, values smaller than half the scale used by PSSUQ (i.e. 3.5) were considered as sufficient.

Even though differences between the groups in performance and usability were identified, this experiment indicates that users with different background knowledge (which includes

experts and non-experts in Semantic Web technologies) can use Juma for the creation of R2RML mappings with **high performance and sufficient usability, confirming the hypothesis H1 of this experiment**. Furthermore, Cronbach's alpha showed a high reliability of the PSSUQ questionnaire used in the experiment, so this adds to the confidence related to usability findings.

5.3.4. Experiment 2: Evaluation of Juma in the creation of uplift mappings

This section presents the second experiment, which also evaluates Juma in the creation of uplift mappings. In contrast with experiment 1, which investigated an application of Juma with users having different background knowledge, this experiment compares the use of an application of Juma with a "manual" approach. The latter required participants to create an R2RML mapping in RDF TURTLE notation using their preferred text editor. The former applied Juma to an application that supports multiple mapping languages, R2RML and SML. For the remainder of this section, we refer to the group of participants creating mappings "manually" in R2RML as the *R2RML group*; and the group of participants using the Juma application as the *Juma group*. As in experiment 1, performance and usability are evaluated in this experiment. An additional characteristic tackled in this experiment is the **cognitive load** of creating an uplift mapping.

5.3.4.1. Hypotheses

There are three hypotheses for this experiment:
- **Hypothesis H1**: the creation of uplift mapping using Juma yields better performance when compared to a "manually" crafted approach.
- **Hypothesis H2**: the perceived usability of users interacting with Juma for the creation of uplift mappings is expected to be better than the usability experienced by users that crafted the same mappings "manually".
- **Hypothesis H3**: the perceived mental workload of users interacting with Juma for the creation of uplift mappings is expected to be lower than the perceived mental workload experienced by users that crafted the same mappings "manually".

5.3.4.2. Methodology

To test the aforementioned hypotheses a user experiment was developed and executed. This experiment was done with a third level class (see details about the participants in Section

5.3.4.4). Participants were advised to bring their computers to be able to take part in the experiment. The class was randomly divided into two groups. One of the groups would use Juma Uplift to create an uplift mapping, the other group should use their preferred text editor to create the same mapping in R2RML. The mapping task is the same used in experiment 1, which was presented in Section 5.3.2. As it was stated in the description of the mapping task, participants could execute the mapping and compare it to a sample provided. For the purpose of this experiment, a specific R2RML engine (Debruyne & O'Sullivan, 2016) was integrated into Juma Uplift. For the group creating mappings "manually", a compacted folder with the same engine and instructions on how to use it was provided. The experiment lasted 50 minutes. During the first 10 minutes, participants were able to access materials about RDF, R2RML and Juma (same material used in experiment 1). Note that only the group using Juma had access to the material about Juma and that this was the first time participants had access to this material. The next 30 minutes were used to execute the mapping task. The last 10 minutes of the experiment were used to answer the post-questionnaires on MWL and usability. To test the hypothesis **H1**, we collected and analysed the mappings created during the experiment. To test **H2**, participants answered the PSSUQ usability questionnaire. To test **H3**, participants answered the MWL questionnaires presented in Section 5.2.2. The group's performance, usability and mental workload are also compared and correlated to each other using statistical tests. The comparison between groups shows whether the differences found can be considered statistically significant. The correlations indicate how each of the aspects being evaluated may influence others.

5.3.4.3. Procedure

The procedure of this experiment had four parts:

- **Informed consent**: in this part, we explained the experiment to participants. At the end, only the participants that consented to that information proceeded with the experiment.
- **Technical debriefing**: participants had the opportunity to watch videos about RDF and R2RML prior to executing the task. The group using the Juma approach also had a presentation and a video about Juma. If they felt comfortable with these technologies, they could skip the material. The material was also available during the execution of the task.
- **Mapping task**: in this part, we asked participants to create one uplift mapping (presented in 5.3.2). Participants could ask questions to clarify any doubts about the experiment. In addition, they were advised to use the material provided first. Any help needed to solve the task was recorded.

- **Post-questionnaire**: after completion of the task, we asked participants to fill out another questionnaire on their perceived usability and mental workload.

5.3.4.4. Participants

Participants of this experiment come from the MSc module in Information and Knowledge Architecture in Trinity College Dublin 2017. The experiment was executed in week 10 of the 12-week module. At this time, the course had covered OWL modeling, RDF and SPARQL (amongst others). They also had one class before the experiment about R2RML, which included exercises.

On the day of the experiment, there were 26 participants in the class. Of these, 12 were assigned to the group using Juma, and the remaining 14 constituted the group that would create the mappings "manually". Participant allocation was done by dividing the classroom. None of these participants were involved in experiment 1.

5.3.4.5. Results

This section is broken down into: task performance, usability, mental workload, correlations, reliability, discussion, and conclusions.

5.3.4.5.1. Task Performance

Table 12 shows the task performance (TP) by participant in the R2RML group. Table 13 shows the same data for the Juma group. As mentioned before, all participants had 30 minutes to execute the mapping task.

Table 12. Task performance (TP) of the R2RML group

Participant	Part 1 (36)	TP %	Part 2 (8)	TP %	Part 3 (9)	TP %	Combined (53)	TP %
#1	36	100.00	0	0.00	0	0.00	36	67.92
#2	0	0.00	0	0.00	0	0.00	0	0.00
#3	36	100.00	8	100.00	9	100.00	53	100.00
#4	27	75.00	0	0.00	0	0.00	27	50.94
#5	36	100.00	8	100.00	9	100.00	53	100.00
#6	0	0.00	0	0.00	0	0.00	0	0.00
#7	27	75.00	0	0.00	0	0.00	27	50.94
#8	0	0.00	0	0.00	0	0.00	0	0.00
#9	0	0.00	0	0.00	0	0.00	0	0.00
#10	0	0.00	0	0.00	0	0.00	0	0.00
#11	18	50.00	0	0.00	0	0.00	18	33.96

#12	0	0.00	0	0.00	0	0.00	0	0.00
#13	0	0.00	0	0.00	0	0.00	0	0.00
#14	36	100.00	8	100.00	9	100.00	53	100.00

Table 13. Task performance (TP) of the Juma group

Participant	Part 1 (36)	TP %	Part 2 (8)	TP %	Part 3 (9)	TP %	Combined (53)	TP %
#1	36	100.00	8	100.00	9	100.00	53	100.00
#2	36	100.00	8	100.00	9	100.00	53	100.00
#3	36	100.00	8	100.00	9	100.00	53	100.00
#4	36	100.00	0	0.00	0	0.00	36	67.92
#5	36	100.00	8	100.00	9	100.00	53	100.00
#6	36	100.00	8	100.00	0	0.00	44	83.02
#7	36	100.00	8	100.00	9	100.00	53	100.00
#8	36	100.00	8	100.00	0	0.00	44	83.02
#9	36	100.00	8	100.00	9	100.00	53	100.00
#10	36	100.00	8	100.00	0	0.00	44	83.02
#11	36	100.00	8	100.00	9	100.00	53	100.00
#12	36	100.00	8	100.00	9	100.00	53	100.00

The performance achieved by the Juma group is higher than the one exhibited by the R2RML group. Comparing the performance between the groups using the independent two sample Welch T-Test results in a p-value of 0.0001, using the Wilcoxon test results in a p-value of 0.001. Taking a confidence level of 0.05, this would indicate that the difference between the groups is statistically significant.

The main issue for participants in the R2RML group was related to syntax errors in the RDF TURTLE document, which resulted in mappings that could not be executed. Participants working with Juma could not create parts of the mapping, such as concatenating values, which resulted in an incomplete, yet executable mapping. Participants in both groups could ask for help during the execution of the task. The types of errors during the experiment were tallied per group and not per participant – a decision made during the experiment as it was not foreseen beforehand. As a consequence, unfortunately there is an inability to correlate the types of errors the individual participants experienced with their performance and survey responses. The R2RML group asked for help a total number of 24 times, while the Juma group needed help a total of 6 times. **Table 14** shows these classified by the type of help needed.

Table 14. Help needed to create mappings per group

Type of help	R2RML	Juma
Syntax of the mapping	8	0
Prefix definition	1	0
Table definition	1	0
Subject definition	4	0
Class definition	2	1
Predicate/object definition	3	0
Concatenation	3	2
Linking	2	3

The final mappings created by participants in the R2RML group had several errors. These were mainly syntactic errors (7 participants). Other errors found were wrong table names, invalid R2RML documents (e.g. `rr:class` inside a predicate map), amongst others. Participants using Juma did not have syntactic errors in their mappings, since Juma "prevents" these. However, Juma does not validate the mappings semantically, which means that participants could still create mappings using the wrong classes or properties. However, these did not occur, and participants that did not achieve 100% performance in the Juma group had incomplete mappings.

5.3.4.5.2. Usability

PSSUQ was used to evaluate the usability of both approaches. **Table 15** and **Table 16** show the PSSUQ responses and scores by participant in the R2RML and Juma groups, respectively. **Table 17** shows the average scores and standard deviation per group. These results show that the Juma group perceived better usability (i.e. smaller values) for every aspect evaluated by the PSSUQ questionnaire.

Table 15. PSSUQ scores per participant for the R2RML group

PSSUQ	Participant													
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
Q1	6	2	2	2	3	3	3	2	5	4	4	3	3	3
Q2	7	2	2	3	4	2	3	2	2	2	4	3	2	3
Q3	3	2	2	3	3	3	3	2	3	5	4	3	3	3
Q4	6	2	2	3	3	4	3	1	3	4	4	4	4	3
Q5	6	2	2	4	2	3	4	1	3	4	4	4	4	3
Q6	5	2	2	4	3	1	2	1	2	1	4	3	3	3
Q7	5	2	2	3	3	1	2	1	1	1	5	3	3	3
Q8	4	2	2	3	3	1	2	2	1	1	4	3	3	3
Q9	3	7	3	4	5	2	4	2	4	4	4	5	2	3
Q10	4	7	3	3	5	2	4	2	4	4	4	4	2	3

Q11	3	7	3	3	4	2	2	2	3	4	4	4	2	3
Q12	3	7	2	3	3	1	2	2	2	4	4	4	2	3
Q13	4	7	2	2	3	1	2	2	2	4	4	3	2	3
Q14	3	7	2	3	3	2	2	2	2	3	4	4	2	3
Q15	4	7	3	4	4	3	2	2	2	4	4	4	2	3
Q16	4	7	3	3	5	3	3	2	2	4	4	4	2	3
Q17	5	7	3	2	6	3	3	2	2	4	4	3	2	3
Q18	5	7	4	3	7	2	3	2	2	4	4	3	2	3
Q19	5	7	2	2	5	2	3	2	2	4	3	4	2	3
SysUse	5.25	2.00	2.00	3.13	3.00	2.25	2.75	1.50	2.50	2.75	4.13	3.25	3.13	3.00
InfoQua	3.43	7.00	2.57	3.14	3.86	1.86	2.57	2.00	2.71	3.86	4.00	4.00	2.00	3.00
IntQua	4.67	7.00	3.33	2.67	6.00	2.67	3.00	2.00	2.00	4.00	4.00	3.33	2.00	3.00
Overall	4.47	4.89	2.42	3.00	3.89	2.16	2.74	1.79	2.47	3.42	4.00	3.58	2.47	3.00

Table 16. PSSUQ scores per participant for the Juma group

PSSUQ	**Participant**											
	#1	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**#10**	**#11**	**#12**
Q1	1	2	2	4	3	2	3	3	2	5	1	2
Q2	1	2	2	4	3	2	3	3	3	5	1	1
Q3	1	1	2	4	3	2	2	3	2	5	1	3
Q4	1	3	2	3	2	3	3	3	3	3	1	2
Q5	1	1	2	3	2	3	1	3	3	5	1	1
Q6	1	2	2	2	3	4	3	3	2	6	1	2
Q7	1	2	2	2	2	3	2	3	2	4	1	1
Q8	1	1	2	3	2	3	1	3	2	6	1	1
Q9	1	1	4	5	5	6	5	2	3	7	1	6
Q10	1	1	2	4	2	5	3	4	3	3	1	5
Q11	1	2	1	2	4	5	3	3	2	5	1	2
Q12	1	3	1	3	3	3	3	4	2	5	1	2
Q13	1	2	1	2	3	2	3	3	2	5	1	3
Q14	1	1	1	2	3	2	3	3	2	5	1	2
Q15	1	4	1	2	3	4	3	3	3	3	1	2
Q16	1	2	1	2	2	5	1	3	2	3	1	1
Q17	1	2	1	3	2	4	1	3	2	5	1	1
Q18	1	3	2	3	3	5	2	3	2	1	1	3
Q19	1	2	2	2	2	2	2	2	2	6	1	2
SysUse	1.00	1.75	2.00	3.13	2.50	2.75	2.25	3.00	2.38	4.88	1.00	1.63
InfoQua	1.00	2.00	1.57	2.86	3.29	3.86	3.29	3.14	2.43	4.71	1.00	3.14
IntQua	1.00	2.33	1.33	2.67	2.33	4.67	1.33	3.00	2.00	3.00	1.00	1.67
Overall	1.00	1.95	1.74	2.89	2.74	3.42	2.47	3.00	2.32	4.58	1.00	2.21

Table 17. Average PSSUQ scores per group

PSSUQ	R2RML		Juma	
	AVG	STD	AVG	STD
SysUse	2.9	0.94	2.34	1.05
InfoQual	3.28	1.41	2.68	1.12
IntQual	3.55	1.31	2.18	1.06
Overall	3.16	0.91	2.43	1.01

The Shapiro-Wilk normality test was applied to the R2RML and Juma groups. **Table 18** shows the p-values for this test. The p-value for information quality is below the threshold of 0.05 in the R2RML group, which suggests that its distribution is not normal.

