Assessing the Quality of R2RML Mappings

Ademar Crotti Junior, Jeremy Debattista, Declan O'Sullivan

ADAPT Centre for Digital Content Platform Research, Knowledge & Data Engineering Group, School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland
{ademar.crotti, jeremy.debattista, declan.osullivan}@adaptcentre.ie

Abstract. This paper presents an approach to assess the quality of mappings used to generate RDF datasets. Data quality is a multidimensional concept determined by many factors which influence the extent by which a dataset is useful for a particular task. Several solutions have been proposed in literature to assess the quality of RDF datasets. Nonetheless, in most cases, these solutions focus on the resulting datasets and not on the artefacts used to generate these. In this paper, we propose the use of metrics commonly used to assess the quality of such datasets to evaluate the mappings used to generate them. The goal is to assist data providers into producing high quality datasets by bringing such quality assessment procedures to also cover the start of the publishing process. We provide an implementation of the approach by extending an existing quality assessment framework, which is then evaluated using real world use cases. Preliminary results shows that the assessment of mappings is capable to identifying quality issues for the observed cases.

Keywords. Mapping Quality; Data Mapping; R2RML.

1 Introduction

Data quality is a complex multidimensional concept involving various aspects by which one can characterize the quality of a dataset for a particular task [1]. Data quality problems, such as inaccuracy, incompleteness, and inconsistency, imply serious limitations to the full exploitation of data [2]. While several quality assessment frameworks have been proposed, in most cases, these focus on the resulting datasets and not on the artefacts used to produce them [3].

In the Linked Data domain, these artefacts are often mappings. Such mappings define the required transformations needed to convert non-RDF resources to RDF [4]. In order to express such transformations, one may avail of the W3C Recommendation RDB-to-RDF mapping language (R2RML) [5], which allows one to declaratively express customized mappings from relational databases to RDF.

In this paper, we propose the use of metrics commonly used to assess the quality of RDF datasets to evaluate the mappings used to generate them. The assessment of mappings used to produce datasets brings quality procedures and their subsequent cleaning and fixing to the start of the publishing process. Since these consume considerable time and resources, such quality mapping assessment is expected to positively impact the economic cost and viability of publishing datasets. More specifically, an error in a declarative mapping may become exponentially larger in the resulting dataset. In order words, each mapping violation would lead to a violation in the final dataset on the number of values contained in the input source. Therefore, an incorrect mapping can be considered a *root cause error* in the respective resulting dataset. Identifying and fixing *poor quality* mappings earlier would also mean that the published dataset is guaranteed to be free of the determined quality deficiencies. The main contributions of this paper can be summarized as follows: i) an approach for assessing mappings that generate RDF datasets; ii) an implementation of the approach which extends Luzzu [6], a data quality assessment framework, which may also be integrated to mapping editors; and iii) an evaluation of the approach using real world use cases together with preliminary results which show that such assessment of mappings is capable of identifying violations and inconsistencies for the observed cases.

The remainder of this paper is organised as follows: Section 2 briefly describes the R2RML mapping language. Section 3 presents our proposed approach to assess the quality of mappings. Section 4 presents an evaluation and its initial results executed with real world use cases. Section 5 describes the related work. Section 6 concludes the paper and discusses future work.

2 R2RML Mapping Language

The RDB to RDF Mapping Language[1] (R2RML) [5] is the W3C Recommendation for the transformation of relational databases to RDF. R2RML mappings consist of one or more *triples maps*. Each *triples map* has one (1) *logical table*, (2) one *subject map*, and (3) zero or more *predicate object maps*; (4) *Graph maps* may be used in *subject maps* or *predicate object maps* to assign triples to named graphs.

1. **Logical Table.** The table, view, or SQL query from which RDF will be generated.
2. **Subject Map.** Define the subjects of the RDF triples. These subjects can be IRIs or blank nodes. You may also define subjects to be instances of zero or more class types.
3. **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 may be IRIs, blank nodes or literals. For literal values, it is possible to define a datatype or a language tag. You may also link the subjects defined in triples maps trough *parent triples map*. A *parent triples map* can have zero or more join conditions.
4. **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 `rr:class` assertions, which play the role of `rdf:type`, 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

[1] `https://www.w3.org/TR/r2rml/`

triples generated by the *predicate object map* are stored in all graphs of the union. Otherwise they are stored in the default graph.

