OQuaRE-KG: an OQuaRE inspired framework for knowledge graph quality assessment

Belén Juanes-Cortés[1,2], Juan Mulero-Hernández[1,2] and Jesualdo Tomás Fernández-Breis[1,2,*]

[1]*Universidad de Murcia, CEIR Mare Nostrum, CP 30100, Murcia, Spain*
[2]*Instituto Murciano de Investigación Biosanitaria Pascual Parrilla, CP 30120, Murcia, Spain*

Abstract

Knowledge graphs play a central role in semantic data integration, knowledge-based applications, and artificial intelligence systems, due to their ability to represent entities, relations, and constraints in a structured and interoperable manner. As knowledge graphs increase in size, complexity, and heterogeneity, ensuring their quality has become a critical challenge for guaranteeing reliability, reuse, and effective reasoning. Existing approaches to knowledge graph quality assurance are largely fragmented, addressing isolated aspects such as schema validation, reasoning, data quality, or ontology evaluation, without offering an integrated and comprehensive assessment framework. In this work, we introduce OQuaRE-KG, a unified quality evaluation framework inspired by ontology quality models and enriched with established data quality principles. OQuaRE-KG integrates concepts from standards and methodologies in data quality assessment, linked data evaluation, and ontology quality evaluation into a coherent model specifically tailored to the dual nature of knowledge graphs, encompassing both schema-level and instance-level characteristics. We present the conceptual design of the framework, including its quality dimensions and assessment structure, and demonstrate its applicability through an evaluation of several biological Knowledge Graphs. The results illustrate OQuaRE-KG's ability to systematically characterize and differentiate knowledge graphs quality profiles, highlighting its potential as a foundational step toward standardized, reproducible knowledge graph quality evaluation.

Keywords

Knowledge graphs, Quality evaluation, Quantitative metrics

1. Introduction

Knowledge Graphs (KGs) have become a central piece in semantic data integration, knowledge-based applications and artificial intelligence systems. Its capacity to represent entities, relations and constraints makes them suitable for reasoning applications, interoperability or advance analysis [1]. However, as graphs are growing in complexity and heterogeneity, guaranteeing their quality became the main challenge to ensure their usefulness and reliability.

Quality assurance in KGs has traditionally been addressed in a fragmented manner. Existing approaches and tools typically focus on specific aspects, such as schema validation languages (SHACL, ShEx) [2, 3] and ontological reasoners (HermiT) [4]. However, current frameworks remain limited to isolated perspectives and do not provide an integrated process for a comprehensive assessment of a graph's overall quality. The work done in related fields such as data engineering, linked data and ontology quality evaluation should be considered in benefit of KG quality evaluation.

The ISO/IEC 25012 standard [5] distinguishes between two categories of data quality: (i) Inherent Data Quality, which includes characteristics such as completeness and consistency, and (ii) System-Dependent Data Quality, which includes characteristics such as availability and understandability. This distinction is useful for classifying different types of data quality requirements. In parallel, the FAIR principles have become a cornerstone for data management and sharing in scientific and industrial

Published in the Proceedings of the Workshops of the EDBT/ICDT 2026 Joint Conference (March 24-27, 2026), Tampere, Finland
*Corresponding author.
belen.juanesc@um.es (B. Juanes-Cortés); juan.mulero@um.es (J. Mulero-Hernández); jfernand@um.es (J. T. Fernández-Breis)
https://webs.um.es/jfernand (J. T. Fernández-Breis)
0000-0003-3641-7548 (B. Juanes-Cortés); 0000-0002-6818-3443 (J. Mulero-Hernández); 0000-0002-7558-2880 (J. T. Fernández-Breis)

contexts [6]. While the focus of FAIR is on enabling data discoverability and reuse, its implementation in KGs is closely tied to quality characteristics such as accessibility, interoperability, and provenance.

From the ontology quality evaluation perspective, OQuaRE [7] provides a quality evaluation model based on the SQuaRE standard. Its strength lies in providing a systematic and reproducible structure; nevertheless, its scope is confined to ontological artefacts, without considering instances typical from KGs [7].

Our hypothesis is that KG evaluation can benefit from the application of a systematic and reproducible framework, and that it makes sense adapting the OQuaRE principles to KGs given the semantic nature of both ontologies and KGs. Therefore, in this work, we introduce OQuaRE-KG, an OQuaRE-inspired framework that integrates principles from ontology quality and data quality assessment into a coherent model tailored for KGs. This article introduces the design of the framework and describes the application to a set of biological KGs to illustrate its capability to detect differences between KGs.

2. Related Work

The quality evaluation in KGs has been addressed from multiple perspectives in the literature, and this section contextualize the approach of OQuaRE-KG regarding previous works.

The framework proposed by Zaveri et al. [8] is considered to be one of the most comprehensive taxonomies, with 18 quality dimensions and 69 metrics. This framework categorizes dimensions such as completeness, availability, accessibility, consistency and provenance. This taxonomy has been essential for comprehending the complexity of evaluating data distributed on the web [8].

In [9] a methodology for assessing Linked Datasets is proposed. It introduces Luzzu, which provides an extensible framework for the automated calculation of Linked Data metrics. However, it should be noted that Luzzu offers only an execution environment and does not include a conceptual model.

