MQIO Design Document

This document provides information about the design and development of the **M**apping **Q**uality **I**mprovement **O**ntology (**MQIO**).

1. Tasks

You are asked to complete the following tasks:

1. Review the feedback questionnaire which will provide an indication of the requested feedback: https://forms.gle/8DdsSXKJz1eGWhWw6
2. Review this document while considering the design methodology followed by MQIO in comparison to the state of the art.
3. Please complete the feedback questionnaire after you have reviewed the document.

Contact Alex Randles (alex.randles@adaptcentre) if you have any questions.

2. Design Methodology

The design of the MQIO followed best practices as recommended by the semantic web community. Ontology design practices were reused from the most prominent ontology design methodologies. The methodologies included the NeON methodology [28], UPON Lite [20], Ontology development 101: A guide to creating your first ontology [21] and LOT: An industrial oriented ontology engineering framework [23].

1. **Identification of aims, objectives, scope:** The design process commenced with the identification of the aims, objectives and scope of the ontology, which are outlined in Table 1 of this document. The template used for the table was retrieved from the methodologies and used to define the ontology requirements specification document. The document outlines requirements and among other things, the aims, objectives and scope of the ontology.
2. **Identify and analyze relevant information:** A review of publications in the state of the art was conducted to identify relevant information. Publications within the state of the art which related to topics within the defined scope were reviewed to facilitate the retrieval of relevant information. Thereafter, the retrieved information was used to formalize competency questions. Table 2 includes references to publications which inspired the creation of each competency question.
3. **Create Use-cases and Competency questions:** Competency questions were created during the design process of the ontology. The questions define the functional requirements of the ontology and were iteratively refined until an accurate representation of the requirements and objectives was conceived. The final iteration of the questions is

shown in Table 2. Use cases were devised in order to refine the requirements of the ontology. The use cases involved projects which uplifted geospatial data (data.geohive.ie) [25] and network monitoring data (Ericsson) [26]. A use case graph generated is available (https://tinyurl.com/2ks5urb9).

4. **Identify Concepts and Relationships:** Concepts and relationships were identified through the state of the art review and the researchers previous experience in the creation of linked data (LD). The concepts and relationships were iteratively defined until the information modeling provided by the ontology satisfied each of the competency questions. In addition, concepts and relationships were reused from existing vocabularies as recommended by the methodologies and the W3C recommendation on Data on the Web Best Practices [16]. **Reused ontologies** included the PROV Ontology (PROV-O) [16] was reused and extended to capture provenance related to mapping information. The Data Quality Vocabulary (DQV) [1] was reused to represent quality metrics utilized during the mapping assessment and validation phase. The reuse in MQIO is demonstrated in the competency questions and ontology documentation.
5. **Progressive iterations:** Steps 2-4 were iteratively repeated until the point when the proposed concepts/relationships provided information which satisfied each requirement defined in the form of a competency question.
6. **Create Ontology:** The ontology was implemented in OWL 2 Web Ontology Language [17]. Concepts and relationships which were defined in the previous step were constructed using Protégé ontology development tool [30]. Furthermore, semantic reasoners were also utilized to detect logical inconsistencies within the ontology.
7. **Evaluate:** The ontology was evaluated with respect of the ability for the defined concepts and relationships to fulfill each competency question. The usage of a semantic reasoner within Protege ensured logical inconsistencies were identified and removed. OOPS! Pitfall Scanner [24] was used to detect common ontology design issues. The quality of metadata and documentation was evaluated through presentation within peer reviewed publications. Feedback received from reviewers allowed us to identify areas for improvement. Peer reviewed publications related to MQIO are outlined in Section 5.
8. **Publication:** Ontology documentation (https://w3id.org/MQIO) was created using WIDOCO [10] which is a tool designed to use ontology metadata to create HTML documents which include descriptions of the classes and properties. Thereafter, the ontology and human readable documentation were published with a permanent identifier as a FAIR resource including an open and permissive license. The documentation contains information about the creation, design, usage, class interaction diagrams and provides various serializations.