The *subject map*, the *predicate map* and the *object map* are called *term maps*. *Term maps* express how an RDF term – which may be an IRI, a blank node or a literal – is generated. A *term map* can be a constant-valued term map which always generates the same RDF term, a reference-valued term map that is the data value of a referenced column attribute from a given *logical table*, or a template-valued term map that is a valid string template that may contain referenced column attributes from a given *logical table*. **Listing 1** presents an example of an R2RML mapping.

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

<#TripleMap1>
  rr:logicalTable [
    rr:tableName "students";
  ];
  rr:subjectMap [
    rr:template "http://example.org/student/{id}";
    rr:class foaf:Person;
  ];
  rr:predicateObjectMap [
    rr:predicate foaf:name;
    rr:objectMap [ rr:column "name";
  ];
];.
```

Listing 1: R2RML mapping definition

In this mapping, a *triples map* defines the *logical table* to be *students* – which may be a table or view from a relational database. The same *triples map* defines the subjects of the triples to have the IRI template string `http://example.org/student/{id}`, where *id* is an attribute column coming from the logical table *students*. In this case, for row with *id* equals to 1, the execution of this mapping would generate the subject `http://example.org/student/1` and so on. The *subject map* also declares the subjects of the triples to be instances of the class `foaf:Person`. Finally, a *predicate object map* relates the subjects to the predicate `foaf:name` where the object of the triples are values coming from the attribute column *name* from the declared logical table *students*.

3 Assessing the Quality of R2RML Mappings

This section presents our proposed approach for assessing the quality of mappings. The main goal is to assist data providers into producing high quality datasets by assessing the quality of the mappings used to generate those datasets. In detail, this section presents a motivating example, a description of the proposed approach, four mapping quality metrics which have been implemented to assess R2RML mappings, a description stating how the quality reports generated by the quality assessment framework can be used to identify erroneous mappings, and finally, a discussion about the general limitations of the proposed approach.

3.1 Motivating Example

An example taken from the DBLP bibliography dataset (which is described and used in the evaluation presented in Section 4) is shown in Listing 2. In this mapping the property `dcterms:partsOf` is used to relate publications.

```
# ... prefixes ...
# ...