Additionally, the empirical study conducted by Debattista et al. [10] evaluated 130 datasets from the LOD Cloud, using 27 metrics. This study builds on the work undertaken in the survey of Zavery et al. [8]. The quality metadata for each assessed dataset is published as Linked Data enabling data consumers to search and filter datasets based on different quality criteria.

Farber et al. [11] proposed a framework to determine the most suitable KG for applications based on data quality criteria and the analysis of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. That work presented 34 data quality criteria classified into 11 dimensions and 4 categories for evaluating KGs.

The framework developed [12] is based on the framework proposed by Zaveri et al., in which a comprehensive study of existing frameworks was conducted to evaluate the quality of KGs. The authors selected a set of quality dimensions and their corresponding metrics, which were mapped to 18 quality requirements and recommended an approach for evaluating each metric considering its feasibility and scalability.

FAIR principles [6] have gained increasing significance in the domain of research data management and reproducibility. To facilitate the adoption of best practices in FAIR by researchers, the scientific community has developed self-assessment tools and validation tools that help to assess the FAIRness of their resource. Self-assessment tools consist of online self-report questionnaires and checklists. The objectives of the questions are to reflect each of the FAIR principles. The questions are grouped according to each principle. An example of this kind of tool is Australian Research Data Commons (ARDC)[1] The FAIR Maturity Evaluation Service (FAIR evaluator) [13] is a fully automated tool based on community-driven efforts to compile discipline specific FAIR maturity indicators (MIs). FAIR Checker [14] also utilises the reference FAIR MIs.

[1]https://ardc.edu.au/resource/fair-data-self-assessment-tool/

3. Materials and Methods

3.1. The OQuaRE-KG Framework

OQuaRE-KG is a quality framework for assessing KGs inspired on the OQuaRE framework for ontology quality [7]. OQuaRE is a systematic framework designed to evaluate the quality of ontologies by adapting the ISO/IEC 25000:2005 SQuaRE standard, for Software Product Quality Requirements and Evaluation (SQuaRE).

The OQuaRE Quality Model Division addresses the internal and external quality of the ontology by defining key characteristics and sub-characteristics. The OQuaRE Quality Metrics Division establishes both base and derived measures for quality evaluation. The OQuaRE quality model consists of 7 characteristics, 39 subcharacteristics, and 19 metrics. Each characteristic has a set of subcharacteristics associated which, in turn, have a set of metrics associated.

In OQuaRE, the raw scores of the metrics are normalized to quality scores on a Likert scale [1,5], with 1 being the lowest quality score, and 5 the highest quality score. Then, a quality score can be calculated for each subcharacteristic and characteristic by averaging the scores of the different associated items. The detailed specification of OQuaRE can be found at https://github.com/tecnomod-um/oquare.

OQuaRE-KG follows the same design principles and adapts them to the evaluation of the quality of KGs. The next subsections describe the OQuaRE-KG quality model, whereas the whole framework can be found at https://github.com/tecnomod-um/oquare-kg.

The OQuaRE KG framework includes 7 characteristics found in the OQuaRE quality model: structural, functional adequacy, compatibility, transferability, operability, reliability, and maintainability. Next, we provide the definition adapted to KGs.

- *Structural*: The structural dimension accounts for the intrinsic design and structure of the KG, independent of the user's specific context.
- *Functional adequacy*: Functional adequacy refers to the capability of the KG to provide concrete functions.
- *Compatibility*: The ability of a KG to function correctly and be integrated with other KGs, data sources, or systems by adhering to common technical standards for data representation and exchange.
- *Transferability*: The degree to which a KG can be deployed, reused, or adapted across different platforms, domains, or technical environments with minimal modification, preserving its data, schema, and semantic definitions.
- *Operability*: The level of effort required by users (e.g., developers, data scientists, domain experts) to effectively access, query, navigate, and update the information stored within the KG to perform their tasks.
- *Reliability*: It refers to the capability of the KG to maintain its level of performance under stated conditions for a given period of time.
- *Maintainability*: The capability of a KG to be modified or extended in response to changes in environments, data sources, requirements, or functional specifications.

The current version of OQuaRE-KG includes 16 quality subcharacteristics. Each subcharacteristic is associated with one characteristic. Table 1 provides the set of subcharacteristics and their association with characteristics. Some subcharacteristics have been adapted from OQuaRE, but some have been adapted from works on Linked Data quality [15, 10] or inspired by the FAIR principles [6].

The current version of OQuaRE-KG also includes 28 quality metrics. Each metric is associated with one or more subcharacteristics. The metrics for the consistency subcharacteristic are shown in table 2. The description of the metrics, and the associations between metrics and the subcharacteristics, are also available in the OQuaRE-KG GitHub repository.

OQuaRE-KG scales the values of the metrics into quality scores in the range [1,5] (see Figure 1). We can classify the OQuaRE-KG metrics into three groups based on their range of values:

Table 1
The OQuaRE-KG quality subcharacteristics.