3. Background

The following section provides information related to the requirements, design and purpose of MQIO.

3.1 Description

MQIO provides an ontology for expressing information relating to the quality assessment, refinement and validation of declarative mapping definitions. The objective is to make this information easier to publish, exchange and consume, thus improving the overall quality of the resulting LD datasets which are created by these mappings. Furthermore, providing data quality information to the users will allow them to assess the suitability of the mapping for their application. The ontology was designed to resolve the gap in the state of the art in relation to an ontology which represents quality assessment, refinement and validation information of LD mappings.

3.2 Requirements

The development of the ontology follows best practices in ontology development methodologies, such as those mentioned. Creating a specification for the ontology provided additional guidance during the development phase. The requirements have been derived from state of the art review and application of the ontology within a framework and use cases. **Table 1** shows the requirements document for MQIO

Table 1: Ontology requirements specification document [20]

Ontology Requirements Specification Document	
1	**Purpose**
	Capture information related to the quality assessment, refinement and validation of mappings used to generate, relate or interlink RDF datasets. Capturing such information is expected to positively impact the quality of the mappings and datasets as well as improve the reuse and maintenance of those mappings.
2	**Scope**
	In scope: • Mapping quality • Mapping agents • Mapping creation • Mapping validation **Out of scope:** • Source data of the mappings • Resulting dataset

3	**Implementation Language (optional)**
	OWL 2 Web Ontology Language
4	**Intended End-Users (optional)**
	Agents involved in the quality assessment, refinement and validation of LD mappings.
5	**Intended Uses**
	Capturing metadata and provenance relating to the quality assessment, refinement and validation of LD mappings. This metadata also allows for the datasets involved to be assessed in terms of its quality.
6	**Ontology Requirements**
	1. Non-Functional Requirements
	Allow the users of the ontology to define and validate quality requirements related to mappings and capture metadata and provenance related to the quality assessment, refinement and validation of LD mappings.
	2. Functional Requirements: Lists or tables of requirements written as Competency Questions and sentences
	Competency questions in Section 4 (next section).
7	**Pre-Glossary of Terms (optional)**
	1. Terms from Competency Questions
	Competency questions in Section 4.
	2. Terms from Answers

	Competency questions in Section 4.
	3. Objects
	Competency questions in Section 4.

4. Competency Questions

Ontology Competency questions define design requirements in natural language form. These questions state information which should be provided by the ontology. The fulfillment of the questions is accomplished by providing a concept/relationship which represents the required information. Most questions were inspired by literature discovered in the state of the art review. However, certain questions were defined through application to use cases (**DTA**) and feedback from experts (**DTF**).**Table *2*** shows the final iteration of competency questions created for MQIO.

The **answer** to each question is structured as <Subject, Relationship, Concept> which represent an RDF triple. A description of each concept and relationship used is available[1]

Table 2: MQIO Competency Questions

#	Question	Relationship	Concept	References
Subject: `mqio:MappingArtefact` A mapping artefact contains rules which link or create linked data datasets.				
1	Who created the mapping?	`mqio:wasCreatedBy`	`prov:Agent` `mqio:MappingRefinement`	[2,3,9,12,18]
2	What was the rationale for creating the mapping?	`mqio:hasPurpose`	`xsd:string`	[4,9,18]
3	What instruments were utilized to define the mapping?	`mqio:usedTool`	`xsd:string`	[4,9,22,27]

[1] https://drive.google.com/file/d/1fWzhZr7UDCXm86Zo9qpHC8egu76_ZypC/view?usp=share_link

