Bringing Data Quality to ODRL: From Policy to Validation

Achraf Hmimou[1], Petar Jovanovic[1], Cristina Gómez[1],
Jose-Norberto Mazón[2], Claudia P. Ayala[1], Oscar Romero[1*]
[1]Universitat Politècnica de Catalunya, Barcelona, Spain [2]Universidad de Alicante, Alicante, Spain

ABSTRACT

Data quality is a governance pillar for trustworthy data sharing in Data Spaces. However, there is an absence of a standardized vocabulary capable of enforcing Data Quality Requirements (DQRs) as machine-readable, binding contracts. Despite ODRL being the predominant policy language in data spaces, we identify structural barriers in its information model that prevents expressing even simple DQRs.

This paper proposes that data quality, like access and usage, should be a core governance aspect and should be also operationalized throughout the policies and contracts. In these regard, we expose some ODRL barrier examples, present a minimal ODRL profile to overcome these, and discuss our current efforts toward an end-to-end operationalization.

KEYWORDS

Data Spaces, Data Quality, Policy Languages, Semantic Web

1 MOTIVATION

The Open Digital Rights Language (ODRL)[1] has become the de facto standard policy language for Data Spaces. Major initiatives, including the IDSA, GAIA-X, and the Data Spaces Support Centre (DSSC) blueprint, mandate ODRL as the interoperable layer for expressing access and usage rights [2, 5]. While recent work focuses on extending ODRL with custom profiles and evaluators to support complex data sharing policies [3, 4], data quality governance remains absent from the standard.

Existing vocabularies, notably W3C DQV[2], describe a dataset's quality (`dqv:QualityPolicy`) retrospectively, but provide no contractual semantics: they cannot negotiate minimum quality thresholds or mandate conformance to a reference standard.

We argue that data quality in data spaces must be governed at the same policy layer as access or usage control. A consumer's willingness to use a data asset depends not only on the right to access it, but also on whether it meets a minimum completeness metric threshold. We believe that it should be contractual commitment, rather than treating these metrics as informal metadata or backend implementation details.

In this direction, this paper presents example structural barriers in ODRL that currently prevent the expression of even simple DQRs. To resolve this, we introduce a minimum data quality validation profile (DQVP) that extends ODRL with validation semantics, grounded in DQ checks catalogued by Papastergios et al. [8] across ISO/IEC 25012[3] quality characteristics, and demonstrate the profile through an illustrative example. To conclude, we further extend our position to the community with a discussion on how we envision an end-to-end operationalisation framework: from a DQR pattern catalogue, to DQVP policies which can be further compiled into data validation code [5].

2 STATE OF THE ART AND CHALLENGES

Current practice confirms an open challenge between DQRs and operational enforcement. Existing frameworks are designed largely for centralized environments and often rely on proprietary quality labels rather than shared, interoperable vocabularies [1, 6]. Despite initiatives like DSSC and SIMPL identifying data quality as a governance pillar, the translation from high-level requirements to automated validation remains manual and fragmented [2].

To evidence this gap, consider this concrete example policy encompassing DQRs for completeness, compliance, and consistency (ISO/IEC 25012) from an example health use case:

"The attribute `ab:nationalHealthcarePatientID` must be non-null in at least 99% of records; `ab:countryOfAffiliation` must conform to the ISO 3166 reference standard in at least 80% of the records; and where `ab:gender` is 'male', `ab:pregnancyHistory` must be completely null."

A data space participant attempting to express these rules over a health dataset natively in ODRL faces an immediate triage of modeling barriers:

- **B1: Column Identity:** In ODRL, `odrl:target` binds to the macroscopic asset (e.g., a `dcat:Dataset`). The `odrl:Constraint` shape lacks a sub-asset slot to target a specific column, which hinders a validator from identifying which column to check without additional context.
- **B2: The Multi-Parameter Arity:** Validating the country code requires two distinct parameters: a reference standard (ISO 3166) and a numeric compliance threshold. ODRL provides only a single `odrl:rightOperand` slot. Overloading this slot with both values would require a custom parsing layer outside the ODRL specification.
- **B3: Conditional Logic:** The conditional check (*"where gender='male'..."*) requires an intra-constraint row filter. ODRL's logical operators (`odrl:and`, `odrl:or`) only compose whole, independent `Constraint` instances at the rule level. There is no native mechanism

*Corresponding author: Achraf Hmimou (achraf.hmimou@upc.edu), LinkedIn: https://www.linkedin.com/in/achraf-hmimou
[1]https://www.w3.org/TR/odrl-model/
[2]https://www.w3.org/TR/vocab-dqv/
[3]https://www.iso.org/standard/35736.html

to assemble a sub-predicate *inside* a single atomic constraint.