Characteristic	Subcharacteristic	Description
Structural	Formalisation	The capability of a knowledge graph's ontology or schema to enable reasoning processes
	Structural accuracy	Structural accuracy refers to the correctness of the use of ontological terms in the graph
	Consistency	Consistency means that two or more entities do not conflict with respect to knowledge representation and inference mechanisms
	Syntactic validity	Syntactic validity is defined as the degree to which an RDF document conforms to the specification of the serialization format
	Redundancy	Redundancy refers to the degree of avoidance of duplicate entities or relationships that could confuse the reasoning and querying of the graph
	Interpretability	The extent to which the information within the knowledge graph is machine-readable, semantically unambiguous and consistent, enabling automated systems to perform valid reasoning
Functional adequacy	Inference	The degree to which the ontology or schema of a knowledge graph can be used by reasoners to make implicit knowledge explicit within the graph
	Understandability	Understandability refers to the ease with which data can be comprehended without ambiguity and be used by a human information consumer
	Trustworthiness	Trustworthiness is defined as the degree to which the information is accepted to be correct, true, real and credible
	Provenance	Provenance refers to the provision of information regarding the origin of the dataset and of the resources within the dataset itself
	Clustering	Clustering refers to the degree to which entities representing the same or closely related concepts are accurately identified, grouped and linked within or across data sources
Compatibility	Interoperability	Interoperability in knowledge graphs refers to the use of formal, accessible, and shared ontologies, controlled vocabularies, and machine-readable formats in the representation of data and metadata, enabling seamless integration, automated reasoning, and cross-system understanding
Transferability	Versatility	Versatility refers to the availability of the data in different representations and in an internationalized way
Operability	Licensing	Licensing is defined as the granting of permission for a consumer to re-use a dataset under defined conditions
Reliability	Accessibility	Accessibility is the degree to which a knowledge graph and its metadata can be reliably retrieved by humans and machines through open, standardized protocols, with clear access conditions and persistent availability of metadata
Maintainability	Reusability	Reusability is the degree to which components of a knowledge graph can be effectively used in multiple contexts to build new systems or enrich other knowledge graphs

- Binary value: They are scaled to 1 or 5.
- Values in the range [0,1] are scaled [1,5] by intervals of 20%.
- Values in other ranges are scaled into [0,1] and then the previous scaling to [1,5] was applied.

In case high values of the metric are a sign of high quality, those high values are mapped onto the higher scores in the range [1,5]. Otherwise, the high values are mapped onto the lower scores. Once the quality scores of the metrics are calculated, the quality scores of the subcharacteristics and characteristics are obtained by averaging the scores of their corresponding associated metrics and

Table 2
Metrics for the consistency subcharacteristic.

Metric	Definition	Score
Classes per instance metric	Mean number of classes per instance to detect missing or multiple rdf:type assignments.	Positive real number; <1 means missing types, >1 means multiple types.
Compatible datatype	Assesses whether the lexical form of a literal matches its declared datatype.	Score from 0 to 1, where 1 indicates full datatype compatibility.
Entities with no type metric	Ratio of nodes lacking rdf:type in the graph.	Score from 0 to 1, where 0 indicates all entities have a type.
Instances with multiple types metric	Ratio of instances with more than one rdf:type.	Score from 0 to 1, where 0 indicates no multi-typed instances.
Misplaced classes or properties	Checks whether classes appear incorrectly in predicate position or properties appear incorrectly in object position.	Score from 0 to 1, where 0 indicates no misplaced terms.
Misused OWL datatype or object properties	Assesses correct usage of predicates according to owl:DatatypeProperty and owl:ObjectProperty axioms. Detects erroneous triples where literals are linked to object properties or entities to datatype properties.	Score from 0 to 1, where 0 indicates no misused properties.
Usage of undefined classes and properties	Checks if entities in the graph use terms not defined in any ontology.	Score from 0 to 1, where 0 indicates no undefined terms.

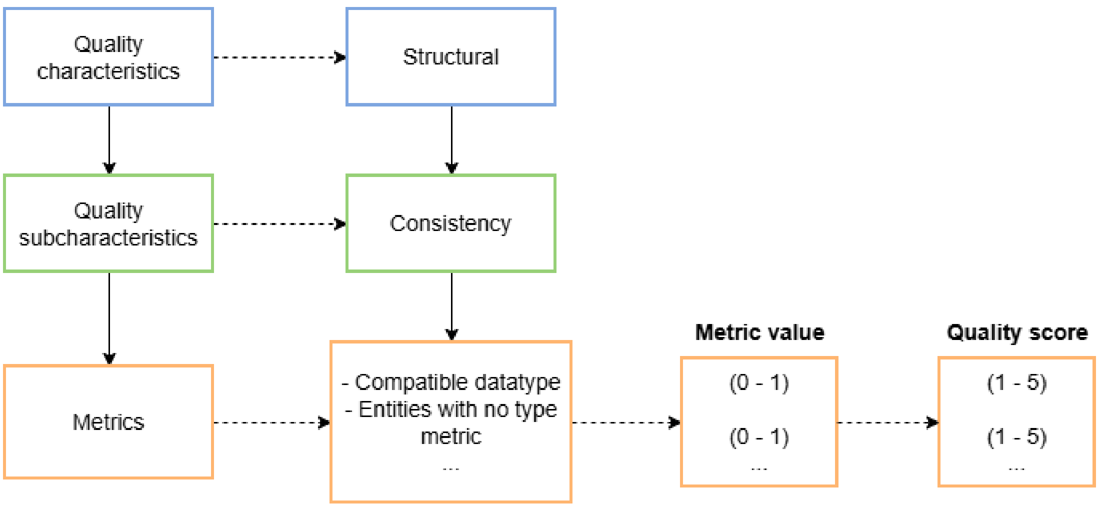

Figure 1: Metric values and their scale to quality scores.

subcharacteristics, respectively.