Table 18. Shapiro-Wilk normality test per group

Shapiro-Wilk	R2RML	Juma
SysUse	0.19	0.30
InfoQua	0.01	0.66
IntQua	0.05	0.21
Overall	0.79	0.79

In **Fig. 46** it can also be seen that information quality has a data point that can be considered an outlier. Moreover, the outlier shown for system usefulness does not affect the normality of the distribution according to the Shapiro-Wilk test. **Fig. 46** and **Fig. 47** show boxplots of the PSSUQ scores for the R2RML and Juma groups, respectively.
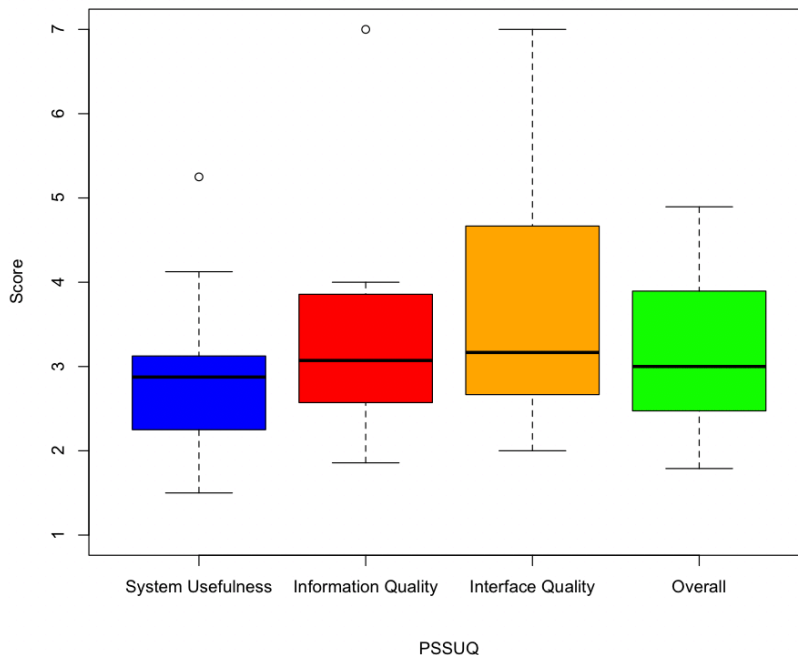
Fig. 46. Boxplot of PSSUQ's responses for the R2RML group
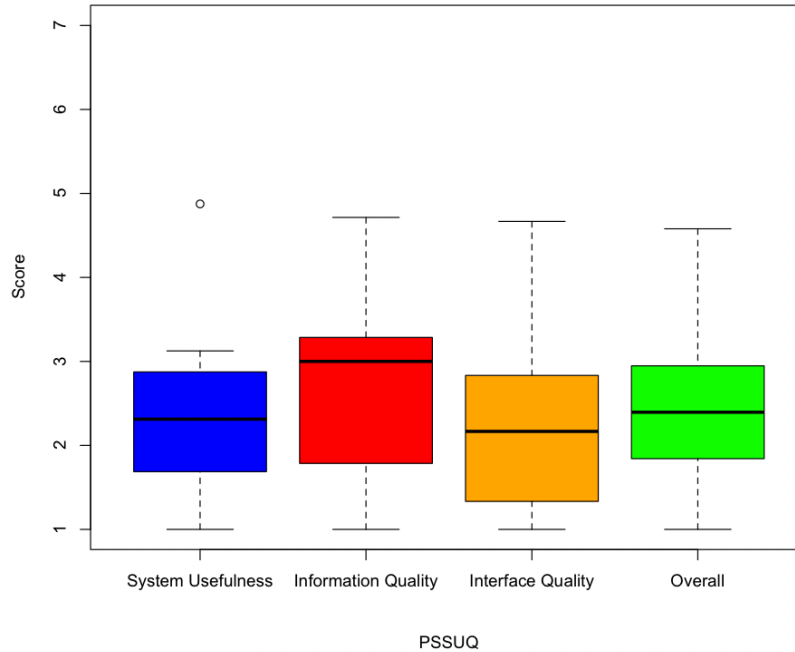
Fig. 47. Boxplot of PSSUQ's responses for the Juma group

As mentioned before, the independent two sample Welch T-Test assumes normality of the data, while the Wilcoxon test does not. Both tests have been applied in the comparison between the groups in every aspect of the PSSUQ questionnaire with all data points (**Table 19**).

Table 19. Welch and Wilcoxon tests between groups for the PSSUQ scores

PSSUQ	Welch	Wilcoxon
SysUse	0.18	0.12
InfoQua	0.23	0.34
IntQua	0.01	0.01
Overall	0.07	0.08

Interface quality shows the only statistical significant difference for both the Welch T-Test and Wilcoxon test. Removing the outliers identified in the boxplot from **Fig. 46** (participants #1 and #2 from R2RML group) and **Fig. 47** (participant #10 from Juma group), results in normality of the data, and in a statistically significant difference for the aspects system usefulness, interface quality and overall for the Welch T-Test; and for system usefulness, and interface quality for the Wilcoxon Test. Information quality, however, still does not show a statistically significant difference. **Table 20** shows the p-values for the independent two sample Welch T-Test and Wilcoxon for the PSSUQ scores between the R2RML and Juma group without outliers.

Table 20. Welch and Wilcoxon test between groups for the PSSUQ scores without outliers

PSSUQ	Welch	Wilcoxon
SysUse	0.04	0.04
InfoQua	0.24	0.37
IntQua	0.03	0.02
Overall	0.04	0.06

Participants in the R2RML group did not leave any comments. Of the 12 participants in the Juma group, 6 have left comments. A summary of the comments are presented below:

- Documentation: 2 users said that the video and presentation were helpful in the execution of the task.

- Interface: 2 users mentioned that the zoom in/out in this application of Juma is not intuitive. One user mentioned that a hover in each block with explanations would be helpful.

- Error messages: 2 users mentioned that the error messages were not clear enough.

- Other comments: 3 users mentioned that Juma requires a small amount of direction to understand it. An example of a comment which we argue to be related to the approach being "*I had to check through the video while doing the first part of the task, but I found the other parts easier after that*".

5.3.4.5.3. Mental Workload Assessment

The cognitive load imposed by the task of creating uplift mappings using the instruments presented in Section 5.2.2, was also used to evaluate Juma. **Table 21** shows the perceived mental workload for the R2RML group per participant. **Table 22** shows the same scores for the Juma group.

Table 21. Perceived mental workload scores for the R2RML group

Participant	WP	NASA-TLX
#1	45.86	65.60
#2	37.86	64.80
#3	41.28	37.80
#4	73.13	51.40
#5	27.86	35.60
#6	32.43	51.60
#7	75.29	56.80
#8	46.29	42.00
#9	71.13	62.80
#10	16.13	34.00
#11	58.43	54.40
#12	63.56	56.00
#13	63.13	73.20
#14	49.56	61.60
AVG	**50.14**	**53.40**
STD	**18.08**	**12.14**

Table 22. Perceived mental workload scores for the Juma group

Participant	WP	NASA-TLX
#1	46.86	52.60
#2	41.29	47.00
#3	36.57	31.20
#4	54.57	51.20
#5	54.72	48.00
#6	45.56	61.80
#7	57.87	57.40
#8	46	34.40
#9	54.86	52.40
#10	64.13	48.20
#11	28.43	26.40
#12	43.29	37.00
AVG	**47.85**	**45.63**
STD	**9.92**	**10.95**

The Shapiro-Wilk normality test was again applied to the R2RML and Juma groups. **Table 23** shows the p-values for this test. Considering a threshold of 0.05, these results suggest that the data is normal.

Table 23. Shapiro-Wilk normality test per group for the MWL scores

Shapiro-Wilk	R2RML	Juma
WP	0.76	0.94
NASA-TLX	0.50	0.50

Table 24 shows the p-values for the independent two sample Welch T-Test and Wilcoxon Test for the perceived mental workload between groups. Since the data follows a normal distribution the Welch T-Test should be sufficient, for consistency, we also applied the Wilcoxon test, which does not assume normality of the data. Assuming a threshold of 0.05, these p-values do not indicate a significant different between the mental workload scores between the R2RML and Juma groups. Nonetheless, the values for NASA-TLX are closer to this threshold.

Table 24. Welch and Wilcoxon test between groups for the MWL scores

MWL	Welch	Wilcoxon
WP	0.69	0.68
NASA-TLX	0.10	0.08

5.3.4.5.4. Correlations

The correlations between performance and usability have been calculated. **Table 25** shows this correlation for the R2RML group and **Table 26** for the Juma group. The Pearson correlation test assumes normality of the data, while the Spearman test does not. The Spearman test is also less sensitive to outliers than the Pearson correlation test. None of the correlations in the R2RML group were found to be statistically significant. The correlations for the Juma group were statistically significant for system usefulness, interface quality and overall. These correlations show negative values i.e. when performance goes up, the usability scores decreases (i.e. the usability improves, since small values in PSSUQ indicate better usability).

Table 25. Correlation coefficient between the PSSUQ scores and TP for the R2RML group

PSSUQ vs TP	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
SysUse	0.26	0.37	0.24	0.40
InfoQua	-0.07	0.80	0.06	0.84
IntQua	0.23	0.43	0.34	0.24
Overall	0.14	0.64	0.19	0.52

Table 26. Correlation coefficient between the PSSUQ and TP for the Juma group

PSSUQ vs TP	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
SysUse	-0.67	0.02	-0.82	0.001
InfoQua	-0.49	0.10	-0.48	0.11
IntQua	-0.66	0.02	-0.79	0.002
Overall	-0.63	0.03	-0.78	0.002

Table 27 and **Table 28** shows the correlation coefficients between performance and perceived mental workload for the R2RML and Juma groups, respectively. These coefficients indicate that as performance goes up, the mental workload score decreases. However, these are not statistically significant considering a confidence level of 0.05.

Table 27. Correlation coefficient between TP and MWL for the R2RML group

MWL vs TP	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
WP	-0.07	0.80	-0.03	0.92
NASA-TLX	-0.24	0.41	-0.17	0.56

Table 28. Correlation coefficient between TP and MWL for the Juma group

MWL vs TP	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
WP	-0.35	0.26	-0.26	0.42
NASA-TLX	-0.23	0.47	-0.21	0.51

Table 29 and **Table 30** show the correlation coefficient between usability and the mental workload instruments WP and NASA-TLX, respectively, for the R2RML group. **Table 31** and **Table 32** show the same correlation coefficients for the Juma group. None of the correlation coefficients are statistically significant in the R2RML group. All values in the Juma group are positive, indicating that high mental workload affects usability negatively, being statistically significant for the PSSUQ aspects of system usefulness, information quality and overall in relation to the WP instrument (**Table 31**).

Table 29. Correlation coefficient between PSSUQ scores and the WP index for the R2RML group

WP vs PSSUQ	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
SysUse	0.18	0.54	-0.15	0.62
InfoQua	-0.25	0.39	-0.53	0.06
IntQua	-0.51	0.06	-0.16	0.59
Overall	-0.22	0.45	-0.15	0.62

Table 30. Correlation coefficient between PSSUQ scores and the NASA-TLX index for the R2RML group

NASA-TLX vs PSSUQ	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
SysUse	0.36	0.21	0.30	0.29
InfoQua	0.09	0.76	-0.01	0.97
IntQua	-0.11	0.70	-0.13	0.66
Overall	0.18	0.54	0.21	0.46

Table 31. Correlation coefficient between PSSUQ scores and the WP index for the Juma group

WP vs PSSUQ	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
SysUse	0.72	0.01	0.59	0.04
InfoQua	0.71	0.01	0.60	0.04
IntQua	0.31	0.33	0.29	0.36
Overall	0.71	0.01	0.58	0.05

Table 32. Correlation coefficient between PSSUQ scores and the NASA-TLX index for the Juma group

NASA-TLX vs PSSUQ	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
SysUse	0.29	0.37	0.27	0.40
InfoQua	0.43	0.16	0.41	0.18
IntQua	0.44	0.15	0.23	0.47
Overall	0.39	0.21	0.37	0.24

5.3.4.5.5. Reliability

Table 33 shows the alpha indexes for the PSSUQ, WP and NASA-TLX self-reporting instruments. These results suggest that the findings produced by these instruments can be reliably considered.

Table 33. Cronbach's alpha index for WP and NASA-TLX

Instrument	Alpha index
PSSUQ	0.95
WP	0.78
NASA-TLX	0.85

5.3.4.6. Discussion

The performance of participants in the Juma group was higher than the one exhibited by the R2RML group (**Table 12** and **Table 13**), and this difference was found to be statistically significant with p-values lower than our chosen significance level of 0.05.

   Table 14 shows that participants in the Juma group requested less help, in comparison to the R2RML group, and these were focused on non-trivial aspects of the mapping, such as concatenating or interlinking. In contrast, the R2RML group requested more help, and these were focused on correcting the RDF syntax of the mapping. Nevertheless, both groups had

the same time constraint of 30 minutes to execute the task, which may indicate that Juma has a smaller learning curve.

The Juma group usability scores were better in every aspect evaluated by the PSSUQ questionnaire, with interface quality having the biggest positive difference (**Table 17**). The analysis of the normality of the data was done using the Shapiro-Wilk test, which shows that the information quality aspect of the R2RML group does not follow a normal distribution (**Table 18**). For this reason, two different tests were used to compare the groups, the independent two sample Welch T-Test, which assumes normality of the data, and the Wilcoxon Test, which does not (**Table 19**). Both tests indicate a statistically significant difference between the R2RML and Juma groups for the interface quality aspect, which is expected since participants in the R2RML were using a text editor of their choice to create their mappings. Removing the outliers identified by the boxplot in **Fig. 46** and **Fig. 47** from both groups resulted in the data being considered normally distributed, and in a statistical difference between groups for the aspects system usefulness, interface quality and overall for the Welch T-Test (**Table 20**). The Wilcoxon Test shows a statistically significant difference for the aspects system usefulness and interface quality – with the overall aspect being near to the threshold of 0.05 (**Table 20**). Information quality still does not show a statistical difference. This can be explained by the material available to participants, which was similar for both groups. Moreover, both groups used the same engine to execute their mappings i.e. errors in executing a mapping would generate the same messages. Note that the Juma Uplift application has an integrated engine which participants could use to run the working version of mapping, and that participants creating mappings "manually" used a command line interface to execute their mappings, for which a command line script was provided. Participants in the R2RML group did not leave any comments, while participants in the Juma group main comments were related to interface improvements and better error messages.

The correlation between performance and usability for the R2RML group showed weak relations, where all p-values are above the threshold of 0.05 (**Table 25**). The Pearson correlation test showed one negative coefficient for information quality, while Spearman test showed a positive value for the same correlation. Even though the correlation coefficients are weak, the positive values indicate that, as a user's performance increase, its usability scores decreases. As mentioned before, in PSSUQ, high values indicate poor scores. This may happen because users with higher performance have a better understanding of the effort involved in creating such mappings. The same correlations for the Juma group show negative values, indicating a tendency that as performance goes up, the usability improves (**Table**

26). In the Juma group, with only 12 participants, Pearson and Spearman showed a statistically strong correlation between task performance and the PSSUQ aspects system usefulness, interface quality and overall. Information quality had a much weaker correlation, which, as mentioned before, can be explained by the material available and feedback given to participants of both groups, which was similar.

The average and standard deviation of the mental workload scores found in the Juma group (**Table 22**) are smaller than the ones found in the R2RML group (**Table 21**). Nonetheless, the difference between the groups was found not to be statistically significant (**Table 24**). Another aspect that can be observed is that even though the cognitive load is not statistically different, participants in the Juma group achieved almost three times the performance in comparison to the R2RML group. The correlation between performance and mental workload scores show weak negative relations, which indicates that as performance goes up the mental workload scores decreases (**Table 27** and **Table 28**). None of the correlation coefficients between usability and mental workload scores for the R2RML group are statistically significant (**Table 29** and **Table 30**). In the Juma group, all the correlation coefficients are positive, indicating a tendency of high mental workload decreasing usability (**Table 31** and **Table 32**). This may indicate that improvements in usability can have a positive effect on perceived mental workload scores. Furthermore, Spearman and Pearson show that this correlation is statistically significant for system usefulness, information quality and overall for the WP instrument (**Table 31**).