4	When was the mapping defined?	prov:generatedAtTime	xsd:dateTime	[9,15, 18]
Subject: mqio:MappingAssessment An activity in which the quality of a mapping document is assessed, generating information on quality issues within the mapping.				
5	Who performed the quality assessment of the mapping?	prov:wasAssociatedWith	prov:Agent	[8,9,15]
6	What mapping is associated with the assessment?	mqio:assessedMapping	mqio:MappingArtefact	[5,8,9,11]
7	What quality metrics were executed during the assessment process?	mqio:wasExecuted	dqv:Metric	[1,8,11–13]
8	What quality measurements resulted from the assessment?	prov:generated	dqv:QualityMeasurement	[1,6]
9	What quality issues were detected?	mqio:hasValidationReport	mqio:MappingValidationReport	[8,11, 12,14, 19]
10	What value is associated with the violation?	mqio:hasObjectValue, mqio:hasLiteralValue	rdfs:Resource, xsd:string	[8,11, 12,14, 19]
11	How are the quality issues described?	mqio:hasResultMessage	mqio:MappingViolation	[8,11, 12,14, 19]
12	When were the quality issues detected?	prov:endedAtTime	xsd:dateTime	[8,11, 12,19]
13	Where can provenance on the issues be accessed?	mqio:hasViolation	mqio:MappingValidationReport	[14,15]

14	What quality metrics were associated with the detected quality issues?	mqio:isDescribedBy	dqv:Metric	[1,7,11,12,19]
15	What quality dimensions represent the metrics?	dqv:inDimension	dqv:Dimension	[1,7,11,12,19]
16	What quality categories represent the dimensions?	dqv:inCategory	dqv:Category	[1,7,11,12,19]
Subject: mqio:MappingRefinement An activity which involves removing quality violations contained within a mapping document.				
17	Who performed the quality refinement of the mapping?	prov:wasAssociatedWith	prov:Agent	[9,15]
18	When was the refinement process completed?	prov:endedAtTime	xsd:dateTime	[8,9,15]
19	What queries are associated with refinements?	mqio:usedQuery	xsd:string	DTA
20	What confidence score did the refinements have?	mqio:hasConfidenceScore	xsd:double	DTA
21	What violations have been refined?	mqio:wasRefinedBy	mqio:Violation	DTA
22	What quality requirements are associated with the mapping?	mqio:hasQualityRequirement	mqio:QualityRequirement	[1,12,19]

Subject: mqio:QualityRequirement A quality requirement is a requirement a mapping should satisfy.				
23	What quality measurements were associated with the requirements?	mqio:hasQualityMeasurement	dqv:QualityMeasurement	[1,6]
24	Are the requirements satisfied?	mqio:isSatisfied	xsd:boolean	[1,6]
Subject: dqv:Metric Represents a standard to measure a quality dimension.				
25	What refinements are associated with the quality metrics?	mqio:hasRefinement	mqio:MappingRefinement	[8,11, 19]

SPARQL query answers to the competency questions are available[2]. Further information on the graph used to execute the queries can be found in the “Description” section of the ontology documentation.

5. Publications

The following peer reviewed publications related to the design and usage of MQIO[3].

1) *Randles, A., Junior, A.C. and O'Sullivan, D., 2020. Towards a vocabulary for mapping quality assessment. In OM@ ISWC (pp. 241-242).*

In this publication we presented a brief overview of the design of MQIO.

2) *Randles, A., Junior, A.C. and O'Sullivan, D., 2021, January. A vocabulary for describing mapping quality assessment, refinement and validation. In 2021 IEEE 15th International Conference on Semantic Computing (ICSC) (pp. 425-430). IEEE*

In this publication we presented a detailed description of the design process followed by MQIO, use case of the ontology and reuse of existing vocabularies. Furthermore, we discussed an

[2] https://drive.google.com/file/d/1zsgX66NokGc3mxXuaRAL3S3bVmkrhMlU/view?usp=share_link
[3] As a note the ontology was previously called the Mapping Quality Vocabulary (MQV) in the publications

application of the ontology within a demonstration walkthrough. Finally, we mentioned related provenance and metadata models.