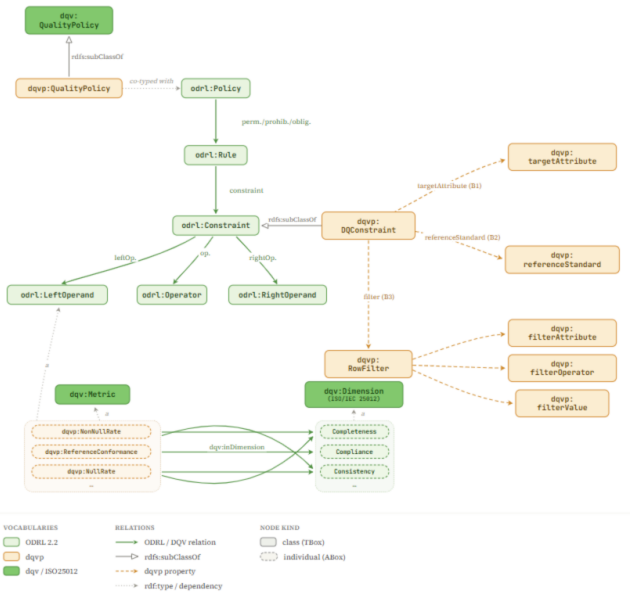
3 TOWARDS A MINIMUM DATA QUALITY VALIDATION PROFILE

To address these barriers, we introduce a minimal, backward-compatible data quality validation profile. Figure 1a presents the ODRL policy model with DQVP additive terms; dashed orange arrows denote new DQVP properties, hollow triangles denote `rdfs:subClassOf`, and solid green arrows denote standard ODRL relations. `dqvp:QualityPolicy` specialises `dqv:QualityPolicy` as a cross-cutting categorisation, co-typed with an ODRL policy subclass (Offer, Agreement, Set) on the same resource; it is intentionally not declared `rdfs:subClassOf odrl:Policy` because ODRL 2.2's policy subclasses are mutually disjoint. `dqvp:DQConstraint` sub-classes [4] `odrl:Constraint` and reuses its native `leftOperand`/`operator`/`rightOperand` triple, the `leftOperand` being a quality metric (an instance of `dqv:Metric` grounded in ISO/IEC 25012 dimensions) and the `rightOperand` the threshold on the measured value. Extneding this, the profile adds exactly three constructs, one per barrier, each generalising a pattern that is recurrent across open source data quality tools: `dqvp:targetAttribute` names the column (B1); `dqvp:referenceStandard` names the external standard a check is validated against (B2), and a `dqvp:RowFilter` operation, which expresses an intra-constraint row predicate (B3). Multiple `DQConstraints` attached to the same rule are conjunctive per ODRL 2.2 semantics: the policy is satisfied only if an evaluator finds every constraint satisfied.

3.1 Example

Figure 1b presents the encoding for the three DQRs from Section 2 as a single typed as `odrl:Agreement` for a "*patient-summary*" data asset. The governance authority (`odrl:assigner`) permits the data provider (`odrl:assignee`) to distribute the asset *under* the stated quality conditions: the DQRs are `odrl:constraints` on an `odrl:Permission`, so the asset may be distributed only if it meets them. Typing the policy as `odrl:Offer` keeps these terms negotiable (i.e., a consumer can counter-propose and both parties can ratify a binding `odrl:Agreement` [2]). Importantly, the policy states what should be guaranteed, while its evaluation should be the job of the governance authority or the system (the operationalisation pipeline of Section 4), not of the provider. The three checks attach to the same `odrl:Permission` as conjunctive `odrl:constraint` values, and each is typed `a dqvp:DQConstraint`: each carries its column identity via `dqvp:targetAttribute`, resolving B1; the second additionally carries `dqvp:referenceStandard` pointing at the ISO 3166 IRI, resolving B2; and the third expresses the conditional row filter via `dqvp:filter` / `dqvp:RowFilter`,