Cronbach's alpha index showed a strong consistency of the items in PSSUQ and the mental workload questionnaires, suggesting that these are reliable instruments for the proposed experiment (**Table 33**).

5.3.4.7. Conclusion

This section presented the second experiment of this thesis, which also evaluated Juma for the task of creating uplift mappings. Participants of this experiment were split into two groups and were asked to create the same mapping using different approaches. One of the groups should use the Juma Uplift application, while the other should use their preferred text editor. This experiment compared the use of Juma Uplift and a text editor with respect to user performance on the proposed task, their perceived usability and their mental workload. The experiment was executed in a third level MSc class in 50 minutes. Participants had the same instructions and had access to mostly the same material. The only difference being that

participants using Juma had one extra presentation and video about Juma. Therefore, participants creating mappings "manually" were not aware of Juma.

The **results** presented in this section **confirm the hypothesis H1** of this experiment, **which stated that participants using Juma achieve higher performance** when compared to creating mapping "manually". Moreover, the difference between the groups were found to be statistically significant using the independent two sample Welch T-Test and the Wilcoxon Test with a significance level of 0.05. Another aspect evaluated was the help needed by participants, while participants in the Juma group could focus on non-trivial aspects of the mapping, participants in the R2RML group had to focus fixing the RDF syntax of the mapping, and other trivial aspects.

The usability for the Juma group was better in every aspect, especially for interface quality, **which confirms the hypothesis H2** of this experiment. Moreover, the differences between the group's usability results, when removing outliers from the data, were statistically significant for the system usefulness, interface quality and overall aspects of PSSUQ for the Welch T-Test; and for system usefulness and interface quality when applying the Wilcoxon test. The difference between the group's information quality scores were found not to be statistically significant. It should be noted that this was expected, as the material and feedback given to participants during the task were similar.

The **mental workload scores of the Juma group were slightly smaller**, but not statistically significant, **which may indicate that the hypothesis H3 is true**. Nonetheless, **our conclusion is that more evidence needs to be gathered**. As mentioned before, even though the mental workload scores are similar, the performance of this group was almost three times higher. Statistically significant correlations were found, for the Juma group, between the mental workload instrument WP and the PSSUQ scores of system usefulness, interface quality and overall. These were positive correlations, meaning that as the WP score increases, the usability aspects decrease (since high values suggest poor usability in PSSUQ). As mentioned in the previous section, this may indicate that improvements in usability can have a positive effect on the perceived mental workload.

Cronbach's alpha indexes highlight a strong internal consistency of the items (questions) in these questionnaires, suggesting that these instruments are reliable for the proposed experiment.

5.4. Findings for the understanding of mappings using the Juma approach

This section presents the user experiment undertaken to evaluate Juma in the context of **understanding** uplift mappings. The introduction is presented in Section 5.4.1. Section 5.4.2

presents the mapping task used in this experiment. Section 5.4.3 presents the execution and results for experiment 3.

5.4.1. Introduction

This section presents experiment 3, which is a user experiment conducted to evaluate the Juma approach with respect to extent to which it facilitates understanding of uplift mappings by users.

Experiment 3 is concerned with comparing the Juma representation of mappings to its equivalent R2RML representation. This experiment was executed through an online survey. The survey would show one of two possible mapping representations. Either a mapping represented in R2RML using the RDF TURTLE notation or the same mapping represented using the Juma Uplift visual representation. Participants that entered the survey should answer the same question, regarding what would be the RDF triples that the mapping presented would generate when executed. The Juma approach was evaluated considering the performance of participants in undertaking the task, and the perceived mental workload that was involved.

5.4.2. Understanding Task

For the task in the experiment, participants were asked to analyze a small database with two tables and a mapping representation. The database is presented in **Table 34** and **Table 35**. The mapping that was represented in R2RML is presented in **Listing 14**. **Fig. 47** presents the same mapping using the Juma Uplift visual representation.

Table 34. Table Person of the database used in experiment 3

Person			
ID	**NAME**	**AGE**	**CITY_FK**
1	Ana	29	100

Table 35. Table City of the database used in experiment 3

City	
CITY_ID	**NAME**
100	Dublin

```
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix ex: <http://example.org/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<#TriplesMapMapping1>
  rr:logicalTable [  rr:tableName "city"; ];
  rr:subjectMap [
      rr:template "http://example.org/city/{city_id}";
```

```
          rr:class ex:City;
];.

<#TriplesMapMapping2>
  rr:logicalTable [ rr:tableName "person"; ];
  rr:subjectMap [
      rr:template "http://example.org/person/{id}";
      rr:class foaf:Person;
 ];

 rr:predicateObjectMap [
   rr:predicateMap [      rr:constant foaf:name;    ];
   rr:objectMap [       rr:column "name";       ];
 ];

 rr:predicateObjectMap [
    rr:predicateMap [ rr:constant foaf:based_near; ];
    rr:objectMap [
      rr:parentTriplesMap <#TriplesMapMapping1>;
      rr:joinCondition [
        rr:child "city_fk";
        rr:parent "city_id";
      ];
    ];
 ];.
```

Listing 14. Mapping used in the understanding task represented in R2RML
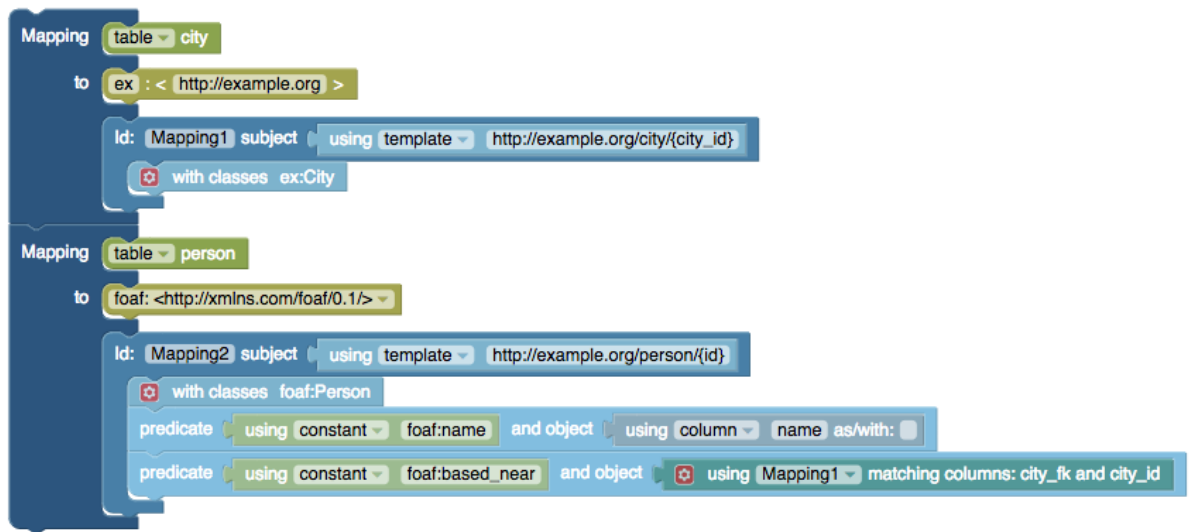
Fig. 48. Mapping used in the understanding task represented using the Juma Uplift application

After analysing the database and given mapping, participants were asked to select all the RDF triples that the presented database and mapping would generate. The multiple-choice question was: *Select all correct triples that will be generated from the mapping presented above*. The question had: 4 correct triples (1 to 4 choices), 4 somewhat correct triples (5 to 8), 4 wrong triples (9 to 12)[51] and an extra choice (13) for *I cannot make sense of the*

---

[51] Somewhat correct triples are more similar to the correct solution than the wrong triples. Nonetheless, none of these triples are the ones that the execution of the mapping would generate.

mapping. These choices were randomized for each participant taking part in the survey. The choices are presented in Listing 15.

```
1. <http://example.org/person/1> a foaf:Person .
2. <http://example.org/person/1> foaf:name "Ana" .
3. <http://example.org/city/100> a ex:City .
4. <http://example.org/person/1> foaf:based_near
   <http://example.org/city/100> .
5. <http://example.org/person/id> class foaf:Person .
6. <http://example.org/person/id> foaf:name "name" .
7. <http://example.org/city/city_id> class ex:City .
8. <http://example.org/person/id> foaf:based_near
   <http://example.org/city/city_id>.
9. foaf:Person a "1" .
10. foaf:Person based_near <#TriplesMapMapping1> .
11. ex:City a "100" .
12. foaf:name a "name" .
13. I cannot make sense of this mapping.
```

Listing 15. Multiple question choices for the understanding task

In order to select the correct triples participants would need to deduce what the mapping representation would generate, and thus demonstrate the extent of understanding of the presented mapping. The understanding task evaluated the following aspects:

- Instance class definitions (`foaf:Person` and `ex:City`).

- Mapping an attribute using a datatype property (`foaf:name`).

- Linking subjects using an object property (`foaf:based_near`).

5.4.3. Experiment 3: Evaluation of Juma in the understanding of uplift mappings

This section presents the third evaluation of the Juma approach, which is concerned with the understanding of uplift mappings. In this experiment, the understanding of mappings represented in Juma to mappings represented in R2RML (RDF TURTLE notation) were compared. Participants of this experiment were presented with an uplift mapping, either represented using the Juma Uplift representation or its R2RML equivalent. Participants were then asked to deduce the RDF triples that this mapping would generate when executed. As in experiment 2, the R2RML group and Juma group are used as labels to distinguish the participants by which each mapping representation was presented.

5.4.3.1. Hypotheses

There are two hypotheses of this experiment:

114

- **Hypothesis H1**: the performance of users in the task of understanding the presented uplift mapping represented in Juma is expected to be higher than the understanding of the same mapping represented in R2RML.
- **Hypothesis H2**: the perceived mental workload is expected to be lower for the task of understanding the presented uplift mapping represented in Juma than for the same mapping represented in R2RML.

For the purpose of this experiment, the understanding of an uplift mapping is defined as the ability of users to deduce the correct output triples from a specific mapping representation. The expected output is defined as the triples generated from the execution of the uplift mapping against its engine. In this case, since the mapping is either represented in R2RML or in an application that applied the Juma approach – which generates an R2RML mapping – the R2RML engine used in the previous experiment was used to generate the expected RDF output (gold standard).

5.4.3.2. Methodology

This user experiment is concerned with evaluating Juma in the context of understanding an uplift mapping. As mentioned before, this experiment was executed online, which means that anyone who received the link to the survey could participate. This also means that participants could pause the survey and come back at any point. For this reason, the link to the survey asked participants to provide a memorable keyword. This keyword was used to redirect participants to one of two questionnaires. The keyword was also elicited so that we could always redirect participants to the same questionnaire, in case participants stopped the experiment and came back at a later stage. The questionnaires within the survey had the same structure and questions, the only different being how the uplift mapping was represented. In one questionnaire the mapping was represented using R2RML in RDF TURTLE notation, in the other, the mapping was represented using the Juma Uplift application. There was no time limit for participants to execute the survey.

As described in Section 5.3.4.4, participants from universities and from the Semantic Web research community (via mailing lists, Twitter, LinkedIn, etc.) were approached. Juma was evaluated considering the participant's performance on the understanding of a certain uplift mapping representation, which is related to the hypothesis **H1** of this experiment. The mental workload was also assessed through the WP and NASA-TLX self-reporting instruments, which is related to the hypothesis **H2**. The group's performance and mental workload scores are also compared and correlated to each other using statistical tests. As stated previously,

statistical comparison tests show whether the differences between groups can be considered statistically significant; statistical correlation tests show the relationship between the aspects being evaluated (i.e. how these aspects may influence others).

5.4.3.3. Procedure

As mentioned before, this experiment was executed as an online survey. The survey was structured as follows:

- **Informed consent**: in this part, we explained the experiment to participants. At the end, participants had to consent to that information to proceed with the experiment. If they answered no, that was the end of the survey.

- **Technical debriefing**: all participants had the opportunity to watch videos about RDF and R2RML prior to executing the task. The group using Juma also had a presentation and a video about the Juma Uplift application. If they felt comfortable with these technologies, they could skip the material. The material was also available during the execution of the task.

- **Pre-task questionnaire**: participants were asked to fill out a pre-questionnaire about their familiarity to Semantic Web (SW) technologies. Participants evaluated their familiarity using a 7 Likert scale from 1 (strongly agree) to 7 (strongly disagree).

- **Understanding task**: in this part, we asked participants to work on the task presented in Section 5.4.2. Participants had access to the material during the task.

- **Post-task questionnaire**: after completion of the task, we asked participants to fill out another questionnaire on their perceived mental workload. The usability was not assessed, as participants did not interact with the mapping representation.

5.4.3.4. Participants

Participants of this experiment were recruited through email and mailing lists. E-mails were sent out to the ADAPT Centre research group, the School of Computer Science and Statistics in Trinity College Dublin, School of Computing in DIT, and to lecturers and professors of related areas asking them to forward the e-mail to their students. E-mails were also sent to the Linked Open Data (public-lod@w3.org) and Semantic Web (semantic-web@w3.org) mailing lists, posted as messages on Twitter and LinkedIn groups related to Semantic Web technologies. Participants were also asked to use their networks and forward the email to anyone who might be interested in taking part in the experiment.

In total, 95 participants entered the survey, but only 55 finished it. Out of these 55 participants, 25 answered the questionnaire with the R2RML representation, the other 30 answered the questionnaire with the Juma Uplift representation.

5.4.3.5. Results

This section is broken down into: task performance, mental workload, correlations, reliability, discussion and conclusions.

5.4.3.5.1. Task Performance

The task performance was calculated using precision (correctness), recall (completeness), and F-measure. Together with performance, the time spent (in minutes) by participants in the understanding task was also measure and presented, as well as the answers for the familiarity pre-questionnaire. As stated before, participants evaluated their familiarity with Semantic Web (SW) technologies and R2RML using a 7-Likert scale from 1 (strongly agree) to 7 (strongly disagree). Table 36 shows these results for the R2RML group. Table 37 shows the same results for the Juma group.