[] rr:predicateObjectMap [
    rr:predicate dcterms:partOf ;
    rr:objectMap [
      rr:parentTriplesMap <#Publications> ;
            rr:joinCondition [
                      rr:child "crossref" ;
                      rr:parent "dblp_key" ;
            ];
      ];
   ] .
```

Listing 2: Excerpt of the DBLP mapping

The execution of this mapping would generate triples using the aforementioned property on the number of rows returned from the input source. Each of these triples would have the same inconsistency where the property `dcterms:partOf` would be discovered to be undefined when dereferenced against its namespace[2]. These inconsistencies found in mappings, as mentioned in Section 1, can be classified as *root cause errors* since they are introduced at mapping level but are often only discovered when the dataset has been generated and published.

Our proposed approach allows one to define mapping quality metrics, where such errors and inconsistencies can be discovered and fixed at mapping design time. In order to assess mappings expressed in R2RML, such as in our example, a quality metric concerned with the usage of undefined properties[3] would access portions of the mapping where such properties are defined. In R2RML, these are found in *predicate maps* within *predicate object maps*, as explained in Section 2. This metric would then analyze each property used in the mapping and report an error when it is not possible to dereference them against their namespace.

3.2 Our approach

Data quality, as stated in Section 1, is a complex multidimensional concept involving various aspects by which one can characterize the quality of a dataset for a particular task [1]. In a survey presented in [1], the authors classify data quality in four categories: accessibility, intrinsic, context, and representational. These categories are further described in 18 dimensions, and 69 metrics. The authors also define a data quality

---

[2] As it is described in Section 4, the correct property is `dcterms:isPartOf`.

[3] The implementation of such metric using our approach is described in Section 3.3.2.

assessment *metric*, *measure*, or *indicator*, as a "procedure for measuring a data quality dimension" [1].

Mappings define how the RDF dataset will be formed, thus assessing the quality of mappings directly correlates to the quality of the resulting dataset. In this work, we propose the use of data quality assessment metrics to evaluate the mappings used to generate RDF datasets. Such assessment would supply data providers with quality information that can be used to identify and solve violations at an earlier stage of the data publishing lifecycle. We argue that the earlier data quality problems are identified and fixed the better, as mappings may be reused multiple times with different input sources. Thus, the quality assessment of mappings would avoid the propagation of violations to all datasets generated by a particular mapping.

In order to assess the quality of mappings, we have extended the Luzzu Framework[4]. Luzzu [7] is a scalable and customizable Linked Data quality assessment framework that is extensible (i.e it allows for new metrics to be added to the framework), and provides interoperable standardized quality metadata and quality problem reports. The latter is used to identify different types of problems in the assessed dataset (or in this case mappings). Even though Luzzu allows for the scalable assessment of RDF datasets by streaming, this is not required for mappings as, in contrast with RDF datasets, these mapping documents are usually much smaller in size. Nonetheless, Luzzu, as mentioned, provides an extensible framework for the implementation of custom quality metrics. Furthermore, Luzzu also allows for the generation of detailed reports together with metadata on the execution of quality metrics which are expressed through an ontology-driven process which allows for its reuse within other semantic frameworks and tools.

Our extension of the Luzzu framework makes use of an R2RML processor [8], which builds an in-memory data structure once the mappings are loaded in Luzzu. This data structure is internally exposed by Luzzu to the third party implemented metrics. This extension was supported by the implementation of four metrics related to the representational category of data quality[5]. These metrics draw inspiration from the ones presented in [2], which have been used to assess Linked Data datasets, being translated to assess mappings in this study. The representational category is concerned with the design of the data. In other words, metrics in these categories evaluate how well the data is represented in terms of best practices and guidelines [1].

3.3 Mapping Quality Metrics

The following subsections present four metrics implemented to assess the quality of mappings. Each metric is presented with the following structure:

- Discussion: a discussion describing the quality metric.
- Definition: how metric is calculated.
- Implementation: how the metric was implemented in the assessment of R2RML mappings.

[4] `https://github.com/Luzzu/Framework/tree/V5`