[4] https://www.w3.org/community/reports/odrl/CG-FINAL-profile-bp-20240808.html



(a) DQVP TBox (rounded rectangles), shown with illustrative ABox individuals (dashed pills). Green nodes are ODRL 2.2 / DQV terms; orange nodes are DQVP extensions.

```
@prefix odrl: <http://www.w3.org/ns/odrl/2/> .
@prefix dqvp: <https://feed-upc.github.io/DQVP-position/dqvp.ttl#> .
@prefix xsd:  <http://www.w3.org/2001/XMLSchema#> .
@prefix ab:   <https://w3id.org/ehds/patsum/attr/> .
@prefix dp:   <https://w3id.org/ehds/patsum/asset/> .
@prefix ex:   <https://example.org/ehds/> .

ex:DataProvider   a odrl:Party .
dp:PatientSummary a odrl:Asset .

ex:dqr-Agreement01
    a dqvp:QualityPolicy, odrl:Agreement ;
    odrl:uid      ex:ps-dq-offer-01 ;
    odrl:profile  <https://feed-upc.github.io/DQVP-position/dqvp.ttl> ;
    odrl:assigner ex:GovernanceAuthority ;
    odrl:assignee ex:DataProvider ;
    odrl:permission [
        a            odrl:Permission ;
        odrl:target  dp:PatientSummary ;
        odrl:action  odrl:distribute  ;
        odrl:constraint
          [ a dqvp:DQConstraint ;
            odrl:leftOperand     dqvp:NonNullRate ;
            dqvp:targetAttribute ab:nationalHealthcarePatientID ;
            odrl:operator        odrl:gteq ;
            odrl:rightOperand    "99"^^xsd:decimal ] ,
          [ a dqvp:DQConstraint ;
            odrl:leftOperand       dqvp:ReferenceConformance ;
            dqvp:targetAttribute   ab:countryOfAffiliation ;
            dqvp:referenceStandard <http://publications.europa.eu/
      resource/authority/country> ;
            odrl:operator          odrl:gteq ;
            odrl:rightOperand      "80"^^xsd:decimal ] ,
          [ a dqvp:DQConstraint ;
            odrl:leftOperand     dqvp:NullRate ;
            dqvp:targetAttribute ab:pregnancyHistory ;
            dqvp:filter [ a dqvp:RowFilter ;
                          dqvp:filterAttribute ab:gender ;
                          dqvp:filterOperator  odrl:eq ;
                          dqvp:filterValue     "male" ] ;
            odrl:operator        odrl:eq ;
            odrl:rightOperand    "100"^^xsd:decimal ]
    ] .
```

(b) A `dqvp:QualityPolicy` typed as `odrl:Agreement`: the provider disributes the patient-summary asset *under* the data-quality conditions, expressed as `odrl:constraints` on the permission.

resolving B3. This example, alongside its operationalisation to data validation checks, can be found in our repository[5].

4 DISCUSSION

The proposed DQVP we introduced above is a minimal baseline based on the set of presented example DQRs. Extending it into a working profile, and scaling it for a real data space deployment, raises questions that go well beyond vocabulary design. The primary open research problem lies in standardizing the end-to-end operationalization pipeline: enabling a paradigm where high-level, natural language DQRs are deterministically translated into declarative DQVP policies, and subsequently compiled into executable validation code. To open the discussion, we share our following efforts with the community.

(Q1) Standardisation path for the profile. We propose that a DQVP-like profile be advanced as a W3C ODRL Community Group Report, given their mandate over the quality governance pillar [2]. We also propose initiating a collaborative discussion to systematize the profile's design. This dialogue will center on example open questions: How can new constraint-grouping mechanisms be integrated? Furthermore, given the highly fragmented landscape of existing data quality tools [8], what is the optimal strategy to ground the profile's semantics to ensure interoperability?

(Q2) From DQR to ODRL. Encoding a DQVP policy by hand is feasible for a handful of constraints, but a data space could include hundreds of DQRs at a time. In our ongoing work, we operationalize elicitation of DQRs by instantiating reusable patterns from a DQR catalogue, and a translation step then compiles each pattern into ODRL. For this stage to be interoperable, the community needs a normative pattern vocabulary and a normative mapping from each pattern to a constraint shape over a DQVP-like profile. Should pattern catalogues be standardised as W3C artefacts, or governed sector-by-sector (health, finance, logistics)? How LLMs can be used to further automate this step? (e.g., [7])

(Q3) From ODRL to data validation code. In [5] (Figure 2), we proposed a workflow that consumes an ODRL policy and generates a *policy checker* (i.e., data validation plan) that can be further mapped to technology-specific validation code (e.g., Great Expectations [8]) via automated code-generation. While existing W3C data validation languages like SHACL provide rich semantics, they natively assume the underlying data is represented as an RDF graph. Modern data spaces, however, routinely exchange heterogeneous data formats (e.g., CSV, Parquet, JSON). Forcing participants to materialize massive tabular datasets into RDF solely to evaluate a quality constraint introduces severe computational bottlenecks. This raises an open question for the community: should the operationalisation of ODRL quality profiles rely exclusively on SHACL-like validators, or must we standardise translation mappings towards native data engineering engines?