Table 36. Familiarity responses, time taken, and performance of the R2RML group

Participant	SW	R2RML	Time	Precision	Recall	F-measure
#1	7	7	3.00	0.00	0.00	0.00
#2	1	7	7.07	1.00	1.00	1.00
#3	3	7	2.68	0.00	0.00	0.00
#4	1	5	11.22	1.00	1.00	1.00
#5	2	6	44.50	0.25	0.25	0.25
#6	2	6	5.12	0.00	0.00	0.00
#7	1	3	5.09	0.67	1.00	0.80
#8	4	4	25.05	0.50	0.50	0.50
#9	7	7	27.49	1.00	0.25	0.40
#10	4	4	3.56	1.00	0.75	0.86
#11	7	7	112.49	0.40	0.50	0.44
#12	4	3	16.59	0.00	0.00	0.00
#13	7	7	25.72	0.00	0.00	0.00
#14	7	7	40.42	0.00	0.00	0.00
#15	4	4	2.39	0.00	0.00	0.00
#16	4	4	4.24	1.00	1.00	1.00
#17	2	2	33.50	0.00	0.00	0.00
#18	7	7	0.47	0.00	0.00	0.00
#19	1	7	5.05	1.00	1.00	1.00
#20	3	3	14.24	0.00	0.00	0.00
#21	4	7	2.16	0.00	0.00	0.00
#22	4	4	5.36	0.00	0.00	0.00

#23	4	4	5.38	0.00	0.00	0.00
#24	4	4	15.91	0.40	0.50	0.44
#25	4	4	11.08	0.50	0.50	0.50
		AVG	17.19	0.35	0.33	0.33
		STD	23.51	0.42	0.41	0.40

Table 37. Familiarity responses, time taken, and performance of the Juma group

Participant	SW	R2RML	Time	Precision	Recall	F-measure
#1	2	2	14.25	1.00	1.00	1.00
#2	7	7	9.98	0.50	0.50	0.50
#3	5	6	2.15	0.29	0.50	0.36
#4	5	7	63.07	0.00	0.00	0.00
#5	2	6	4.28	1.00	1.00	1.00
#6	3	3	4.56	0.60	0.75	0.67
#7	1	1	5.55	1.00	1.00	1.00
#8	6	6	11.73	0.25	0.50	0.33
#9	2	5	2.19	1.00	0.75	0.86
#10	4	4	4.58	0.25	0.25	0.25
#11	1	5	30.36	1.00	1.00	1.00
#12	1	6	14.08	1.00	1.00	1.00
#13	2	3	23.55	0.44	1.00	0.62
#14	1	5	27.54	1.00	1.00	1.00
#15	1	4	4.20	1.00	1.00	1.00
#16	1	6	6.97	1.00	1.00	1.00
#17	1	7	6.56	1.00	1.00	1.00
#18	2	6	10.52	0.67	0.50	0.57
#19	1	7	40.04	1.00	1.00	1.00
#20	4	7	5.29	1.00	1.00	1.00
#21	1	6	4.07	1.00	1.00	1.00
#22	1	7	1.98	1.00	1.00	1.00
#23	1	7	11.80	1.00	1.00	1.00
#24	2	2	1.65	0.00	0.00	0.00
#25	1	2	1.81	0.50	0.25	0.33
#26	2	2	4.99	1.00	1.00	1.00
#27	2	3	2.34	0.67	1.00	0.80
#28	1	3	3.17	0.60	0.75	0.67
#29	3	4	2.96	0.43	0.75	0.55
#30	4	4	17.17	1.00	1.00	1.00
		AVG	11.45	0.74	0.78	0.75
		STD	13.62	0.34	0.32	0.32

Examining these tables it can be seen that the performance (F-measure) achieved by the Juma group is higher than the one obtained by the R2RML group. Comparing the performance of the groups using the participants F-measure results in a p-value of 0.0001 for the independent two sample Welch T-Test and for the Wilcoxon test. These results suggest a statistically significant difference between the performance of the groups.

It should be noted that 8 participants from the R2RML did not attempt to answer the question task, that is these selected the option 13 which states *I cannot make sense of this*

mapping. These participants are #3, #6, #12, #13, #14, #17, #18, and #20. All participants in the Juma group attempted to answer the question.

Table 38 shows the average F-measure by grouping participants based on their answer to the familiarity pre-questionnaire. The number of participants in each group is also shown. Those participants with responses between 1 to 4 were considered as being familiar with the technology. Participants were considered to belong in one group only. For example, if a participant declared to be familiar with Semantic Web technologies and with R2RML then the participant was considered to belong to the R2RML group. If a participant declared to be familiar with Semantic Web technologies but not with R2RML then the participant was considered to be in the SW group. Participants who declared themselves not familiar with Semantic Web technologies or R2RML ended up in the non-experts group. Note that all participants familiar with R2RML declared themselves familiar with SW as well.

Table 38. F-measure averages per group

Groups	R2RML		Juma	
	Number of participants	Average F-measure	Number of participants	Average F-measure
SW	7	0.46	13	0.96
R2RML	12	0.34	13	0.66
Non-experts	6	0.14	4	0.30

The results show that the performance achieved by the Juma group is higher than the ones observed in the R2RML group, regardless of the participant's familiarity to Semantic Web technologies or to the R2RML mapping language.

Tailoring the performance per aspect evaluated by the task, which is described in Section 5.4.2, and only considering the correct triples selected, in the R2RML group there were: 26% of correct class mappings; 36% for attribute mapping; and 44% for linking; with overall performance of 33%. In the Juma group the performance was: 78.33% for class mappings; 86.67% for attribute mapping; and 70% for linking; with overall performance of 78.33%.

5.4.3.5.2. Mental Workload Assessment

The perceived mental workload was also evaluated in this experiment. **Table 39** shows the perceived mental workload for the R2RML group per participant. **Table 40** shows the same scores for the Juma group. These results show that the average mental workload is smaller for the Juma group, considering both the WP and NASA-TLX instruments.

Table 39. Perceived mental workload scores for the R2RML group

Participant	WP	NASA-TLX
#1	46.25	67
#2	35.00	38.2
#3	57.63	83.6
#4	40.88	51.8
#5	43.13	59
#6	56.25	66.2
#7	45.63	64
#8	70.38	40.8
#9	52.50	66
#10	52.00	56.8
#11	56.13	50.6
#12	63.63	87.4
#13	71.13	69.4
#14	65.88	82.8
#15	67.75	89.4
#16	58.88	73.4
#17	54.38	74.6
#18	42.88	39.6
#19	65.63	86.8
#20	64.50	79.6
#21	41.75	25.4
#22	73.13	80
#23	63.50	82.2
#24	61.75	68.6
#25	59.13	70.6
AVG	**56.39**	**66.15**
STD	**10.68**	**17.29**

Table 40. Perceived mental workload scores for the Juma group

Participant	WP	NASA-TLX
#1	34.00	57
#2	50.75	42.4
#3	43.63	45.6
#4	90.25	76.2
#5	30.13	45.8
#6	45.00	56.2
#7	46.38	45.8
#8	49.38	72.6
#9	15.00	32.6
#10	44.38	43
#11	45.75	54
#12	51.88	51
#13	49.38	55.4
#14	64.00	54.6
#15	46.38	55
#16	38.63	41
#17	55.25	57.8
#18	35.00	37.4
#19	44.25	41.6
#20	56.50	46.8
#21	31.38	31
#22	51.50	47.2
#23	51.38	46.6
#24	34.13	8.6
#25	49.38	43
#26	22.13	31.2
#27	26.13	41
#28	52.38	31.2
#29	58.25	67.4
#30	44.25	46.6
AVG	**45.23**	**46.85**
STD	**13.98**	**13.35**

The Shapiro-Wilk normality test was applied to both groups. **Table 41** shows the p-values for this test for the MWL scores, which suggests that the data follows a normal distribution.

Table 41. Shapiro-Wilk normality test p-values for the MWL scores per group

Shapiro-Wilk	R2RML	Juma
WP	0.37	0.06
NASA-TLX	0.14	0.20

Table 42 shows the p-values for the independent two sample Welch T-Test and Wilcoxon Test for the perceived mental workload between groups. As mentioned before, the Welch T-Test is considered sufficient since the data follows a normal distribution. For consistency, the Wilcoxon test was also applied, which does not assume the normal distribution of the data. Assuming a threshold of 0.05, these results suggest that the difference between the Juma group's and R2RML group's perceived mental workload scores is statistically significant.

Table 42. Mental workload p-values between the Juma and R2RML groups

MWL	Welch	Wilcoxon
WP	0.001	0.001
NASA-TLX	0.000003	0.0001

5.4.3.5.3. Correlations

Table 43 and **Table 44** show the correlation between F-measure and the familiarity questionnaire responses for the R2RML and Juma groups, respectively. These results indicate that as the F-measure goes up, so does the familiarity to Semantic Web technologies. Moreover, this correlation is statistically significant for both groups, assuming a threshold of 0.05. The correlations between F-measure and the participant's familiarity with R2RML were found not to be statistically significant.

Table 43. Correlation between F-measure and familiarity responses for the R2RML group

F-measure correlation	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
SW	-0.44	0.03	-0.39	0.05
R2RML	-0.04	0.86	-0.02	0.92

Table 44. Correlation between F-measure and familiarity responses for the Juma group

F-measure correlation	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
SW	-0.56	0.001	-0.59	0.0005
R2RML	0.12	0.52	0.16	0.40

Table 45 shows the correlation between F-measure and familiarity responses to the WP mental workload instrument for the R2RML group. **Table 46** shows the same correlations

for the NASA-TLX instrument. Assuming a threshold of 0.05, none of these correlations were found to be statistically significant.

Table 45. Correlation between F-measure and familiarity responses to WP for the R2RML group

WP correlations	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
F-measure	-0.31	0.13	-0.29	0.16
SW	0.24	0.26	0.26	0.21
R2RML	-0.30	0.15	-0.25	0.23

Table 46. Correlation between F-measure and familiarity responses to NASA-TLX for the Juma group

NASA-TLX correlations	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
F-measure	-0.26	0.22	-0.36	0.87
SW	-0.03	0.88	-0.01	0.94
R2RML	-0.33	0.11	-0.30	0.14

Table 47 shows the correlation between F-measure and familiarity responses to the WP mental workload instrument for the Juma group. **Table 48** shows the same correlations for the NASA-TLX instrument. These correlations were also found not to be statistically significant when considering a threshold of 0.05.

Table 47. Correlation between F-measure and familiarity responses to WP for the Juma group

WP correlations	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
F-measure	-0.29	0.12	-0.08	0.67
SW	0.24	0.20	-0.06	0.75
R2RML	0.33	0.08	0.35	0.06

Table 48. Correlation between F-measure and familiarity responses to NASA-TLX for the Juma group

NASA-TLX correlations	Pearson		Spearman	
	Coefficient	p-value	Coefficient	p-value
F-measure	-0.02	0.90	0.03	0.87
SW	0.30	0.11	0.17	0.36
R2RML	0.25	0.18	0.17	0.38

5.4.3.5.4. Reliability

Table 49 shows the Cronbach's alpha index for the WP and NASA-TLX questionnaires applied in this experiment, suggesting that the results are reliable.

Table 49. Cronbach's alpha index for WP and NASA-TLX

MWL	Alpha index
WP	0.72
NASA-TLX	0.78

5.4.3.6. Discussion

The performance of participants in the Juma group were higher, as is shown by the F-measures presented in **Table 36** (R2RML group) and **Table 37** (Juma group). Furthermore, this difference was found to be statistically significant with a confidence level of 0.05. As it was stated in the results section, 8 participants in the R2RML group did not attempt to answer the main question of the experiment, which asked participants to select the triples that the presented mapping would generate when executed. Note that all participants presented with the Juma approach did attempt to answer the same question. This suggests to the author of this thesis that the Juma representation is more engaging than the R2RML representation.

The time taken by participants to execute the task has a lot of variation in both groups. Participants could stop the survey and come back at any point, which may explain the high values. Small values might show that participants gave up easily, or that they closed the survey window and came back at a later stage, which resets the timer. In the R2RML group, 4 participants spent 3 minutes or less to execute the task, 2 of which would be considered familiar with Semantic Web technologies, all with F-measure of 0.00 (**Table 36**). In the Juma group, 7 participants spent 3 minutes or less in the execution of the task, 2 of which would be considered familiar with Semantic Web technologies, and 4 would be familiar with R2RML (**Table 37**). Of these, 1 participant familiar with R2RML had 0.00 F-measure.

In the R2RML group, 2 participants achieved maximum F-measure, and these considered themselves familiar with Semantic Web technologies. In the Juma group, only 2 participants got 0.00 F-measure, and 1 of these declared himself to be familiar with R2RML. Other participants had higher scores with similar familiarity responses. Nevertheless, the number of participants with 0.00 F-measure scores is much higher in the R2RML group.

Table 38 shows that participants that declared themselves familiar with Semantic Web technologies achieved better results than the ones who considered themselves familiar with R2RML, in both groups. It is important to note that all participants who declared themselves to be familiar with R2RML have also declared themselves to be familiar with Semantic Web technologies. Nonetheless, participants in the Juma group had at least double the F-measure when compared to the R2RML group, in any of the familiarity groups defined. Juma group participants also had higher performance when considering the type of the uplift mapping definition being evaluated (class, attribute, and linking).

The correlation between F-measure and their familiarity with Semantic Web technologies was found to be statistically significant in both groups (**Table 43** and **Table 44**). This suggests that as F-measure goes up, their familiarity with Semantic Web technologies increases, which is expected. However, the correlations between F-measure and the familiarity with R2RML was found not to be statistically significant.

The mental workload was smaller for the Juma group for both the WP and NASA-TLX instruments (**Table 39** and **Table 40**). Moreover, the difference between the group's mental workload were found to be statistically significant, as it is shown in **Table 42**. The correlations, however, between F-measure, familiarity responses, and mental workload scores were not statistically significant (**Table 45**, **Table 46**, **Table 47** and **Table 48**).

Cronbach's alpha indexes highlight a strong internal consistency of the mental workload questionnaires used in this experiment (**Table 49**).

5.4.3.7. Conclusion

This section presented the third experiment of this thesis, which evaluated the Juma approach with respect to a task that required understanding of uplift mapping representations. As stated before, the understanding of an uplift mapping in this experiment has been defined as the ability of users to deduce the correct output triples from a specific mapping representation.

The **results** presented in this section **confirm the hypothesis H1** of this experiment, which stated that the **performance** of participants **is expected to be higher for mappings represented using the Juma approach**. Moreover, the difference between the group's performances was found to be statistically significant. The performance of the participants in the Juma group was also higher when considering the participant's familiarity to Semantic Web technologies and R2RML; and when considering the type of the uplift mapping definition. **The mental workload scores in the Juma group were also smaller**. Moreover, the difference between the groups was also found to be statistically significant, **confirming the hypothesis H2**, which stated that the mental workload perceived by participants understanding a mapping represented using the Juma approach is lower than the same mapping represented in R2RML.

Finally, Cronbach's alpha suggest that the questionnaires used in this experiment are reliable.

5.5. Findings for the expressiveness of Juma

This section presents two experiments developed to evaluate the expressiveness of the Juma approach in the representation of mappings in Linked Data. Section 5.5.1 presents the introduction. Sections 5.5.2 and 5.5.3 present experiments 4 and 5, respectively.

5.5.1. Introduction

This section presents two experiments conducted to evaluate the expressiveness of Juma in the visual representation of uplift and semantic mappings.

Experiment 4 focuses on the expressiveness of uplift mappings. In this experiment, it is shown that Juma visualization is capable of representing mappings using the R2RML and SML uplift mapping languages. Experiment 5 focuses on evaluating whether Juma visualization is capable of representing semantic mappings.

5.5.2. Experiment 4: Evaluation of the expressiveness of Juma in the representation of uplift mappings

This section presents the fourth evaluation of the Juma approach and is concerned with the expressiveness of Juma in the representation of uplift mappings. Juma R2RML reflects the R2RML vocabulary and is fully compliant with the R2RML specification, as was stated in Section 4.3.1. For this reason, the Juma Uplift application was instead used (by the author of this thesis) to evaluate the expressiveness of Juma in the visual representation of uplift mappings. Juma Uplift has a higher level of abstraction and is designed to be capable of generating mappings using both the R2RML and SML mappings languages.