3) *Randles, A. and O'Sullivan, D., Assessing Quality of R2RML Mappings for OSi's Linked Open Data Portal.*

In this publication we presented an overview of the MQI framework applied to geospatial R2RML mappings within a current research project. The reports generated during the application were expressed in MQIO.

4) *Randles, A., O'Sullivan, D., Keeney, J. and Fallon, L., Applying a Mapping Quality Framework in Cloud Native Monitoring.*

In this publication we presented an overview of the MQI framework applied to mappings designed to uplift time series metric data utilized within cloud native monitoring. The reports generated during the application were expressed in MQIO.

5) *Randles, A. and O'Sullivan, D., 2022. Evaluating Quality Improvement Techniques Within the Linked Data Generation Process. In Towards a Knowledge-Aware AI (pp. 21-35). IOS Press.*

In this publication we presented a detail description of the framework and the usability evaluation which was conducted on the MQI framework. Furthermore, we discuss realizations of the results and outline respective improvements. The reports generated during the application were expressed in MQIO.

Bibliography

[1] Riccardo Albertoni and Antoine Isaac. 2016. Data on the web best practices: Data quality vocabulary. *W3C Work. Draft* 19, (2016). Retrieved from https://www.w3.org/TR/vocab-dqv/

[2] Ademar Crotti, Christophe Debruyne, Rob Brennan, and Declan O'Sullivan. 2017. An evaluation of uplift mapping languages. *Int. J. Web Inf. Syst.* 13, 4 (2017), 405–424. DOI:https://doi.org/10.1108/IJWIS-04-2017-0036

[3] Ademar Crotti, Christophe Debruyne, and Declan O'Sullivan. 2018. Juma Uplift: Using a Block Metaphor for Representing Uplift Mappings. In *Proceedings - 12th IEEE International Conference on Semantic Computing, ICSC 2018*, Institute of Electrical and Electronics Engineers Inc., 211–218. DOI:https://doi.org/10.1109/ICSC.2018.00037

[4] Ademar Crotti Junior, Christophe Debruyne, and Declan O'Sullivan. 2017. Juma: An Editor that Uses a Block Metaphor to Facilitate the Creation and Editing of R2RML Mappings. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 87–92. DOI:https://doi.org/10.1007/978-3-319-70407-4_17

[5] Souripriya Das, Seema Sundara, and Richard Cyganiak. 2012. R2RML: RDB to RDF Mapping Language. *W3C Recomm.* (2012). DOI:https://doi.org/10.1017/CBO9781107415324.004

[6] Jeremy Debattista, Christoph Lange, and Sören Auer. 2014. Daq, an ontology for dataset

quality information. In *CEUR Workshop Proceedings*.

[7] Jeremy Debattista, Christoph Lange, Sören Auer, and Dominic Cortis. 2018. Evaluating the quality of the LOD cloud: An empirical investigation. *Semant. Web* 9, (March 2018), 1–43. DOI:https://doi.org/10.3233/SW-180306

[8] Anastasia Dimou, Dimitris Kontokostas, Markus Freudenberg, Ruben Verborgh, Jens Lehmann, Erik Mannens, Sebastian Hellmann, and Rik Van de Walle. 2015. Assessing and refining mappings to RDF to improve dataset quality. In *Lecture Notes in Computer Science*, Springer Verlag, 133–149. DOI:https://doi.org/10.1007/978-3-319-25010-6_8

[9] Anastasia Dimou, Tom De Nies, Ruben Verborgh, Erik Mannens, and Rik de Walle. 2016. Automated Metadata Generation for Linked Data Generation and Publishing Workflows. In *Proceedings of the 9th Workshop on Linked Data on the Web* (CEUR Workshop Proceedings).

[10] Daniel Garijo. 2017. WIDOCO: a wizard for documenting ontologies. In *International Semantic Web Conference*, 94–102.