3.2. Use case

In this work, we have applied OQuaRE-KG to evaluate gene regulation KGs. In particular, we have used data from the BioGateway knowledge network available at https://github.com/juan-mulero/cisregEA [16]. This resource integrates data related to different domains of relevance in the biomedical context, such as proteins, genes and diseases. From the gene regulation perspective, it has a particular focus on enhancers, which are the most widely studied cis-regulatory modules (CRM) [17]. These sequences were modelled using the CisReg schema, which was used in BioGateway to integrate data from 25 different databases, modelling information from enhancer sequences and their relations with other entities. For the evaluation of OQuaRE-KG, we have used five graphs offered by the BioGateway knowledge network, which are described in Table 3.

Table 3
Details of the knowledge graphs used in the evaluation.

Graph	Description	Triples
crm	Cis Regulatory Modules (CRM)	1,483,949
crm2tfac	Relations between CRM and transcription factors	12,933
crm2phen	Relations between CRM and phenotypes	45,794
crm2gene	Relations between CRM and target genes	82,491
all	Union of the previous four graphs	1,622,550

4. Results

We have calculated the values of the OQuaRE-KG quality metrics for the five aforementioned graphs and scaled the values into score at the level of metrics, subcharacteristics and characteristics. In this section we mainly focus on the main results at the level of characteristics and subcharacteristics, whereas the complete ones are available at https://github.com/tecnomod-um/oquare-kg, including the raw value of the metrics, and the quality score (scaled value) of the metrics, subcharacteristics and characteristics.

4.1. Characteristics

Figure 2 shows the radar chart for each KG at the level of characteristics. The analysis of the scores shows differences between the scores of graphs for the structural and functional adequacy characteristics, whereas all graphs get the same quality score for compatibility (3), transferability (1), operability (1), reliability (2), and maintainability (1). These characteristics account for aspects which are more related to the metadata and the process followed to generate the graph. For example, operability is linked to the availability of machine-readable and human-readable licenses, and compatibility is liked to the reuse of terms and vocabularies, serialization formats, and validity of the formats. Given that these graphs are part of the same project, the achievement of the same scores means homogeneity in those aspects. Hence, all graphs would have the same pros and cons with respect to those characteristics.

Regarding the structural and functional adequacy characteristics, a common pattern is observed. On the one hand, the *crm* and *all* graphs have the same quality scores. This makes sense since the *crm* graph represents more than 90% of the *all* graph data. On the other hand, the *crm2tfac* and *crm2gene* graphs have the same score for both characteristics, whereas the *crm2phen* graph has the lowest score for functional adequacy. The scores of these three graphs are not related to their size, since *crm2gene* is the largest and *crm2tfac* the smallest. Further information can be obtained by analysing the scores at the level of subcharacteristics.

4.2. Subcharacteristics

In this section we analyse the results of subcharacteristics which exhibit differences between graphs, namely, structural and functional adequacy. Regarding the structural category (see Figure 3), all graphs get the same score for formalisation, structural accuracy, syntactic validity, and redundancy. The differences between graphs are exhibited for consistency and interpretability. Further analysis requires to work at the level of metrics. In this case, one metric is responsible for the differences:

- *Entities with no type metric*: this metric is associated with both consistency and interpretability. It accounts for the ratio of nodes lacking an rdf:type in the graph. The *crm* and *all* graphs have the highest score, whereas the other three graphs exhibit the general pattern observed at the level of characteristics, that is, *crm2tfac* and *crm2gene* have the same score.

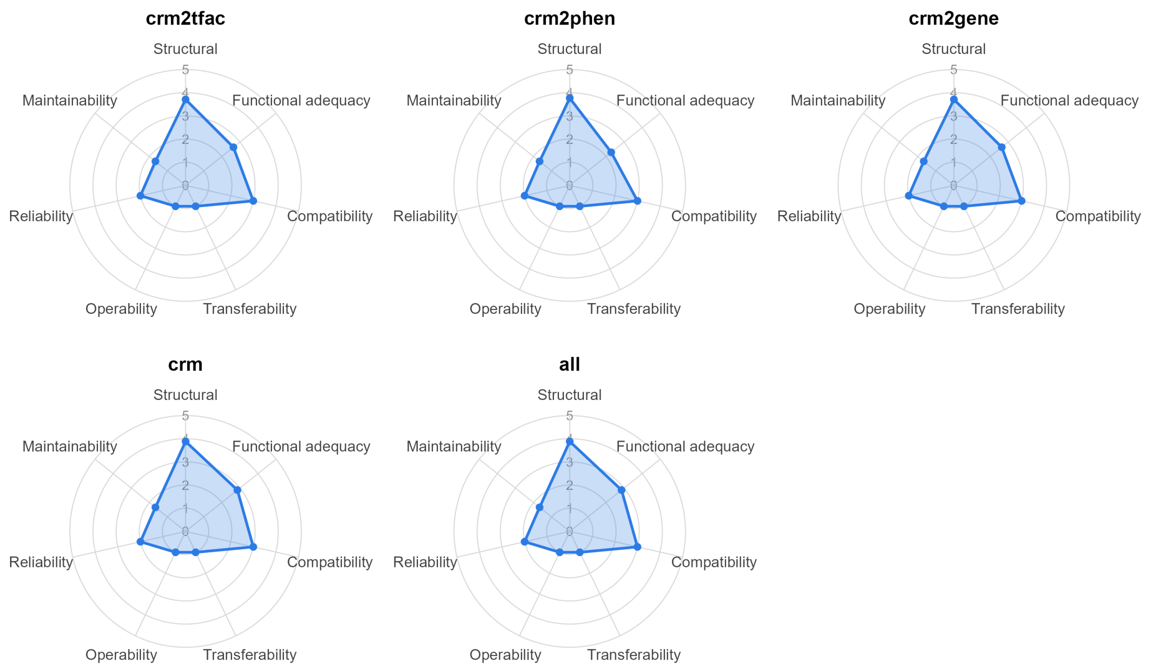

Figure 2: The radar chart at the level of characteristics.