5.5.2.1. Hypothesis

The hypothesis of this experiment is:
- **Hypothesis H1**: Juma is able to express uplift mappings that generate accurate mappings in the R2RML and SML mapping languages for common uplift use cases.

5.5.2.2. Methodology

This experiment is concerned with evaluating the expressiveness of Juma in the visual representation of uplift mappings. The Juma Uplift application, which can generate

mappings using the R2RML and SML mapping languages, was used in this experiment. For this experiment, 10 uplift mapping use cases were developed based on the work presented in (J. Sequeda et al., 2012). These use cases are presented in Section 5.5.2.4. Juma Uplift is then used to generate a mapping for each use case. To test the hypothesis **H1** of this experiment, the mappings generated by Juma are executed against an R2RML engine[52] and an SML engine[53]. The outputs are then compared to an expected gold standard output.

5.5.2.3. Procedure

The procedure of this experiment (undertaken by the author in the lab) is as follows:

- Common uplift mapping use cases were created based on the uplift mapping patterns presented in (J. Sequeda et al., 2012). For this experiment, some of these mapping patterns were combined into more general use cases. In order to explore the expressiveness of the Juma Uplift application two use cases were also added: one for the mapping of attributes with a language tag, and one for the use of a transformation function. In this step, an expected gold standard RDF output was created for each use case. The use cases and gold standard created for this experiment was also verified by 2 Semantic Web experts.
- For each mapping use case, Juma Uplift was then used to create an uplift mapping.
- The R2RML and SML mappings generated by Juma Uplift are then executed against its respective engines.
- In the final part of this experiment, the RDF datasets generated are compared against the expected gold standard RDF output using Jena[54] models.

5.5.2.4. Use cases

Table 50 and **Table 51** show the relational database created for this experiment.

Table 50. Table *person* of the database used in the experiment for the expressiveness of Juma

ID	NAME	AGE	CITY_FK
1	Ana	29	100
2	John	25	100
3	Mary	30	200

---

[52] The engine is available at `https://opengogs.adaptcentre.ie/debruync/r2rml`, accessed in June 2018.

[53] The engine is available at `https://github.com/AKSW/Sparqlify`, accessed in June 2018.

[54] Jena is a Java framework for building Semantic Web applications available at `https://jena.apache.org/`, accessed in June 2018.

Table 51. Table *city* of the database used in the experiment for the expressiveness of Juma

CITY_ID	NAME
100	Dublin
200	London

As stated before, 10 uplift mapping use cases were defined (**Table 52**). These use cases were defined to cover common cases when uplifting data to RDF and also to explore the Juma visual representation.

Table 52. Mapping use cases used in the experiment for the expressiveness of Juma

Mapping	Description
#1	Mapping one table to one class
#2	Mapping one table to two classes
#3	Mapping two tables to one class each
#4	Mapping one table and one attribute
#5	Mapping one table and two attributes
#6	Mapping one table and one attribute with a language tag
#7	Mapping one table and one attribute as a resource
#8	Mapping two tables and one attribute each
#9	Mapping one table and one attribute as the result of a data transformation function
#10	Mapping two tables with one class and one attribute each and a link between them

For example, the mapping use case #10 represented using the Juma Uplift application can be seen in **Fig. 49**. Listing 16 and Listing 17 show the corresponding R2RML and SML mappings generated by Juma Uplift. In this example, the table *person* is mapped to the class foaf:Person and the column name to the predicate foaf:name, The example also maps the table *city* to be instances of the class ex:City with the column name to the predicate foaf:name, Finally, a link between the subjects using the predicate foaf:based_near is defined.
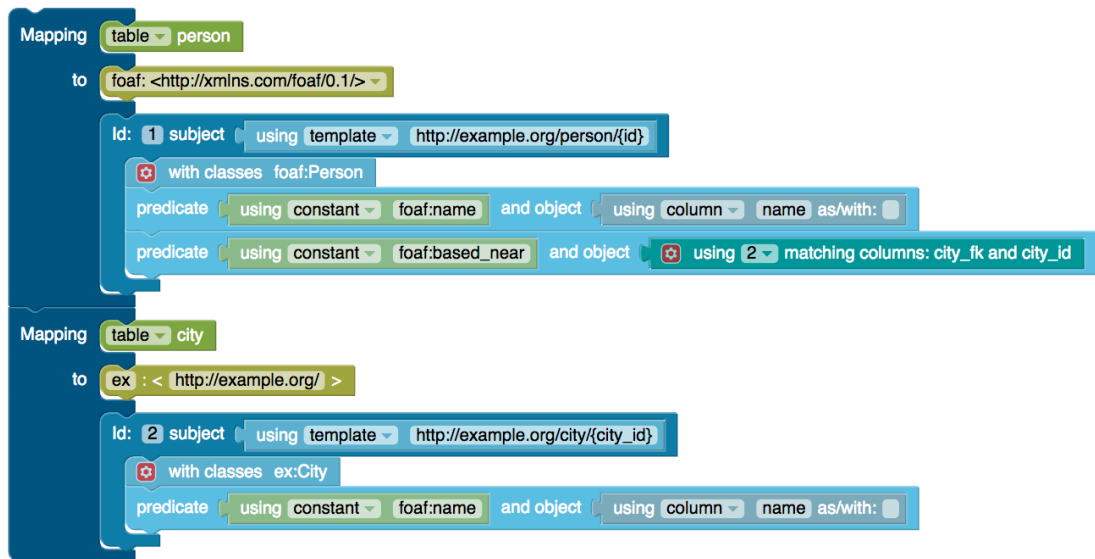
Fig. 49. Mapping use case #10 in Juma Uplift

```
@prefix rr: <http://www.w3.org/ns/r2rml#> .
@prefix rrf: <http://kdeg.scss.tcd.ie/ns/rrf#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ex: <http://example.org/> .


<#TriplesMap1>
  rr:logicalTable [ rr:tableName "person";];
  rr:subjectMap   [
    rr:template "http://example.org/person/{id}"; rr:class
    foaf:Person;
  ];
  rr:predicateObjectMap [
    rr:predicateMap [rr:constant foaf:name;];
     rr:objectMap [rr:column "name";];
  ];
  rr:predicateObjectMap [
    rr:predicateMap [rr:constant foaf:based_near;];
    rr:objectMap [
     rr:parentTriplesMap <#TriplesMap2>;
     rr:joinCondition [rr:child "city_id";rr:parent "city_id";];];
  ];.
<#TriplesMap2>
  rr:logicalTable [ rr:tableName "city"; ];
  rr:subjectMap   [
    rr:template "http://example.org/city/{city_id}";
    rr:class ex:City;
  ];
  rr:predicateObjectMap [
    rr:predicateMap [ rr:constant foaf:name;];
    rr:objectMap [rr:column "name";];
  ];.
```

Listing 16. Mapping use case #10 in R2RML

129

```
   Prefix foaf: <http://xmlns.com/foaf/0.1/>
   Prefix ex: <http://example.org/>

   Create View view2 As
    Construct {
       ?s1  a foaf:Person.
       ?s1   foaf:name   ?o1. }
    With
     ?s1 = uri(concat('http://example.org/person/', ?id))
     ?o1 = plainLiteral(?name)
    From person

   Create View view1 As
    Construct { ?s1 foaf:based_near ?o1 . }
    With
   ?s1 = uri(concat('http://example.org/person/', ?id))
   ?o1 = uri(concat('http://example.org/city/', ?city_id))
   From
    [[
     SELECT * FROM city AS child, person AS parent WHERE
    child.city_id = parent.city_fk
     ]]

   Create View view3 As
    Construct {
       ?s1  a ex:City.
       ?s1   foaf:name   ?o1. }
    With
     ?s1 = uri(concat('http://example.org/city/', ?city_id))
     ?o1 = plainLiteral(?name)
    From
     city
```

Listing 17. Mapping use case #10 in SML

5.5.2.5. Results

It was found that Juma Uplift visual representation of the mappings were capable of creating the R2RML and SML representations that generate the expected RDF output. The full experiment data is available[55].

5.5.2.6. Discussion

Through an experiment, common use cases have been defined for uplifting data from relational databases to RDF. For every use case a visual representation of the mapping was created using Juma Uplift. The visual representation automatically generates R2RML and SML mappings. These mappings were then executed using their respective engines and compared to an expected output. The results showed that Juma was able to create mappings that generated the expected output.

---

[55] http://www.tara.tcd.ie/handle/2262/85064, accessed in October 2018.

5.5.2.7. Conclusion

This section presented the fourth experiment of this thesis, which evaluated the expressiveness of Juma in the representation of uplift mappings.

The results presented in this section **confirm the hypothesis H1** of this experiment, which states that **Juma is expressive enough to generate accurate R2RML and SML mappings** for the designed use cases.

5.5.3. Experiment 5: Evaluation of the expressiveness of Juma in the representation of semantic mappings

This section presents the fifth evaluation of the Juma approach. This experiment evaluates if Juma can be used to represent semantic mappings. This experiment is also concerned with the expressiveness of Juma approach in the representation of such semantic mappings

5.5.3.1. Hypothesis

The hypothesis of this experiment is:

- **Hypothesis H1**: Juma is capable of expressing semantic mappings that generate accurate mappings as SPARQL CONSTRUCT queries from a real use case scenario.

5.5.3.2. Methodology

This experiment is concerned with evaluating the expressiveness of Juma in the representation of semantic mappings. The Juma Interlink application was used in this experiment. The dataset used in this experiment is described in Section 5.5.3.4). To test the hypothesis **H1** of this experiment, Juma Interlink was used (by the author of this thesis) to generate the mappings as SPARQL CONSTRUCT queries, and compared with the independently published mappings.

5.5.3.3. Procedure

The procedure of this experiment is as follows:

- Juma Interlink was used to create each of the mappings in the dataset.

- The mappings generated, as SPARQL CONSTRUCT queries, were manually compared to the mappings from the dataset.

5.5.3.4. Dataset

The dataset used in this experiment was devised by the authors of the R2R framework and were published in (Bizer & Schultz, 2010), using the R2R mapping language. Such mappings have been used in other experiments (Meehan, Brennan, Lewis, & O'Sullivan, 2014b), where the authors have converted, validated and published them as SPARQL CONSTRUCT queries. Juma Interlink generates mappings as SPARQL CONTRUCT queries, thus these are the mappings used in this experiment. The dataset consists of 72 mappings between DBpedia and 11 other datasets. There are 52 simple mappings and 20 complex mappings. These mappings contain class mappings, property mappings, value transformations, amongst others. A description of the dataset is available in (Bizer & Schultz, 2010). The mappings as SPARQL CONSTRUCT queries are available[56].

An example of a simple mapping from this dataset was shown in Fig. 38 (Chapter 4). **Fig. 50** shows an example of a complex mapping in Juma Interlink. In this example, the property `dbpedia:runtime` is mapped to `movie:runtime` through a transformation function. This transformation function converts seconds to minutes. The mapping generated by Juma Interlink as a SPARQL CONSTRUCT query is shown in **Listing 18**.
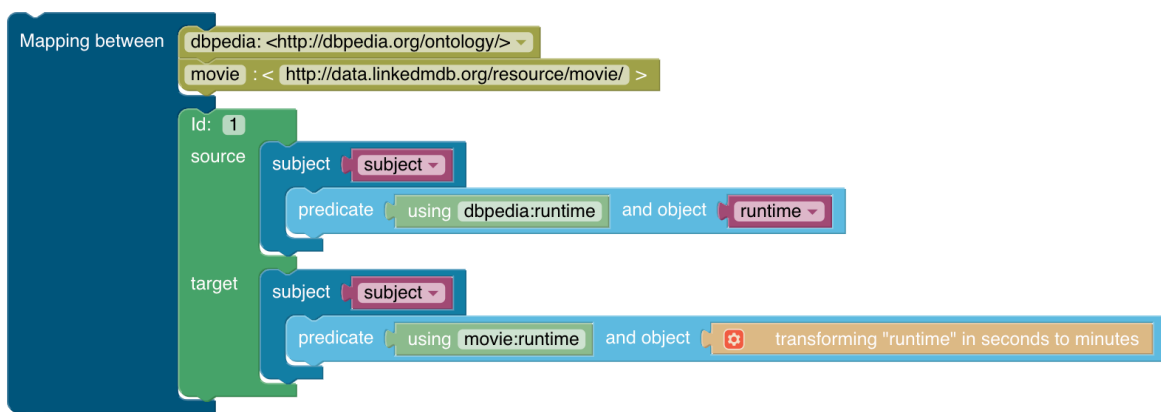

Fig. 50. Mapping use case in Juma Interlink

```
prefix dbpedia: <http://dbpedia.org/ontology/>
```

---

[56] http://www.tara.tcd.ie/handle/2262/85066, accessed in October 2018.

```
prefix movie: <http://data.linkedmdb.org/resource/movie/>

#Id1
CONSTRUCT {
      ?subject  movie:runtime   ?result1 .
} WHERE {
      ?subject  dbpedia:runtime   ?runtime .
      BIND ((?runtime/60) AS ?result1)
}
```

Listing 18. Mapping use case as a SPARQL CONSTRUCT query

5.5.3.5. Results

It was found that Juma Interlink was able to represent the entire set of mappings contained in the dataset.

5.5.3.6. Discussion

This experiment evaluated the extent to which Juma could represent semantic mappings. To do this, Juma Interlink application was used to create a visual mapping for each of the use case mappings presented in the dataset described in Section 5.5.3.4. It was found that Juma was able to represent all set of mappings.

5.5.3.7. Conclusion

This section presented the fifth experiment of this thesis, which evaluated Juma's expressiveness in the representation of semantic mappings. The results presented in this section **confirm the hypothesis H1** of this experiment, which states that **Juma is able of expressing semantic mappings** for a real use case scenario.

5.6.   Overall Conclusions

This section presents overall conclusions from the experiments undertaken to evaluate the Juma approach through use of the three Juma applications that have been developed.

The performance and usability results suggest that the process of creating, editing and understanding Linked Data mappings is facilitated through the Juma approach for different types of users. Even though participants using Juma achieved high performance the most difficult aspect of the uplift mapping task seems to be the linking entities. In the comparison between R2RML and Juma, it was seen that Juma was able to aid users in the creation and

understanding of those links, however, it was still the characteristic where users achieved less performance.

To the authors knowledge, this thesis presents the first experiments evaluating the cognitive load of creating and understanding uplift mappings. As stated in 5.2.2, mental workload instruments quantify the cognitive load of performing a task. The mental workload assessment indicates that creating and understanding mappings in R2RML is more demanding when compared to the Juma approach. It is noted that, for the task of creating uplift mappings, the difference between using Juma and R2RML was found not to be statistically significant – maybe due to the number of participants in this experiment. Nonetheless, for the task of understanding mappings this difference between the group's mental workload scores were found to be statistically significant.