[5] `https://github.com/ademarcrotti/r2rml-quality-metrics`

3.3.1 Usage of undefined classes

The aim is to assess the use of undefined classes in a mapping.

Discussion. The use of classes without a formal definition is undesirable, as agents would not be able to understand how the data should be interpreted, for instance, during reasoning [1]. Errors leading to such invalid usage include syntactic errors, the use of inexistent classes, or schema dereferenceability[6] issues.

Classes are considered undefined when it is not possible to dereference them against their namespace.

Definition. This metric is defined as one minus the sum of all undefined classes used in a mapping divided by the total sum of classes in a mapping.

Implementation. This metric considers classes associated to *subject maps* through the property `rr:class`, and *predicate object maps* relating the property `rdf:type` to a constant *object map* through the property `rr:constant` (or its shortcut `rr:object`).

The use of template and column term maps, which are valid in R2RML, are not considered in this metric as this would require to access the input data in order to form those class type IRIs. In other words, this would mean that to fully assess the usage of undefined classes one would need to generate all class IRI types referencing columns in each *logical table* given in a mapping.

3.3.2 Usage of undefined properties

The aim is to assess the use of undefined properties in a mapping.

Discussion. In similar way to the metric assessing the usage of undefined classes, this metric is also related to syntactic errors, inexistent properties, and dereferencability issues.

Undefined properties are the ones where it is not possible to dereference them against their namespace.

Definition. This metric is defined as one minus the sum of all undefined properties used in a mapping divided by the total number properties.

Implementation. This metric considers properties defined using a constant *term map* through a *predicate map* (or the shortcut property `rr:predicate`). As mentioned, one may use template or column term maps in order to define how RDF terms are generated in R2RML, however, these refer to attribute columns in the *logical table*.

3.3.3 Usage of blank nodes

The aim is to assess the use of blank nodes in a mapping.

Discussion. The use of blank nodes is undesirable because they cannot be externally referenced. The scope of blank nodes, thus, is limited to the RDF documents in which they appear.

It is important to note that even though discourage, blank nodes may be necessary in a number of datasets, which would allow, for instance, the RDF representation of more complex structures [2]. Thus, this metric may be informative on how blank nodes have

[6] Linked Data principles recommend that all IRIs in a dataset are deferenceable i.e. that HTTP clients are able of accessing and receiving the resources identified by such IRIs [2].

been used to design a dataset, or deployed to identify use cases in which an IRI should have been used. Finally, users may also decide not to assess their mappings through this metric,

Definition. This metric is defined as the sum of all resources defined as blank nodes definitions divided by the total number resources in a mapping.

Implementation. This metric considers all resources defined as blank nodes in a mapping. R2RML allows one to define the subject and object of the triples as blank nodes by associating the *term map* generating an RDF term with a blank node term type (`rr:BlankNode`).

3.3.4 Usage of RDF reification

The aim is to evaluate the use of the RDF reification model in mappings[7].

Discussion. The usage of the RDF reification model, even with the introduction of property paths in SPARQL 1.1, is discouraged due to their complex syntax and semantics. Another issue is related to this data structure often being used in combination with blank nodes, which is also discouraged.

Similar to the usage of blank nodes, even though undesirable, a number of use cases may require the use of the RDF reification model. As an example, provenance may be designed to use RDF reification for a particular dataset [2].

Definition. This metric is defined as the sum of all classes and properties related to the use of reification divided by the total number of classes and properties defined in a mapping.

Implementation. This metric consider all classes in a mapping defined to be instances of `rdf:Statement`, and *predicate maps* defined with any of the properties `rdf:subject`, `rdf:predicate`, or `rdf:object`.

3.4 Mapping Quality Reports

The execution of quality metrics in Luzzu results in two ontology-driven reports, one on the problems found in the dataset/mappings, and another one related to quality metadata. In this paper, we are mostly interested in the generated problem reports. The idea is to help mapping engineers to easily identify quality problems in their mappings that will affect the whole dataset during its generation. Listing 3 shows an excerpt of a quality problem report generated by Luzzu that identifies `dcterms:partOf` as an unidentified property.