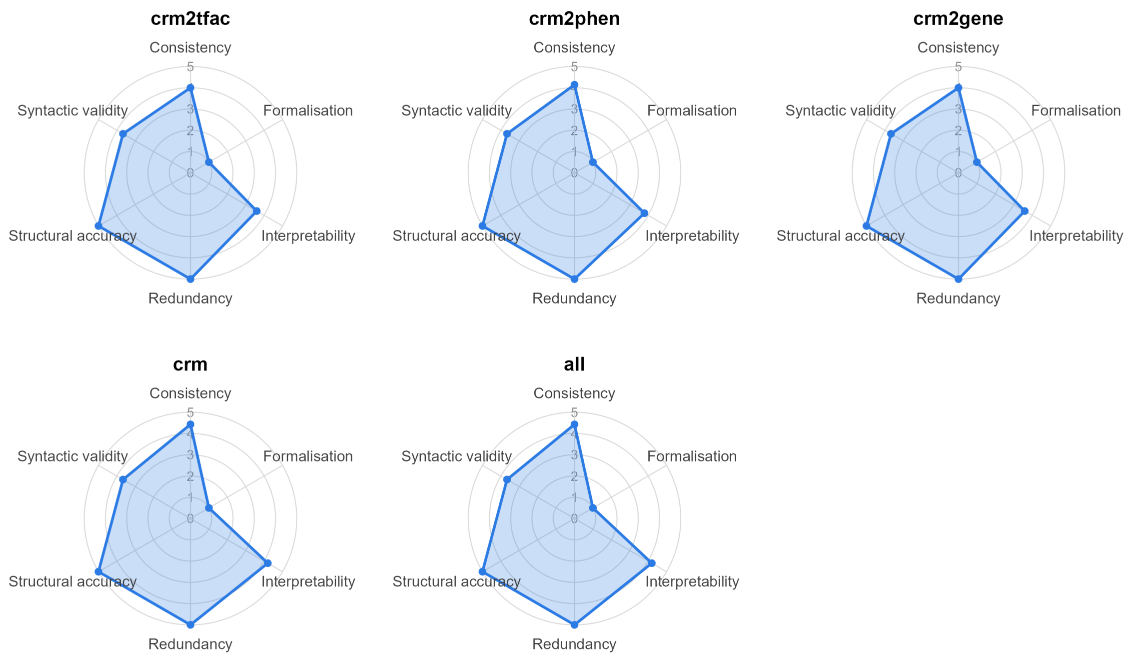

Figure 3: The radar chart for the structural subcharacteristics.

It should be noted that having the same quality score for a metric does not mean having the same metric value. For example, the value of the metric *Entities with no type metric* for *crm* and *all* are 0.0239 and 0.0756 respectively. Since they are in the interval 0-0.20, they are scaled to the same score (5).

Regarding functional adequacy (see Figure 4), the difference in quality scores happen for the following subcharacteristics: inference, trustworthiness and clustering. The pattern is like in the previous cases except for trustworthiness, where all graphs obtain a quality score 3 except for *crm2phen*, which gets a score 1.

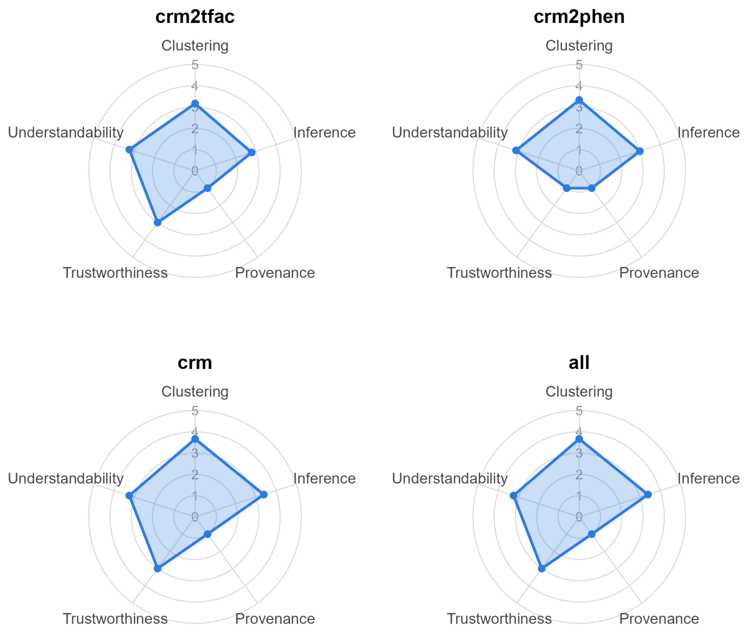

Figure 4: The radar chart for the functional adequacy subcharacteristics.