Overall, the evaluations presented in this chapter indicate that the Juma approach is effective in representing uplift and semantic mappings. Even though the experiments involving users presented in this thesis focus on uplift mappings, the application Juma Uplift and Juma Interlink apply the same approach in the representation of Linked Data mappings, and have a uniform representation for common mapping constructs. These characteristics lead us to believe that Juma can aid users in creating, editing and understanding mappings in Linked Data, regardless of its type.

5.7. Chapter Summary

In summary, this chapter has presented five evaluations of the Juma approach. The evaluations were grouped based on three different aspects: creation, understanding and expressiveness. Experiments 1 and 2 evaluated Juma in the creation of uplift mappings. Experiment 3 evaluated the understanding of mappings, while experiment 4 and 5 evaluated the expressiveness of our approach.

The first experiment evaluated users with different background knowledge creating one R2RML mapping. As mentioned before, the participants of this experiment were selected to be considered by the authors as more likely to create and publish Linked Data datasets. The results have shown that, regardless of their expertise, participants were able to use Juma to create mappings with high performance and sufficient usability.

The objective of the second experiment was again to evaluate Juma for the creation of uplift mappings. In contrast with the first experiment, this evaluation compares two different approaches in the creation of uplift mappings, one using Juma applied to uplift mappings and the use of RDF TURTLE to write R2RML mappings, to test whether Juma performs

better. In this experiment, we also evaluated the cognitive load of creating uplift mappings using the two aforementioned approaches. The results have shown that participants Juma using achieved higher performance, better usability and slightly less mental workload.

The third evaluation had the objective of testing the understanding of uplift mappings represented in Juma and R2RML. This experiment was executed as an online survey, and participants were evaluated on their understanding of these uplift mapping representations, and the mental workload imposed by this task. The results have shown that participants were more likely to interact with Juma, with higher performance and smaller mental workload.

The fourth experiment evaluated the expressiveness of Juma in the representation of uplift mappings. In this experiment, common uplift mapping use cases were defined. Juma was then used to create a mapping for each of these use cases. The generated R2RML and SML mappings were then executed and compared to an expected output. Results have shown that Juma generated accurate mappings for all use cases.

Finally, the fifth experiment evaluated the expressiveness of Juma in the representation of semantic mappings. For this experiment, as mentioned before, we have used Juma to generate the 72 semantic mappings created independently of the research of this thesis. Results have shown that Juma was able to represent all sets of mappings.

6. Conclusion

This chapter draws conclusions from the research presented in this thesis. Section 6.1 discusses to what extent the research objectives of this thesis (as set out in Chapter 1) have been achieved. In Section 6.2 the contributions of the research are revisited. Section 6.3 presents topics for possible further work based on the research presented in this thesis. Final remarks are presented in Section 6.4.

6.1. Research Objectives

The extent to which each research objective posed to address the research question of this thesis are analysed in this section.

6.1.1. RO1: Perform a state-of-the-art review of existing visual representations for mappings in Linked Data

The first research objective was to *perform a state-of-the-art review of existing visual representations for uplift and semantic mappings*.

This research objective was achieved through the review of the state of the art described in Chapter 3. This review examined approaches that apply visual representations for mappings that generate or interlink Linked Data, which are respectively named uplift and semantic mappings. These approaches were classified based on their primary visual representation as graph-based and tree-based visual representations. The key characteristics examined are related to the research question of this thesis, being the visual technique applied by these approaches, the mapping type and mapping languages supported, how users can interact with it in order to create and edit mappings, and if user experiments were conducted to evaluate them. The state of the art analysis has shown that existing approaches support one type of mapping (uplift or semantic), and one mapping language only. This analysis has also shown that some approaches offer editing approaches that are separate from the visual representation of mappings, which we believe not to be intuitive. Finally, we have observed that user evaluations are often neglected.

The state of the art review was used to identify requirements for a visual representation for mappings in Linked Data, which were presented in Section 4.1. In the same section, the support for requirements provided by the state of the art approaches were also presented.

6.1.2. RO2: Propose a visual representation for mappings in Linked Data

The second research objective was to *propose a visual representation applicable to both uplift and semantic mappings*.

This research objective was achieved by proposing the **J**igsaw **Pu**zzles for Representing **Ma**ppings (Juma) approach, which was described in Chapter 4. Juma adopts the block metaphor in order to facilitate the creation, editing and understanding of mappings that generate and interlink Linked Data datasets. The block metaphor was chosen as it takes advantage of the user's familiarity with jigsaw puzzles, fosters users to explore the combinations of blocks, and is accessible to experts and non-experts alike. Juma supports the requirements derived from the state of the art analysis and the research question by supporting the representation of both uplift and semantic mappings, offering an intuitive editing approach through elements represented using blocks, guiding users in the creation and editing of (at least syntactically) valid mappings, and by generating mappings independently of its visual representation, which allows for the serialization of mappings using multiple distinct mapping languages.

6.1.3. RO3: Apply, implement and evaluate the proposed visual representation to uplift mappings

The third research objective was to *apply, implement and evaluate the visual representation defined in RO2 to uplift mappings*.

This research objective was achieved by applying the Juma approach to two applications that can represent uplift mappings, being Juma R2RML (Section 4.3.1) and Juma Uplift (Section 4.3.2). Juma R2RML applies the Juma approach to the R2RML mapping language. Each block in Juma R2RML has been designed to represent an R2RML statement that automatically generates a correspondent R2RML construct. Juma R2RML was evaluated in an experiment 1, which was presented in Section 5.3.3. There were three groups of users in this experiment: web developers, knowledge engineers, and users familiar with the R2RML mapping language. Our intuition is that these users are more likely to create and publish Linked Data datasets. The results have shown that, regardless of their expertise, participants were able to use Juma to create mappings with high performance and sufficient usability.

Juma Uplift has been developed with a higher level of abstraction in order to have the capability to generate mappings using distinct mapping languages, as per user choice. The development of Juma Uplift involved identifying and abstracting uplift mapping constructs that would be used to generate mappings in distinct mapping languages. Juma Uplift was evaluated in experiment 2 (Section 5.3.4), experiment 3 (Section 5.4.3), and experiment 4

(Section 5.5.2). Experiment 2 evaluated the Juma approach in the creation of uplift mappings by comparing the use of Juma Uplift to creating mappings "manually" i.e. through a text editor. For this experiment, participants were split into two groups to work on the same mapping task. The results have shown that participants Juma using achieved higher performance, better usability and slightly less mental workload. Experiment 3 evaluated the Juma approach for the understanding of uplift mappings. This experiment was executed as an online survey where the same mapping was shown to participants using two different mapping representations, R2RML in RDF TURTLE syntax and Juma Uplift. Participants were then asked to answer the same multiple-choice question. In this question, participants should select all triples that the mapping representation would generate when executed. The results have shown that participants using Juma achieved higher performance and smaller mental workload. Experiment 4 evaluated the expressiveness of the Juma approach in the representation of uplift mappings. Ten uplift mapping use cases were derived from the research presented in (J. Sequeda et al., 2012). Juma Uplift was then used to create a mapping for each of the use cases. The mappings generated by Juma Uplift were then executed and compared to expected RDF outputs. Results have shown that Juma generated accurate mappings for all use cases.

6.1.4. RO4: Apply, implement and evaluate the proposed visual representation to semantic mappings

The fourth research objective was to *apply, implement and evaluate the visual representation defined in RO2 for semantic mappings*.

This research objective was achieved by applying the Juma approach to the representation of semantic mappings, in an application called Juma Interlink. Juma Interlink (Section 4.3.3) was developed to support semantic mappings that automatically generate executable mappings in the form of SPARQL CONSTRUCT queries. The development of Juma Interlink also involved identifying general principles found between semantic and uplift mappings, that were used to define a uniform representation for equivalent constructs between these types of mappings. Juma Interlink was evaluated in experiment 5 (Section 5.5.3). This experiment evaluated the expressiveness of the Juma approach in the representation of semantic mappings. For this experiment, the Juma Interlink application was used to represent 72 semantic mappings. The mappings used in this experiment were devised by the R2R Framework research (Bizer & Schultz, 2010). Results have shown that Juma was able to represent all sets of mappings.

6.2. Contributions

This section briefly revisits the contributions from the research of this thesis, which were initially presented in Section 1. This section also presents the impact of the research of this thesis.

The major contribution of this thesis is the Jigsaw Puzzles for Representing Mappings (Juma) approach. Juma applies the block metaphor in order to facilitate the creation, editing and understanding of mappings in Linked Data. The Juma approach advances the state of the art by offering a visual representation for Linked Data mappings that can be applied to both uplift and semantic mappings, and that is focused on the different types of users (experts and non-experts) involved in the mapping process. The user experiments conducted to evaluate Juma considered three aspects. The performance of participants in creating and understanding mappings. Usability, through a standard instrument that explicitly assess interface quality, which, as mentioned, is of particular importance as this thesis proposes a visual representation. Finally, to the authors knowledge, this thesis advances the state of the art by presenting the first experiments evaluating the perceived mental workload of creating and understanding uplift mappings.

As well as advancing the state of the art, it is envisaged that the Juma approach would have an impact on the adoption of Linked Data by widening the types of users that could get involve in the creation, editing and understanding of mappings that can generate and interlink Linked Data datasets.

The minor contributions yielded from the research of this thesis are three applications that apply the Juma approach in the representation of Linked Data mappings. The **first minor contribution is the Juma R2RML** application. Juma R2RML applies the Juma approach to the R2RML mapping language.

The **second minor contribution is the Juma Uplift** application. Juma Uplift has a higher level of abstraction in order to be able to generate mappings that are not only compliant with R2RML, but also with other mapping languages (in this case, the SML mapping language).

The **third minor contribution is the Juma Interlink** application. Juma Interlink is able of representing simple and complex semantic mappings that are automatically translated to SPARQL CONSTRUCT queries.

The Juma Uplift and Juma Interlink applications also offer a uniform representation for equivalent mapping constructs that can be found in both uplift and semantic mappings,

which is argued by the author to facilitate the interpretation of both types of mappings found in the Linked Data domain.

6.2.1. Uptake

As stated in Section 1, this research is already having impact within the research community with publications in well-known venues such as the Extended Semantic Web Conference 2018, IEEE International Conference on Semantic Computing 2018, amongst others.

The Juma Uplift application is also currently being used in the Linked Open Statistics (LOS) project[57]. The LOS project itself is part of the wider European Statistical System's (ESS) DIGICOM project[58]. The objective of the DIGICOM project is to enable further advancements in two key areas of the ESS vision: (1) focus on users and (2) improve dissemination and communication. The LOS project fits into the DIGICOM *Open Data Dissemination* work package where it aims to produce tools, standards and guidelines for the production of an open Linked Data representation of statistical data – with the goal of allowing easier publication and analysis of ESS wide statistical data. There are four National Statistical Organisations (NSO) partners taking part in the LOS project:

- The *National Statistical Institute* of Bulgaria.
- The *Institut National de la Statistique et des Etudes Economiques* (INSEE) of France.
- The *Istituto Nazionale Di Statistica* (ISTAT) of Italy.
- The *Central Statistics Office* (CSO) of Ireland.

Juma is currently deployed in a Statistical Linked Data publication and analysis pipeline. It is utilised for the creation of uplift mappings for the purpose of the conversion of statistical data, from the four NSO partners, to RDF Linked Data. This Statistical Linked Data is then made available for users to download, visualise and analyse through other specific tools made available in the pipeline.

Finally, Juma applications were also requested to be used in third level modules related to Linked Data and Semantic Web technologies to teach how mappings can be used to generate and integrate Linked Data datasets

---

[57] `https://ec.europa.eu/eurostat/cros/content/linked-open-statistics_en`, accessed in September 2018.

[58] `https://ec.europa.eu/eurostat/web/ess/digicom`, accessed in September 2018.

6.3. Further Work

This section discusses potential further work that could be undertaken based on the findings of this thesis.

In this thesis we have shown that a visual representation can be expressive enough to support different mapping languages. Nonetheless, other use cases may need the support of other mapping languages, which may be added in future versions of the provided implementations.

Another possible further work is related to assessing the quality of mappings and the generated RDF data. Although beyond the scope of the research of this thesis, quality assessment can be used to validate and refine mappings as well as the generated data. Similar work has been conducted in (Dimou et al., 2015), which can be incorporated or used as inspiration for a data quality assessment process applied to the Juma approach.

As stated in the state of the art chapter, studies have been conducted to compare graph-based and tree-based visual representations. Nonetheless, another possible route of further work would be to conduct user experiments to compare different visual representations applied to the task of creating and editing Linked Data mappings.

6.4. Final Remarks

It is hoped by the author of this thesis that the Juma approach, which provides a visual representation for mappings in Linked Data, can be of benefit to users who already generate, publish, and interlink Linked Data datasets, by integrating Juma into their mapping process; and to users who want to begin the process of mapping their data using Linked Data principles.

It is also hoped that the Juma approach would benefit the research community. Researchers can employ Juma in their mapping processes, use the findings presented in this thesis in their research, and apply their expertise to contribute to the approach and its implementations.

References

Alexiev, V. (2016). RDF by Example: rdfpuml for True RDF Diagrams, rdf2rml for R2RML Generation. In *Semantic Web in Libraries 2016 (SWIB 16)*. Bonn, Germany.

Arenas, M., Bertails, A., Prud'hommeaux, E., & Sequeda, J. (2012). A direct mapping of relational data to RDF. *W3C Recommendation*.

Aryan, P. R., Ekaputra, F. J., Kiesling, E., & Tjoa, A. M. (2017). RMLx : Mapping Interface for Integrating Open Data with Linked Data Exploration Environment. In *1st International Conference on Informatics and Computational Sciences (ICICoS)* (pp. 113–118).

Aumueller, D., Do, H.-H., Massmann, S., & Rahm, E. (2005). Schema and ontology matching with COMA++. In *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. ACM. http://doi.org/10.1145/1066157.1066283

Bąk, J., Blinkiewicz, M., & Ławrynowicz, A. (2017). User-friendly visual creation of R2RML mappings in SQuaRE. In *Proceedings of the Third International Workshop on Visualization and Interaction for Ontologies and Linked Data co-located with the 16th International Semantic Web Conference ISWC 2017, Vienna, Austria, October 22, 2017.* (Vol. 1947, pp. 139–150).

Barrasa, J., Corcho, Ó., & Gómez-pérez, A. (2004). R2O, an Extensible and Semantically based Database-to-Ontology Mapping Language. In *In Proceedings of the 2nd Workshop on Semantic Web and Databases (SWDB2004)* (pp. 1069–1070). Springer.

Bart, A. C., Tibau, J., Kafura, D., A. Shaffer, C., & Tilevich, E. (2017). Design and Evaluation of a Block-based Environment with a Data Science Context. *IEEE Transactions on Emerging Topics in Computing*, *6750*(c), 1–12. http://doi.org/10.1109/TETC.2017.2729585

Beckett, D., Berners-Lee, T., Prud'hommeaux, E., & Carothers, G. (2014). RDF 1.1 Turtle - Terse Rdf Triple Language. W3C Recommendation. *World Wide Web Consortium (Feb 2014), Available at Http://Www. W3. Org/TR/Turtle*.