```
@base <https://w3id.org/lodquator/resource/> .
# ... other prefixes ...

<ba4e8bf9-7e40-4e19-9b62-fb96fce429d2>
    a qpro:QualityProblem ;
    qpro:isDescribedBy dqm:UndefinedPropertiesMetric ;
```

[7] This metric only assesses the use of reification in mappings as R2RML does not natively supports other RDF data structures, such as containers and collections. We do note that R2RML extensions supporting these data structures exist, and that the metric may be extended to support these in future work.

```
    qpro:problemStructure qpro:ModelContainer ;
    qpro:problematicThing <469a3186-8d9f-48e3-9027-
8458d887dca8> .


<469a3186-8d9f-48e3-9027-8458d887dca8>
  qpro:exceptionDescription dqm-prob:UndefinedProperty ;
  ex:undefinedProperty dcterms:partOf ;
  ex:onMapping <../TriplesMapPublications> ; ... .
```

Listing 3: Excerpt of a problem report generated by Luzzu

This semantically structured report can be used within mapping editing frameworks, such as the Jigsaw Puzzle for Representing Mappings (Juma) [12]. Juma is capable of representing Linked Data mappings through a block-based metaphor. Where the Juma Uplift application [13], one of its implementations, allows for the definition of mappings through an abstract block-based interface which generates mappings in multiple distinct mapping languages. Figure 1 illustrates a snippet from the Juma Uplift application [13], presenting how problematic parts of the mappings may be highlighted to the user in order to show that the current mapping will output datasets of lesser quality. Combining Luzzu's semantic quality problem reports with Juma will enable data providers to identify and improve their mappings in order to produce high quality datasets at design time.

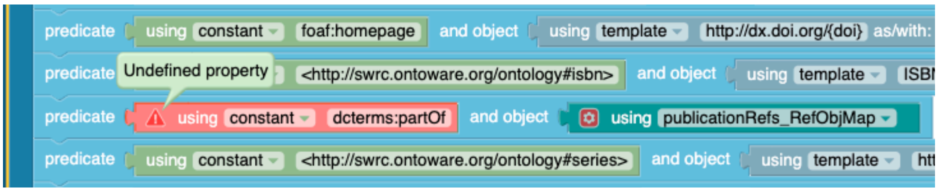

Figure 1: Snippet from the Juma Uplift application showing that the predicate `dcterms:partOf` is an undefined property.

3.5 Discussion

It is important to note that not all inconsistencies found in a particular dataset can be identified and fixed at mapping level. For instance, a number of issues depend on the input data (such as missing information) or is related to information that is not available in the mapping (such as the available serialization formats). As an example, a mapping may define the object of the triples to have a specific datatype. By only considering the information contained in the mapping no issues can be identified, however, violations may still occur when the mapping is executed to produce the final dataset. In this sense, the proposed approach does not replace existing quality assessment frameworks, and instead extends them to also cover the data generation and publication process. As mentioned, the main goal of our approach is to allow data providers in discovering and fixing mistakes by considering all the information available in the mapping so that any violations may be repaired before these are propagated to the final datasets.

4 Evaluation

This section presents preliminary results of an experiment evaluating our proposed approach to assess R2RML mappings. Two real world use cases were used in this evaluation, one set of mappings devised by the MusicBrainz project, and another used in the DBLP bibliography dataset.

4.1 Use cases

We have evaluated the proposed framework using the following two use cases.

MusicBrainz. MusicBrainz[8] is an open music encyclopedia containing information about artists, releases and recordings. A set of 12 R2RML mappings used to generate the dataset are also available[9].

DBLP. The Computer Science bibliography (DBLP) collects open bibliographic information from major computer science journals and proceedings. The original DBLP mappings are defined using D2RQ [10], which have been converted to R2RML and published in [3].

4.2 Results

Table 1 presents the results of the data quality metrics described in Section 3 applied for the use cases mappings of the MusicBrainz and DBLP use cases.

Mapping Quality Metric	MusicBrainz	DBLP
Usage of undefined classes	66.6%	40%
Usage of undefined properties	82.6%	76.9%
Usage of blank nodes	100%	100%
Usage of RDF reification	100%	100%

Table 1: Mapping quality assessment results

These results show that none of the mappings in our use cases make use of the RDF reification model or blank nodes. The results related to metrics evaluating the usage of undefined classes and properties show that, for the MusicBrainz use case, 33.4% of the classes and 17.4% of the properties are undefined (i.e. classes and properties which were not possible to dereference them against their namespace). For the DBLP use case, 60% of the classes and 23.1% of the properties were found to be undefined.

Upon further inspection, the mapping expressing the conversion of music records according to tags contains all the undefined classes and properties identified by our quality assessment. This mapping uses the Modular Unified Tagging Ontology [11], which provides a vocabulary to describe tags. The IRI (`http://purl.org/muto/core#`) used in the mapping, however, does not return the formal specification of the ontology.