The variation for inference and clustering is due to the entities with no type metric, which reveals as a distinctive metric for these graphs (see Figures 5, 6). These results are consistent with the fact that these graphs contain untyped entities, because the definition of these entities is carried out in other graphs such as *crm* or *gene*.

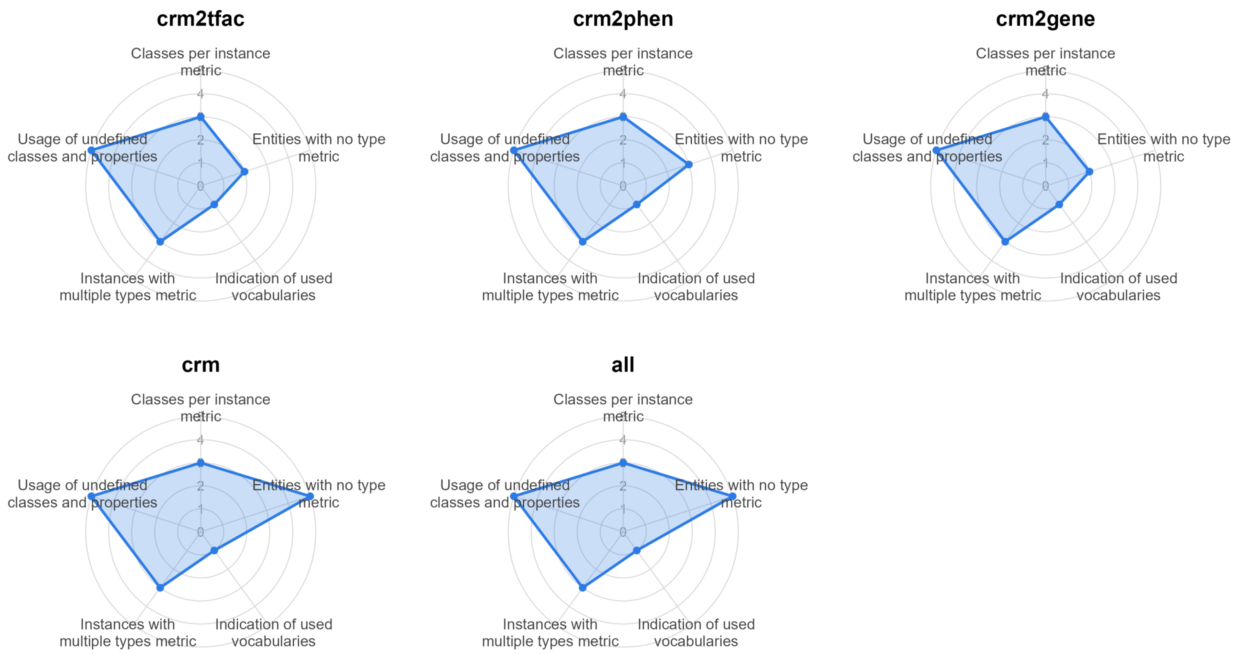

Figure 5: The radar chart for the inference metrics.

The difference in trustworthiness is due to the evidence metric, which verifies whether graph's assertions have terms for capturing evidence. This is calculated by considering a predetermined list of

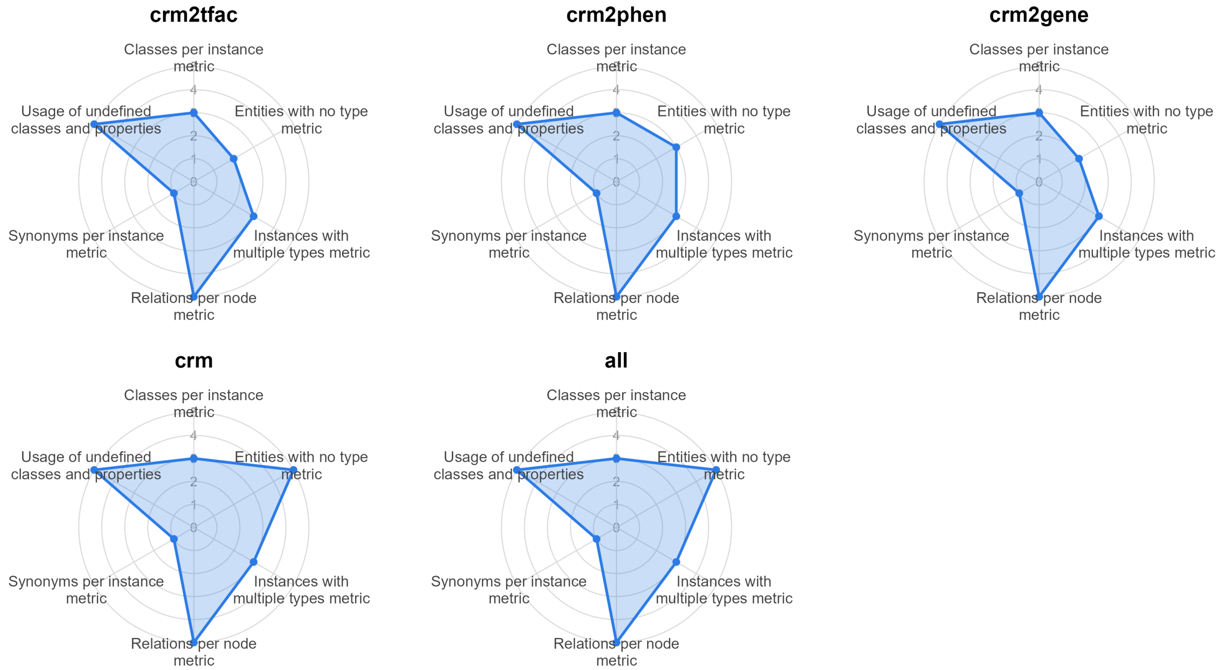

Figure 6: The radar chart for the clustering metrics.