Bederson, B., & Shneiderman, B. (2003). *The craft of information visualization: readings and reflections*. Morgan Kaufmann.

Bizer, C., & Cyganiak, R. (2006). D2R Server - Publishing Relational Databases on the Semantic Web. In *5th International Semantic Web Conference* (p. 26).

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, *5*(3), 1–22. http://doi.org/10.4018/jswis.2009081901

Bizer, C., & Schultz, A. (2010). The R2R framework: Publishing and discovering mappings on the web. *Proceedings of the First International Workshop on Consuming Linked Data, Shanghai, China, November 8, 2010*. http://doi.org/urn:nbn:de:0074-665-8

Bizer, C., & Seaborne, A. (2004). D2RQ – Treating Non-RDF Databases as Virtual RDF Graphs. In *Proceedings of the 3rd International Semantic Web Conference (ISWC2004)* (p. 26). http://doi.org/10.1.1.126.2314

Blinkiewicz, M., & Bak, J. (2016). SQuaRE: A Visual Support for OBDA Approach. In *Proceedings of the Second International Workshop on Visualization and Interaction for Ontologies and Linked Data co-located with the 15th International Semantic Web Conference, Kobe, Japan, October 17, 2016.* (pp. 41–53).

Brickley, D., & Guha, R. V. (2014, February). RDF Schema 1.1. *World Wide Web Consortium*. http://doi.org/10.1016/B978-0-12-373556-0.00006-X

Brooke, J. (1996). SUS: A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194).

Cain, B. (2007). *A Review of the Mental Workload Literature. Defence research and development Toronto (Canada)*.

Card, S., Mackinlay, J., & Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision To Think. Information Visualization - IVS.*

Ceriani, M., & Bottoni, P. (2017). SparqlBlocks : Using Blocks to Design Structured Linked Data Queries. *Journal of Visual Languages and Sentient Systems*, *3*, 1–21.

Crotti Junior, A., Debruyne, C., Brennan, R., & O'Sullivan, D. (2017). An evaluation of uplift mapping languages. *International Journal of Web Information Systems*, *13*(4), 405–424. http://doi.org/10.1108/IJWIS-04-2017-0036

Crotti Junior, A., Debruyne, C., Longo, L., & O'Sullivan, D. (2018). On the Mental Workload Assessment of Uplift Mapping Representations in Linked Data. In *Human Mental Workload: Models and Applications - Second International Symposium, H-WORKLOAD 2018*.

Crotti Junior, A., Debruyne, C., & O'Sullivan, D. (2017). Juma: An Editor that Uses a Block Metaphor to Facilitate the Creation and Editing of R2RML Mappings. In *The Semantic Web: ESWC 2017 Satellite Events - ESWC 2017 Satellite Events, Portorož, Slovenia, May 28 - June 1, 2017, Revised Selected Papers* (Vol. 10577 LNCS, pp. 87–92). Springer. http://doi.org/10.1007/978-3-319-70407-4_17

Crotti Junior, A., Debruyne, C., & O'Sullivan, D. (2018a). An Editor that Uses a Block Metaphor for Representing Semantic Mappings in Linked Data. In *The Semantic Web: ESWC 2018 Satellite Events - ESWC 2017 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*.

Crotti Junior, A., Debruyne, C., & O'Sullivan, D. (2018b). Juma Uplift : Using a Block Metaphor for Representing Uplift Mappings. In *12th IEEE International Conference on Semantic Computing, ICSC 2018, Laguna Hills, CA, USA, January 31 - February 2, 2018* (pp. 211–218). IEEE Computer Society. http://doi.org/10.1109/ICSC.2018.00037

Crotti Junior, A., Debruyne, C., & O'Sullivan, D. (2018c). Using a Block Metaphor for Representing R2RML Mappings. In *Proceedings of the Third International Workshop on Visualization and Interaction for Ontologies and Linked Data co-located with the 16th International Semantic Web Conference , Vienna, Austria, October 22, 2017.* (Vol. 1947, pp. 211–218). CEUR-WS.org. http://doi.org/10.1109/ICSC.2018.00037

Cruz, I. F., Antonelli, F. P., & Stroe, C. (2009). AgreementMaker : Efficient Matching for Large Real-World Schemas and Ontologies. *Pvldb*, *2*(2), 1586–1589. http://doi.org/10.14778/1687553.1687598

Das, S., Sundara, S., & Cyganiak, R. (2012). R2RML: RDB to RDF Mapping Language, W3C Recommendation 27 September 2012. *W3C Recommendation*.

David, J., Euzenat, J., Scharffe, F., & dos Santos, C. (2011). The alignment API 4.0. *Semantic Web*, *2*(1), 3–10.

De Meester, B., Dimou, A., Verborgh, R., & Mannens, E. (2016). An ontology to semantically declare and describe functions. In *The Semantic Web - ESWC 2016 Satellite Events, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers* (Vol. 9989 LNCS, pp. 46–49). http://doi.org/10.1007/978-3-319-47602-5_10

De Meester, B., Heyvaert, P., Dimou, A., & Verborgh, R. (2018). Towards a Uniform User Interface for Editing Data Shapes. In *Proceedings of the Fourth International Workshop on Visualization and Interaction for Ontologies and Linked Data co-located with the 17th International Semantic Web Conference, Monterey, CA, USA, October 8, 2018.* (pp. 13–24). Retrieved from http://ceur-ws.org/Vol-2187/paper2.pdf

Debruyne, C., Beyan, O. D., Grant, R., Collins, S., Decker, S., & Harrower, N. (2016). A

semantic architecture for preserving and interpreting the information contained in Irish historical vital records. *Int. J. on Digital Libraries*, *17*(3), 159–174. http://doi.org/10.1007/s00799-016-0180-8

Debruyne, C., & O'Sullivan, D. (2016). R2RML-F: Towards Sharing and Executing Domain Logic in R2RML Mappings. *Proceedings of the Workshop on Linked Data on the Web, LDOW 2016, Co-Located with 25th International World Wide Web Conference (WWW 2016)*.

Debruyne, C., Walshe, B., & O'Sullivan, D. (2015). Towards a project centric metadata model and lifecycle for ontology mapping governance. In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications &Services - iiWAS '15* (pp. 1–10). http://doi.org/10.1145/2837185.2837201

Dimou, A., Kontokostas, D., Freudenberg, M., Verborgh, R., Lehmann, J., Mannens, E., … Van de Walle, R. (2015). Assessing and Refining Mappings to RDF to Improve Dataset Quality. *The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II*.

Dimou, A., Sande, M. Vander, Colpaert, P., Verborgh, R., Mannens, E., & Van De Walle, R. (2014). RML: A generic language for integrated RDF mappings of heterogeneous data. *Proceedings of the Workshop on Linked Data on the Web Co-Located with the 23rd International World Wide Web Conference (WWW 2014)*.

Dzbor, M., Motta, E., Buil, C., Gomez, J. M., Görlitz, O., & Lewen, H. (2006). Developing Ontologies in OWL: an Observational Study. In *Proceedings of the OWLED*06 Workshop on OWL: Experiences and Directions, Athens, Georgia, USA, November 10-11, 2006*. Retrieved from http://ceur-ws.org/Vol-216/submission_30.pdf

Euzenat, J. (2004). An API for Ontology Alignment. In *The Semantic Web – ISWC 2004* (Vol. 3298, pp. 698–712). http://doi.org/10.1007/978-3-540-30475-3_48

Euzenat, J., Loup, D., Touzani, M., & Valtchev, P. (2004). Ontology alignment with OLA. In *3rd ISWC2004 workshop on Evaluation of Ontology-based tools (EON)* (pp. 59–68). Retrieved from http://ceur-ws.org/Vol-128/EON2004_EXP_Euzenat.pdf

Euzenat, J., & Shvaiko, P. (2007). *Ontology matching*. Springer. http://doi.org/10.1007/978-3-540-49612-0

Falconer, S. M., & Storey, M. A. (2007). A cognitive support framework for ontology mapping. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007*. (Vol. 4825 LNCS, pp. 114–127). http://doi.org/10.1007/978-3-540-76298-0_9

Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I. F., & Couto, F. M. (2013). The AgreementMakerLight ontology matching system. In *On the Move to Meaningful Internet Systems: OTM 2013 Conferences - Confederated International Conferences: CoopIS, DOA-Trusted Cloud, and ODBASE 2013, Graz, Austria, September 9-13, 2013. Proceedings* (Vol. 8185 LNCS, pp. 527–541). http://doi.org/10.1007/978-3-642-41030-7_38

Fraser, N. (2018). Blockly - Google Developers. Retrieved August 20, 2018, from https://developers.google.com/blockly/

Fu, B., Noy, N. F., & Storey, M. A. (2013). Indented tree or graph? A usability study of ontology visualization techniques in the context of class mapping evaluation. In *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I* (Vol. 8218 LNCS, pp. 117–134). http://doi.org/10.1007/978-3-642-41335-3_8

García-Zubía, J., Angulo, I., Martínez-Pieper, G., Orduña, P., Rodríguez-Gil, L., &

Hernandez-Jayo, U. (2018). Learning to Program in K12 Using a Remote Controlled Robot: RoboBlock. In *Online Engineering & Internet of Things* (pp. 344–358). http://doi.org/10.1007/978-3-319-64352-6_33

Ghoniem, M., Fekete, J. D., & Castagliola, P. (1997). On the readability of graphs using node-link and matrix-based representations: A controlled experiment and statistical analysis. *Information Visualization*, *4*(2), 114–135. http://doi.org/10.1057/palgrave.ivs.9500092

Granitzer, M., Sabol, V., Onn, K. W., Lukose, D., & Tochtermann, K. (2010). Ontology Alignment—A Survey with Focus on Visually Supported Semi-Automatic Techniques. *Future Internet*, *2*(3), 238–258. http://doi.org/10.3390/fi2030238

Haase, P., Lewen, H., Studer, R., & Tran, D. (2008). The neon ontology engineering toolkit. In *Www* (pp. 4–6). http://doi.org/10.1.1.141.4163

Harris, S., Seaborne, A., & Prud'hommeaux, E. (2013). SPARQL 1.1 query language. *W3C Recommendation*, *21*(10).

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, pp. 904–908). http://doi.org/10.1177/154193120605000909

Hayes, P. J., & Patel-Schneider, P. F. (2014). RDF 1.1 Semantics. W3C Recommendation, February 2014. *World Wide Web Consortium. Retrieved from Https://Www. W3. Org/TR/2014/REC-Rdf11-Mt-20140225*.

Heyvaert, P., Dimou, A., De Meester, B., Seymoens, T., Herregodts, A. L., Verborgh, R., … Mannens, E. (2018). Specification and implementation of mapping rule visualization and editing: MapVOWL and the RMLEditor. *Journal of Web Semantics*, *49*, 31–50. http://doi.org/10.1016/j.websem.2017.12.003

Heyvaert, P., Dimou, A., Herregodts, A. L., Verborgh, R., Schuurman, D., Mannens, E., & Van de Walle, R. (2016). RMLEditor: A graph-based mapping editor for linked data mappings. In *The Semantic Web. Latest Advances and New Domains - 13th International Conference, ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016* (Vol. 9678, pp. 709–723). http://doi.org/10.1007/978-3-319-34129-3_43

Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (2012, December). OWL 2 Web Ontology Language Primer . W3C Recommendation 11 December 2012. Retrieved from http://www.w3.org/TR/owl2-overview/

Hoefler, P., Granitzer, M., Veas, E., & Seifert, C. (2014). Linked data query wizard: A novel interface for accessing sparql endpoints. In *Proceedings of the Workshop on Linked Data on the Web co-located with the 23rd International World Wide Web Conference (WWW 2014)* (Vol. 1184).

Horrocks, I., Patel-schneider, P. F., Boley, H., Tabet, S., Grosof, B., & Dean, M. (2004). SWRL : A Semantic Web Rule Language Combining OWL and RuleML. *W3C Member Submission 21*, *21*(May 2004), 1–20.

Katifori, A., Torou, E., Halatsis, C., Lepouras, G., & Vassilakis, C. (2006). A Comparative Study of Four Ontology Visualization Techniques in Protègé: Experiment Setup and Preliminary Results. In *10th International Conference on Information Visualisation, IV 2006, 5-7 July 2006, London, UK* (pp. 417–423). http://doi.org/10.1109/IV.2006.3

Knoblock, C. A., Szekely, P., Ambite, J. L., Goel, A., Gupta, S., Lerman, K., … Mallick, P. (2012). Semi-automatically mapping structured sources into the semantic web. In *The Semantic Web: Research and Applications - 9th Extended Semantic Web Conference, ESWC 2012, Heraklion, Crete, Greece, May 27-31, 2012.* (Vol. 7295 LNCS, pp. 375–390). http://doi.org/10.1007/978-3-642-30284-8_32

Lambrix, P., & Ivanova, V. (2013). A unified approach for debugging is-a structure and

mappings in networked taxonomies. *Journal of Biomedical Semantics*, *4*(1), 10. http://doi.org/10.1186/2041-1480-4-10

Lambrix, P., & Tan, H. (2006). SAMBO-A system for aligning and merging biomedical ontologies. *Web Semantics*, *4*(3), 196–206. http://doi.org/10.1016/j.websem.2006.05.003

Lanzenberger, M., & Sampson, J. (2006). AlViz - A tool for visual ontology alignment. In *Proceedings of the International Conference on Information Visualisation* (pp. 430–438). IEEE Computer Society. http://doi.org/10.1109/IV.2006.18

Lembo, D., Rosati, R., Ruzzi, M., Savo, D. F., & Tocci, E. (2014). Visualization and Management of Mappings in Ontology-based Data Access (Progress Report). In *Proceedings of the 27th International Workshop on Description Logics, Vienna, Austria, July 17-20, 2014.* (pp. 595–607). Retrieved from http://ceur-ws.org/Vol-1193/paper_77.pdf

Lewis, J. R. (1992). Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 36, pp. 1259–1260). http://doi.org/10.1177/154193129203601617

Liu, Z., & Stasko, J. (2010). Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 999–1008. http://doi.org/10.1109/TVCG.2010.177

Lohmann, S., Link, V., Marbach, E., & Negru, S. (2015). WebVOWL: Web-based visualization of ontologies. In *Knowledge Engineering and Knowledge Management - EKAW 2014 Satellite Events, VISUAL, EKM1, and ARCOE-Logic, Linköping, Sweden, November 24-28, 2014. Revised Selected Papers.* (Vol. 8982, pp. 154–158). http://doi.org/10.1007/978-3-319-17966-7_21

Lohmann, S., Negru, S., Haag, F., & Ertl, T. (2016). Visualizing ontologies with VOWL. *Semantic Web*, *7*(4), 399–419. http://doi.org/10.3233/SW-150200

Longo, L. (2015). A defeasible reasoning framework for human mental workload representation and assessment. *Behaviour and Information Technology*, *34*(8), 758–786. http://doi.org/10.1080/0144929X.2015.1015166

Longo, L. (2016). Mental Workload in Medicine: Foundations, Applications, Open Problems, Challenges and Future Perspectives. In *2016 IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 106–111). http://doi.org/10.1109/CBMS.2016.36

Marnette, B., Mecca, G., Papotti, P., Raunich, S., & Santoro, D. (2011). ++ Spicy: an OpenSource Tool for Second-Generation Schema Mapping and Data Exchange. *Proceedings of the VLDB Endowment*, *4*(12), 1438–1441. Retrieved from http://www.vldb.org/pvldb/vol4/p1438-marnette.pdf

Marshall, M. S., Boyce, R., Deus, H. F., Zhao, J., Willighagen, E. L., Samwald, M., … Stephens, S. (2012). Emerging practices for mapping and linking life sciences data using RDF — A case series. *Web Semantics: Science, Services and Agents on the World Wide Web*, *14*, 2–13. http://doi.org/10.1016/J.WEBSEM.2012.02.003

Meehan, A., Brennan, R., Lewis, D., & O'Sullivan, D. (2014a). Mapping Representation based on Meta-data and SPIN for Localization Workflows. In *Joint Proceedings of the Second International Workshop on Semantic Web Enterprise Adoption and Best Practice and Second International Workshop on Finance and Economics on the Semantic Web Co-located with 11th European Semantic Web Conference, WaSABi-FEOSW*. Retrieved from http://ceur-ws.org/Vol-1240/wasabi2014-paper4.pdf

Meehan, A., Brennan, R., Lewis, D., & O'Sullivan, D. (2014b). Mapping Representation based on Meta-data and SPIN for Localization Workflows. In *Joint Proceedings of the Second International Workshop on Semantic Web Enterprise Adoption and Best Practice and Second International Workshop on Finance and Economics on the Semantic Web Co-located with 11th European Semantic Web Conference, WaSABi-FEOSW*.