[5] https://github.com/feed-upc/DQVP-position

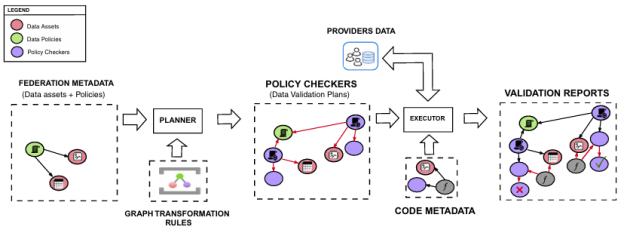


Figure 2: Data Validation Workflow [5]

To address this, we advocate for knowledge graphs as the core backbone for policy orchestration. This approach enables the expression of DQR patterns, ODRL policies, and validator metadata within a single graph, utilizing rule-based reasoning to connect them while preserving provenance metadata. The workflow established in [5] represents one concrete realisation. As we continue to iterate on this framework, we offer it as a starting point for the wider community, aiming to foster discussion and convergence toward a shared operationalisation stack rather than fragmented, deployment-specific solutions.

5 CONCLUSION

To operationalise data governance effectively, data quality should be negotiated as a primary concern alongside access and usage policies. Currently, the lack of standardised contractual vocabulary forces implementers to rely on ad hoc, case-specific validation solutions. The DQVP provides a structured alternative, overcoming existing barriers within the ODRL information model to introduce data quality semantics. We propose this profile to invite standardisation stakeholders into a broader discussion. It serves as a starting point to express DQRs within ODRL and operationalize its validation.

ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministerio de Ciencia e Innovación under project PID2024-161707OB-I00 / AEI /10.13039/501100011033 (FeeD).

REFERENCES

[1] Marcel Altendeitering, Tobias Moritz Guggenberger, and Frederik Möller. 2024. A design theory for data quality tools in data ecosystems: Findings from three industry cases. *Data & Knowledge Engineering* 153 (2024), 102333. doi:10.1016/j.datak.2024.102333

[2] Claudia P Ayala, Besim Bilalli, Cristina Gómez, Jose-Norberto Mazón, Oscar Romero, et al. 2025. Challenges to enforce data quality in data spaces. *DOLAP 2025* (2025), 85–87.

[3] Andrea Cimmino, Juan Cano-Benito, and Raúl García-Castro. 2025. Open Digital Rights Enforcement framework (ODRE): From descriptive to enforceable policies. *Computers & Security* 150 (2025), 104282.

[4] Delaram Golpayegani, Beatriz Esteves, Harshvardhan J. Pandit, and Dave Lewis. 2024. AIUP: an ODRL Profile for Expressing AI Use Policies to Support the EU AI Act. In *Proceedings of the 20th International Conference on Semantic Systems (SEMANTiCS 2024)*. IOS Press, Amsterdam, Netherlands, 1–6.

[5] Achraf Hmimou, Petar Jovanovic, Sergi Nadal, Oscar Romero, and Anna Queralt. 2025. Operationalizing and automating data validation in data spaces. *Data Science and Engineering* (2025), 1–21. doi:10.1007/s41019-025-00317-7

[6] Mario Köbis-Riedel, Marcel Altendeitering, and Christian Beecks. 2025. Data Quality Scoring: A Conceptual Model and Prototypical Implementation. In *Proceedings of the 14th International Conference on Data Science, Technology and Applications (DATA 2025)*. SCITEPRESS, Setúbal, Portugal, 329–338. doi:10.5220/0013461300003967

[7] Daham M Mustafa, Abhishek Nadgeri, Diego Collarana, Benedikt T Arnold, Christoph Quix, Christoph Lange, and Stefan Decker. 2025. From instructions to ODRL usage policies: An ontology guided approach. *arXiv preprint arXiv:2506.03301* (2025).

[8] Vasileios Papastergios, Lisa Ehrlinger, and Anastasios Gounaris. 2026. Unfolding data quality dimensions in practice: A survey. *ACM Journal of Data and Information Quality* 18, 1 (2026), 1–37. doi:10.1145/3786328