annotation properties, which includes properties from Dublin Core (source, references), RDFs (isDefinedBy), OWL (sameAs), schema.org (evidenceLevel, evidenceOrigin), and from GO-LEGO (evidence, evidence-with).

5. Discussion

According to the results of the experiments, the graphs may be classified in three groups by their quality scores. Group 1 includes *crm* and *all*, which have the same and highest quality scores. Group 2 includes *crm2phen* and Group 3 include *crm2tfac* and *crm2gene*, which also have the same but lowest quality scores.

The analysis at the level of characteristics shows that all the graphs have the same quality score for the characteristics more related to the methodological process, which is an expected and desired result, independently of the concrete score. That means that the homogeneity is captured by those characteristics. The differences are in characteristics more related to the modelling of the data and the functional adequacy, which may vary more between graphs, since it may have also links to the quality of the data itself. If we interpret the quality scores for the characteristics, the graphs obtained scores ≥3 for the structural and compatibility characteristics, whereas the scores are ≤3 for the rest of the characteristics. This information can be used by the developers of the KGs to identify strengths and weaknesses.

The analysis at the level of subcharacteristics has shown that the pattern identified at the level of characteristics. Either all the graphs have the same scores or the same three groups are identified. In the current version of the framework, five characteristics have only one subcharacteristic associated, but the structural characteristic and the functional adequacy ones have 6 and 5 subcharacteristics, respectively. This permits to a more detailed analysis for those characteristics.

Regarding the structural category, for which we have shown the radar chart, the difference in the quality scores for the subcharacteristics is due to aspects related to interpretability and consistency. This can be further inspected at the level of metrics, which reveals that it is due to the ratio of missing type definitions. These findings can be used by the developers of the knowledge graph to review their decisions.

Additional findings come from the analysis of the functional adequacy subcharacteristics. The evidence metric score for *crm2phen* is 1, whereas the score of this metric is 5 for the rest of graphs. This may be due to the nature of the data of this graph. In these graphs, the evidence properties are used for asserting from which database the data comes from and for associating experimental methods as evidence. The latter is not applicable to *crm2phen*, since this information is absent in the original data sources, and this may justify this difference.

These results also underscore the relevance of the approach adopted in this work on KGs, since the schema used for graph modelling can include information that is subsequently present or absent in the instantiated KGs according to the sources used. Therefore, the metrics and results derived from the ontology evaluation are not directly transferred to KG quality. The results also suggest that improving the quality of the graphs would require greater effort in metadata. For example, including licenses and data formats, such as language, would be helpful.

Table 4 summarizes the levels of the quality scores of the metrics for each graph. The graphs obtain a very similar number of metrics with the same quality scores. The same three groups of graphs can be made based on the values observed in the table. At the levels of characteristics and subcharacteristics, *crm2phen* got higher scores than *crm2tfac* and *crm2gene*. However, the analysis of the levels of the metric reveals that *crm2phen* has more metrics with 1 and less with 5, although the difference is small. This is another example of how the framework permits to analyse the KGs at different levels, which provide complementary information.

Table 4
Summary of the quality score levels for each KG, where 1 represents the lowest quality score and 5 the highest. The numbers in each cell represent the number of metrics that achieved the quality score associated with the corresponding column. For example, 15 metrics have a quality score 5 for the graph *crm2tfac*.

Graph	**1**	**2**	**3**	**4**	**5**
crm2tfac	9	1	3	0	15
crm2phen	10	0	4	1	13
crm2gene	9	1	3	0	15
crm	9	0	3	0	16
all	9	0	3	0	16

In the present article, we do not intend to make a strong interpretation of the quality scores, because we have used a simple scaling function to map metrics values into quality scores. This is a limitation of the study. Future work should improve the mapping function. For example, our reference framework for ontology evaluation, OQuaRE, proposed two different scaling functions, one static function based on best practices and another dynamic function based on the distribution of real data [18, 19]. For this purpose, the Evaluome framework developed by our research group will be used [20]. We will also carry out further experiments with different KGs to optimize the framework, including resources such as Wikidata [21] or DBPedia [22]. Research will also be done to extend the framework to include additional aspects that impact on the quality of the KG, such as the quality of the ontologies used or the quality of the data sources employed for the generation of the KG. Finally, we plan to apply OQuaRE-KG to perform a metrics-based comparison of KGs developed by traditional processes against KGs generated by LLMs or automated methods, as done with OQuaRE for ontologies [23].

6. Conclusions

In this work we have presented OQuaRE-KG, a framework for assessing the quality of knowledge graphs that follows principles previously applied for the assessment of the quality of ontologies. The experiments reported in this article show that it is able to detect differences in the quality scores of different knowledge graphs from the gene regulation domain. Further experiments in different domains will help to define levels of quality based on the scores of metrics, subcharacteristics, and characteristics.