Michel, F., Djimenou, L., Faron-Zucker, C., & Montagnat, J. (2015). Translation of relational and non-relational databases into RDF with xR2RML. *Proceedings of the 11th International Conference on Web Information Systems and Technologies (WEBIST 2015)*, 443–454. http://doi.org/10.5220/0005448304430454>

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM, 38*(11), 39–41. http://doi.org/10.1145/219717.219748

Munzner, T. (2014). *Visualization analysis and design*. AK Peters/CRC Press.

Neto, L. E. T., Vidal, V. M. P., Casanova, M. A., & Monteiro, J. M. (2013). R2RML by assertion: A semi-automatic tool for generating customised R2RML mappings. *The Semantic Web: ESWC 2013 Satellite Events - ESWC 2013 Satellite Events, Montpellier, France, May 26-30, 2013, Revised Selected Papers, 7955 LNCS*, 248–252. http://doi.org/10.1007/978-3-642-41242-4_33

Ngo, D., & Bellahsene, Z. (2012). YAM++: A multi-strategy based approach for ontology matching task. In *Proceedings of the Knowledge Engineering and Knowledge Management - 18th International Conference, EKAW 2012, Galway City, Ireland, October 8-12, 2012* (Vol. 7603 LNAI, pp. 421–425). http://doi.org/10.1007/978-3-642-33876-2_38

Nielsen, J. (1994). *Usability engineering*. Elsevier.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Constellation.

Noy, N. F., & Musen, M. A. (2003). The PROMPT suite: Interactive tools for ontology merging and mapping. *International Journal of Human Computer Studies, 59*(6), 983–1024. http://doi.org/10.1016/j.ijhcs.2003.08.002

Nunnally, J. C. (1979). *Psychometric Theory* (Vol. null). McGraw-Hill New York.

Pietriga, E., Bizer, C., Karger, D., & Lee, R. (2006). Fresnel: A Browser-Independent Presentation Vocabulary for RDF, 158–171. http://doi.org/10.1007/11926078_12

Preece, J., Rogers, Y., & Sharp, H. (2001). *Beyond Interaction Design: Beyond Human-Computer Interaction*. New York, NY, USA: John Wiley & Sons, Inc.

Rahm, E. (2011). Towards Large-Scale Schema and Ontology Matching. *Schema Matching and Mapping*, 3–27. http://doi.org/10.1007/978-3-642-16518-4

Rietveld, L., & Hoekstra, R. (2017). The YASGUI family of SPARQL clients 1. *Semantic Web, 8*(3), 373–383. http://doi.org/10.3233/SW-150197

Rivero, C., & Hernández, I. (2011). Generating SPARQL executable mappings to integrate ontologies. *Conceptual Modeling*, 118–131. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-24606-7_10

Rodriguez, J. B., & Gómez-Pérez, A. (2006). Upgrading relational legacy data to the semantic web. In *Proceedings of the 15th international conference on World Wide Web - WWW '06* (p. 1069). http://doi.org/10.1145/1135777.1136019

Sana E Zainab, S., Saleem, M., Mehmood, Q., Zehra, D., Decker, S., & Hasnain, A. (2015). FedViz: A visual interface for SPARQL queries formulation and execution. In *Proceedings of the International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data co-located with 14th International Semantic Web Conference (ISWC 2015), Bethlehem, Pennsylvania, USA, October 11, 2015.* (Vol.

1456, pp. 49–60).

Scharffe, F., Atemezing, G., Troncy, R., Gandon, F., Villata, S., Bucher, B., … Vatant, B. (2012). Enabling linked data publication with the Datalift platform. In *Semantic Cities, Papers from the 2012 AAAI Workshop, Toronto, Ontario, Canada, July 22-23, 2012.*

Schulz, H. J. (2011). Treevis.net: A tree visualization reference. *IEEE Computer Graphics and Applications*, *31*(6), 11–15. http://doi.org/10.1109/MCG.2011.103

Sequeda, J. F., & Miranker, D. P. (2017). A pay-as-you-go methodology for ontology-based data access. *IEEE Internet Computing*, *21*(2), 92–96. http://doi.org/10.1109/MIC.2017.46

Sequeda, J., Priyatna, F., & Villazón-Terrazas, B. (2012). Relational database to RDF mapping patterns. *Proceedings of the 3rd Workshop on Ontology Patterns, Boston, USA, November 12, 2012*, *929*, 97–108.

Severo, B., Trojahn, C., & Vieira, R. (2014). VOAR: A Visual and Integrated Ontology Alignment Environment. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014.* (p. 6).

Severo, B., Trojahn, C., & Vieira, R. (2017). VOAR 3.0: A configurable environment for manipulating multiple ontology alignments. In *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks co-located with 16th International Semantic Web Conference ISWC 2017, Vienna, Austria, October 23rd - to - 25th, 2017.* (Vol. 1963). Retrieved from http://ceur-ws.org/Vol-1963/paper550.pdf

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6th ed.). Pearson.

Shvaiko, P., & Euzenat, J. (2013). Ontology matching: State of the art and future challenges. *IEEE Transactions on Knowledge and Data Engineering*, *25*(1), 158–176. http://doi.org/10.1109/TKDE.2011.253

Sicilia, Á., Nemirovski, G., & Nolle, A. (2017). Map-On: A web-based editor for visual ontology mapping. *Semantic Web*, *8*(6), 969–980. http://doi.org/10.3233/SW-160246

Slepicka, J., Yin, C., Szekely, P., & Knoblock, C. (2015). KR2RML: An Alternative Interpretation of R2RML for Heterogenous Sources. In *Proceedings of the 6th International Workshop on Consuming Linked Data (COLD 2015)* (pp. 1–12). Retrieved from http://ceur-ws.org/Vol-1426/paper-08.pdf

Stadler, C., Unbehauen, J., Westphal, P., Sherif, M. A., & Lehmann, J. (2015). Simplified RDB2RDF Mapping. In *Proceedings of the Workshop on Linked Data on the Web, LDOW 2015, co-located with the 24th International World Wide Web Conference (WWW 2015).*

Stoilos, G., Trivela, D., Vassalos, V., Venetis, T., & Xarchakos, Y. (2017). Enabling data integration using MIPMap. In *Proceedings of Data Integration in the Life Sciences - 12th International Conference, DILS 2017, Luxembourg, Luxembourg, November 14-15, 2017* (Vol. 10649 LNBI, pp. 86–93). http://doi.org/10.1007/978-3-319-69751-2_9

Thakker, D., Dimitrova, V., Lau, L., Yang-Turner, F., & Despotakis, D. (2013). Assisting user browsing over linked data: Requirements elicitation with a user study. In *Proceedings of Web Engineering - 13th International Conference, ICWE 2013, Aalborg, Denmark, July 8-12, 2013.* (Vol. 7977 LNCS, pp. 376–383). http://doi.org/10.1007/978-3-642-39200-9_31

Thiéblin, É., Amarger, F., Haemmerlé, O., Hernandez, N., & Trojahn, C. (2016). Rewriting SELECT SPARQL queries from 1:n complex correspondences. *Proceedings of the*

11th International Workshop on Ontology Matching Co-Located with the 15th International Semantic Web Conference (ISWC 2016), *1766*, 49–60.

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, *39*(3), 358–381. http://doi.org/10.1080/00140139608964470

Tufte, E. (1990). *Envisioning Information*. Cheshire, CT, USA: Graphics Press.

Ware, C. (2012). *Information Visualization: Perception for Design*. Elsevier.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance. Engineering psychology and human performance* (Vol. 2nd). Psychology Press.

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. http://doi.org/10.1518/001872008X288394

Wolber, D., Abelson, H., Spertus, E., & Looney, L. (2014). *App Inventor: Create your Own Android Apps* (2nd ed.). O'Reilly Media, Inc.

Ziemkiewicz, C., & Kosara, R. (2008). The shaping of information by visual metaphors. *IEEE Transactions on Visualization and Computer Graphics*, *14*(6), 1269–1276. http://doi.org/10.1109/TVCG.2008.171

Appendix A. PSSUQ questionnaire

#	Question
1	Overall, I am satisfied with how easy it is to use this system
2	It was simple to use this system
3	I could effectively complete the tasks and scenarios using this system
4	I was able to complete the tasks and scenarios quickly using this system
5	I was able to efficiently complete the tasks and scenarios using this system
6	I felt comfortable using this system
7	It was easy to learn to use this system
8	I believe I could become productive quickly using this system
9	The system gave error messages that clearly told me how to fix problems
10	Whenever I made a mistake using the system, I could recover easily and quickly
11	The information (such as on-line help, on-screen messages, and other documentation) provided with this system was clear
12	It was easy to find the information I needed
13	The information provided for the system was easy to understand
14	The information was effective in helping me complete the tasks and scenarios
15	The organization of information on the system screens was clear
16	The interface of this system was pleasant
17	I liked using the interface of this system
18	This system has all the functions and capabilities I expect it to have
19	Overall, I am satisfied with this system

Appendix B. Workload Profile questionnaire

Label	Question
WP_1	How much attention was required for activities like remembering, problem-solving, decision-making, perceiving (detecting, recognizing, identifying objects)?
WP_2	How much attention was required for selecting the proper response channel (manual - keyboard/mouse, or speech - voice) and its execution?
WP_3	How much attention was required for spatial processing (spatially pay attention around)?
WP_4	How much attention was required for verbal material (eg. reading, processing linguistic material, listening to verbal conversations)?
WP_5	How much attention was required for executing the task based on the information visually received (eyes)?
WP_6	How much attention was required for executing the task based on the information auditorily received?
WP_7	How much attention was required for manually respond to the task (eg. keyboard/mouse)?
WP_8	How much attention was required for producing the speech response (eg. engaging in a conversation, talking, answering questions)?

Appendix C. NASA-TLX questionnaire

Label	Question
NT_1	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
NT_2	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
NT_3	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
NT_4	How hard did you have to work (mentally and physically) to accomplish your level of performance?
NT_5	How successful do you think you were in accomplishing the goals, of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
NT_6	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Appendix D. Informed consent

This is the informed consent signed by participants of the user experiments presented in this thesis.

LEAD RESEARCHER: Ademar Crotti Junior

BACKGROUND OF RESEARCH:
The majority of data in the Web still resides in other formats than the Resource Description Framework (RDF). RDF is a W3C recommendation for representing information in the Web, facilitating data exchange, data integration and others. One of the main tasks when upgrading legacy systems to the Semantic Web is the conversion of data. The process of converting data in any format into RDF is called uplift . The key stakeholders in this process are web developers, software programmers specialized in the development of systems for the web, and ontology engineers, experts in semantic web technologies such as ontologies, RDF and so on.
Several solutions have been proposed, however, these still focus on Semantic Web experts. To facilitate the uplift process and to make the technology available to a wider set of stakeholders, I have developed a method to represent uplift mappings visually. The method draws inspiration from visual programming languages such as Google's Blockly. Blockly has been used in many projects, such as code.org's introduction courses. In the visual representation, blocks represent a mapping that automatically generates an uplift mapping. In this experiment, I aim to investigate if such a visual representation: (i) facilitates the creation of accurate uplift mappings; (ii) eases the understandability of uplift mappings and (iii) imposes an optimal mental workload on users.

PROCEDURES OF THIS STUDY: You have been chosen for this study because: (i) you responded to the lead researcher's email requesting your participation. This email was sent to you because you are a part of the same research group as the lead researcher in this study (note that this may create a conflict of interest). (ii) You have knowledge on Semantic Web technologies or more specifically on uplifting data into RDF. (iii) You are a student of technical disciplines in Computer Science.

The experiment should take about 30 minutes where:
- The researcher will present basic information about relevant technologies related to this study.
- You will be asked to fill a pre-questionnaire.
- You will be asked to work on 1 task.
- Finally, you will be asked to fill questionnaires about the use of the tool.

Participation in this study is entirely voluntary. If, for any reason, you wish to terminate your participation, you are free to do so.

In the extremely unlikely event that illicit activity is reported I will be obliged to report it to appropriate authorities.

PUBLICATION: It is my intention to publish the results of this evaluation in conferences and/or scientific journals. It is also my intention to use these results in my PhD thesis.

Individual results may be aggregated anonymously and research reported on aggregate results.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that my participation is fully anonymous and that no personal details about me will be recorded.
- <If the research involves viewing materials via a computer monitor> I understand that if I or anyone in my family has a history of epilepsy then I am proceeding at my own risk.
- I have received a copy of this agreement.

PARTICIPANT'S NAME:
PARTICIPANT'S SIGNATURE:
Date:

Statement of investigator's responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

RESEARCHERS CONTACT DETAILS:
Email: crottija@scss.tcd.ie
Phone: XXX XXXXXXX (omitted for publication of the thesis)

INVESTIGATOR'S SIGNATURE:

Date:

Appendix E. USB Flash Drive Contents

The accompanying USB flash drive with this thesis makes available the code for the applications applying the Juma approach in the representation of Linked Data mappings and all the data used in experiments 4 and 5.

All data used in Experiment 4 and 5 can be found in the `experiment_data/experiment_4` and `experiment_data/experiment_5` folders, respectively.

The folders `juma-r2rml.zip`, `juma-uplift.zip`, and `juma-interlink.zip` contain the code for Juma R2RML, Juma Uplift and Juma Interlink applications, respectively.