In relation to the DBLP use case, we identified that one of the vocabularies (`http://swrc.ontoware.org/ontology#`), which is used in the version of the mapping published in [3], is not available. The quality metric assessing the usage

[8] `https://musicbrainz.org/`

[9] `https://github.com/metabrainz/MusicBrainz-R2RML`

of undefined properties has also identified two other violations. These are related to the mapping defining the ontology properties `dcterms:partOf` and `dcterms:tableOfContent` which were discovered to be undefined when dereferencing them against their namespace. After analyzing the DC Terms vocabulary (`http://purl.org/dc/terms/`), we have noticed that the mapping contains a typo and that the properties `dcterms:isPartOf` and `dcterms:tableOfContents` should have been used instead.

5 Related Work

A number of quality assessment frameworks have been proposed in literature to assess the Web of Data. For instance, Flemming [14] provides a simple web user interface and a step-by-step guide that aids users in assessing the data quality of a resource using a set of pre-defined metrics. LiQuate [15] is a quality assessment tool based on Bayesian Networks. The tRDF framework [16] provides a number of trust assessment metrics that determine the trustworthiness of RDF statements. LinkQA [17] is an assessment tool that measures the quality of a dataset using network analysis measures. Sieve [11] uses metadata about named graphs to assess data quality. RDFUnit [18] provides test-driven quality assessment for Linked Data through the SPARQL query language. SHACL [19] is the W3C Recommendation language for validating datasets against a set of conditions. Finally, Luzzu [7] is a quality assessment framework whose rationale is to provide an integrated platform that is scalable, extensible, interoperable and customizable. Moreover, Luzzu has been used to evaluate a number of datasets available in the Linked Open Data cloud [2].

The aforementioned approaches were designed to assess the quality of the (resulting) RDF datasets. Nonetheless, a number of errors found in such datasets may have been introduced at mapping level [3]. A single mapping error could have a great impact on the resulting dataset, where, for instance, an error applied to an input source containing 1000 rows would result in potentially 1000 violations in the final dataset. To our knowledge, only one approach has proposed the quality assessment of mappings [3], which is also done through the use of an existing quality assessment framework. This approach uses RDFUnit to create test cases which validate a mapping against the vocabularies and ontologies defined in the mapping. As mentioned, RDFUnit relies on the SPARQL query language in order to execute its test cases. In other words, not all quality metrics can be assessed by RDFUnit, such as the ones previously described in this paper. As an example, the assessment of undefined classes and properties cannot be computed by RDFUnit as SPARQL does not deference resources natively.

The novelty of our approach lies in proposing the use of quality metrics commonly used to evaluate RDF datasets to assess the processes that produced these datasets i.e. the mappings. Moreover, by extending Luzzu, we also allow for others to implement their own mapping quality metrics while reporting on the results of the quality assessment of their mappings through an ontology-driven approach. Similarly to the work presented in [20], the semantic description of quality reports can be used in combination with visualization and editing tools – where problems identified in ontologies and datasets can be presented to users. When mapping non-RDF resources to RDF, one may use, as mentioned in Section 3, the Juma [12] approach. In this case, the quality reports generated by the proposed Luzzu extension can be integrated to mapping editors. This integration would allow for violations and inconsistencies identified by the assessment of mappings to be presented to users prior to the generation and publishing of RDF datasets. Moreover, this integration would also allow for the quality of mappings to be assessed at design time (i.e. as the mappings are created by data providers with immediate feedback).

6 Conclusions and Future Work

While several quality assessment frameworks have been proposed, in most cases, such approaches remain independent of the mapping process, being executed by third parties rather than the data providers. This study tackles this issue by proposing an approach which allows one to assess the quality of the artefacts commonly used to generate RDF datasets – the mappings. This mapping quality assessment allows data publishers in evaluating their datasets prior to its generation and publishing, where each mapping violation potentially leads to multiple violations in the resulting dataset. We have also presented an implementation of the approach which extends an existent Linked Data quality assessment Framework, namely Luzzu. Preliminary results from an evaluation using two real world use cases proves the feasibility of the approach for the implemented quality metrics.