7. Acknowledgments

This research is part of the grant PID2024-155257OB-I00 funded by MICIU/AEI/10.13039/501100011033/ and by ERDF,EU (FONDOS FEDER).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] C. Peng, F. Xia, M. Naseriparsa, F. Osborne, Knowledge graphs: Opportunities and challenges, Artificial intelligence review 56 (2023) 13071–13102.
[2] H. Knublauch, D. Kontokostas, Shapes Constraint Language (SHACL), W3C Recommendation, World Wide Web Consortium (W3C), 2017. URL: https://www.w3.org/TR/2017/REC-shacl-20170720/.
[3] E. Prud'hommeaux, J. E. Labra Gayo, H. Solbrig, Shape expressions: an rdf validation and transformation language, in: Proceedings of the 10th International Conference on Semantic Systems, 2014, pp. 32–40.
[4] B. Glimm, I. Horrocks, B. Motik, G. Stoilos, Z. Wang, Hermit: an owl 2 reasoner, Journal of automated reasoning 53 (2014) 245–269.
[5] I. O. for Standardization ISO/IEC 25012: 2008 (E), Software engineering-software product quality requirementsand evaluation (square)-data quality model, 2008.
[6] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9.
[7] A. Duque-Ramos, J. T. Fernández-Breis, R. Stevens, N. Aussenac-Gilles, Oquare: A square-based approach for evaluating the quality of ontologies, Journal of research and practice in information technology 43 (2011) 159–176.
[8] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann, S. Auer, Quality assessment for linked data: A survey: A systematic literature review and conceptual framework, Semantic web 7 (2015) 63–93.
[9] J. Debattista, S. Auer, C. Lange, Luzzu–a framework for linked data quality assessment, in: 2016 IEEE Tenth International Conference on Semantic Computing (ICSC), IEEE, 2016, pp. 124–131.
[10] J. Debattista, C. Lange, S. Auer, D. Cortis, Evaluating the quality of the lod cloud: An empirical investigation, Semantic Web 9 (2018) 859–901.
[11] M. Färber, F. Bartscherer, C. Menne, A. Rettinger, Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago, Semantic Web 9 (2018) 77–129.
[12] H. Chen, G. Cao, J. Chen, J. Ding, A practical framework for evaluating the quality of knowledge graph, in: Knowledge Graph and Semantic Computing: Knowledge Computing and Language Understanding: 4th China Conference, CCKS 2019, Hangzhou, China, August 24–27, 2019, Revised Selected Papers 4, Springer, 2019, pp. 111–122.
[13] M. D. Wilkinson, M. Dumontier, S.-A. Sansone, L. O. Bonino da Silva Santos, M. Prieto, D. Batista, P. McQuilton, T. Kuhn, P. Rocca-Serra, M. Crosas, et al., Evaluating fair maturity through a scalable, automated, community-governed framework, Scientific data 6 (2019) 174.
[14] A. Gaignard, T. Rosnet, F. De Lamotte, V. Lefort, M.-D. Devignes, Fair-checker: supporting digital resource findability and reuse with knowledge graphs and semantic web standards, Journal of Biomedical Semantics 14 (2023) 7.
[15] J. Lehmann, A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, S. Auer, Quality assessment for linked data: A survey., Semantic Web (1570-0844) 7 (2016).
[16] J. Mulero-Hernández, G. Almagro-Hernández, J. T. Fernández-Breis, Graph alignment methods

for the development of interoperable gene regulation knowledge graphs, IEEE Access 14 (2025) 792–814.

[17] J. Mulero-Hernández, V. Mironov, J. A. Miñarro-Giménez, M. Kuiper, J. T. Fernández-Breis, Integration of chromosome locations and functional aspects of enhancers and topologically associating domains in knowledge graphs enables versatile queries about gene regulation, Nucleic Acids Research 52 (2024) e69–e69.

[18] A. Duque-Ramos, M. Quesada-Martínez, M. Iniesta-Moreno, J. T. Fernández-Breis, R. Stevens, Supporting the analysis of ontology evolution processes through the combination of static and dynamic scaling functions in oquare, Journal of biomedical semantics 7 (2016) 63.

[19] M. Franco, J. M. Vivo, M. Quesada-Martínez, A. Duque-Ramos, J. T. Fernández-Breis, Evaluation of ontology structural metrics based on public repository data, Briefings in bioinformatics 21 (2020) 473–485.

[20] J. A. Bernabé-Díaz, M. Franco, J.-M. Vivo, J. T. Fernández-Breis, Optimizing clustering-based analytical methods with trimmed and sparse clustering, Computers in Biology and Medicine 194 (2025) 110436.

[21] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Communications of the ACM 57 (2014) 78–85.

[22] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, Z. Ives, Dbpedia: A nucleus for a web of open data, in: international semantic web conference, Springer, 2007, pp. 722–735.

[23] M. Val-Calvo, M. E. Aranguren, J. Mulero-Hernández, G. Almagro-Hernández, P. Deshmukh, J. A. Bernabé-Díaz, P. Espinoza-Arias, J. L. Sánchez-Fernández, J. Mueller, J. T. Fernández-Breis, Ontogenix: Leveraging large language models for enhanced ontology engineering from datasets, Information Processing & Management 62 (2025) 104042.