Future work includes the implementation of other metrics related to other quality categories and dimensions, such as the ones described in [1], reports explaining the issues identified in a mapping, and metrics that specifically evaluate the quality of mappings, such as mapping completeness – which would assess, for instance, the extent to which a dataset is being mapped. Future work also includes supporting mappings expressed in other mapping languages, such as RML [9]. RML is an extension of the R2RML mapping language to support other data formats, such as XML, JSON, amongst others. The integration of the assessment of mappings to editors and visualization tools, such as Juma, is also left as future work.

Acknowledgements. This paper was supported by the Science Foundation Ireland (Grant 13/RC/2106) as part of the ADAPT Centre for Digital Content Technology (http://www.adaptcentre.ie/) at Trinity College Dublin.

References

1. Zaveri, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J., Auer, S.: Quality assessment for Linked Data: A Survey. In: Semantic Web (2016). https://doi.org/10.3233/SW-150175.
2. Debattista, J., Lange, C., Auer, S., Cortis, D.: Evaluating the quality of the LOD cloud: An empirical investigation. Semant. Web. (2018). https://doi.org/10.3233/SW-180306.
3. Dimou, A., Kontokostas, D., Freudenberg, M., Verborgh, R., Lehmann, J., Mannens, E., Hellmann, S., Van de Walle, R.: Assessing and Refining Mappings to RDF to Improve Dataset Quality. Semant. Web - ISWC 2015 - 14th Int. Semant. Web Conf. Bethlehem, PA, USA, Oct. 11-15, 2015, Proceedings, Part II. (2015).
4. Crotti Junior, A., Debruyne, C., Brennan, R., O'Sullivan, D.: An evaluation of uplift mapping languages. Int. J. Web Inf. Syst. 13, 405–424 (2017). https://doi.org/10.1108/IJWIS-04-2017-0036.
5. Das, S., Sundara, S., Cyganiak, R.: R2RML: RDB to RDF Mapping Language, W3C

Recommendation 27 September 2012. W3C Recomm. (2012).
6. Debattista, J., Auer, Sö., Lange, C.: Luzzu—A Methodology and Framework for Linked Data Quality Assessment. J. Data Inf. Qual. (2016). https://doi.org/10.1145/2992786.
7. Debattista, J., Auer, S., Lange, C.: Luzzu-A Framework for Linked Data Quality Assessment. In: Proceedings - 2016 IEEE 10th International Conference on Semantic Computing, ICSC 2016 (2016). https://doi.org/10.1109/ICSC.2016.48.
8. Debruyne, C., O'Sullivan, D.: R2RML-F: Towards Sharing and Executing Domain Logic in R2RML Mappings. Proc. Work. Linked Data Web, LDOW 2016, co-located with 25th Int. World Wide Web Conf. (2016).
9. Dimou, A., Sande, M. Vander, Colpaert, P., Verborgh, R., Mannens, E., Van De Walle, R.: RML: A generic language for integrated RDF mappings of heterogeneous data. Proc. Work. Linked Data Web co-located with 23rd Int. World Wide Web Conf. (WWW 2014). (2014).
10. Bizer, C., Seaborne, A.: D2RQ – Treating Non-RDF Databases as Virtual RDF Graphs. In: Proceedings of the 3rd International Semantic Web Conference (ISWC2004). p. 26 (2004). https://doi.org/10.1.1.126.2314.
11. Lohmann, S., Díaz, P., Aedo, I.: MUTO: the modular unified tagging ontology. Proc. 7th Int. Conf. Semant. Syst. - I-Semantics '11. (2011).
12. Crotti Junior, A., Debruyne, C., O'Sullivan, D.: 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. pp. 1–12. CEUR-WS.org (2017).
13. Crotti Junior, A., Debruyne, C., O'Sullivan, D.: 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 (2018). https://doi.org/10.1109/ICSC.2018.00037.
14. Flemming, A.: Quality characteristics of linked data publishing datasources, (2010).
15. Ruckhaus, E., Vidal, M.E., Castillo, S., Burguillos, O., Baldizan, O.: Analyzing linked data quality with LiQuate. In: OTM Confederated International Conferences: On the Move to Meaningful Internet Systems (2014). https://doi.org/10.1007/978-3-319-11955-7_72.
16. Hartig, O.: Trustworthiness of Data on the Web. n STI Berlin CSW PhD Work. (2008).
17. Guéret, C., Groth, P., Stadler, C., Lehmann, J.: Assessing linked data mappings using network measures. In: The Semantic Web: Research and Applications. ESWC 2012. Lecture Notes in Computer Science, vol 7295. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30284-8_13.
18. Kontokostas, D., Westphal, P., Auer, S., Hellmann, S., Lehmann, J., Cornelissen, R., Zaveri, A.: Test-driven evaluation of linked data quality. In: Proceedings of the 23rd international conference on World Wide Web (WWW) (2014). https://doi.org/10.1145/2566486.2568002.
19. Knublauch, H., Kontokostas, D.: Shapes constraint language (SHACL). W3C Recomm. (2017).
20. Mc Gurk, S., Debattista, J., Abela, C.: OntoQAV: A pipeline for visualising ontology quality. In: International Semantic Web Conference (Posters, Demos & Industry Tracks) (2017).