[11] Pieter Heyvaert, Ben De Meester, Anastasia Dimou, and Ruben Verborgh. 2019. Rule-driven inconsistency resolution for knowledge graph generation rules. *Semant. Web* 10, 6 (2019). DOI:https://doi.org/10.3233/SW-190358

[12] Ademar Crotti Junior, Jeremy Debattista, and Declan O'Sullivan. 2019. Assessing the Quality of R2RML Mappings. In *Joint Proceedings of the International Workshop On Semantics For Transport and on Approaches for Making Data Interoperable co-located with 15th Semantics Conference, Karlsruhe, Germany* (CEUR Workshop Proceedings), CEUR-WS.

[13] Pavel Klinov and Dmitry Mouromtsev. 2015. Measuring the Quality of Relational-to-RDFMappings. In *Communications in Computer and Information Science*, Springer Verlag. DOI:https://doi.org/10.1007/978-3-319-24543-0

[14] Holger Knublauch and Dimitris Kontokostas. 2017. Shapes Constraint Language (SHACL), W3C Recommendation 20 July 2017. *URL https//www. w3. org/TR/shacl* (2017).

[15] Timothy Lebo, Satya Sahoo, Deborah McGuinness, Khalid Belhajjame, James Cheney, David Corsar, Daniel Garijo, Stian Soiland-Reyes, Stephan Zednik, and Jun Zhao. 2013. PROV-O: the prov ontology. w3c recommendation, 30 April 2013. *World Wide Web Consort.* (2013). Retrieved from https://www.w3.org/TR/prov-o/

[16] Bernadette Farias Lóscio, C Burle, and N Calegari. 2016. Data on the web best practices. W3C recommendation. *World Wide Web Consort.* (2016). Retrieved from https://www.w3.org/TR/dwbp/

[17] Deborah L McGuinness, Frank Van Harmelen, and others. 2004. OWL web ontology language overview. *W3C Recomm.* 10, 10 (2004), 2004.

[18] Alan Meehan, Rob Brennan, Dave Lewis, and Declan O'sullivan. *Mapping Representation based on Meta-data and SPIN for Localization Workflows*. Retrieved from http://www.w3.org/TR/its20/

[19] Benjamin Moreau and Patricia Serrano-Alvarado. 2020. Assessing the Quality of RDF Mappings with EvaMap. In *17th Extended Semantic Web Conference (ESWC2020)*. 164–167. DOI:https://doi.org/10.1007/978-3-030-62327-2_28

[20] Antonio De Nicola and Michele Missikoff. 2016. A Lightweight Methodology for Rapid Ontology Engineering. *Commun. ACM* 59, 3 (February 2016), 79–86. DOI:https://doi.org/10.1145/2818359

[21] Natalya F Noy, Deborah L McGuinness, and others. 2001. Ontology development 101: A guide to creating your first ontology.

[22] Christoph Pinkel, Carsten Binnig, Peter Haase, Clemens Martin, Kunal Sengupta, and Johannes Trame. 2014. How to best find a partner? An evaluation of editing approaches to construct R2RML mappings. In *Lecture Notes in Computer Science (including*

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 675–690. DOI:https://doi.org/10.1007/978-3-319-07443-6_45

[23] María Poveda-Villalón, Alba Fernández-Izquierdo, Mariano Fernández-López, and Raúl García-Castro. 2022. LOT: An industrial oriented ontology engineering framework. *Eng. Appl. Artif. Intell.* 111, (2022), 104755. DOI:https://doi.org/https://doi.org/10.1016/j.engappai.2022.104755

[24] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. 2014. OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation. *Int. J. Semant. Web Inf. Syst.* 10, 2 (2014), 7–34.

[25] Alex Randles and Declan O'Sullivan. 2021. Assessing quality of R2RML mappings for OSi's Linked Open Data portal. *4th Int. Work. Geospatial Linked Data ESWC 2021* (2021).

[26] Alex Randles, Declan O'Sullivan, John Keeney, and Liam Fallon. 2022. Applying a Mapping Quality Framework in Cloud Native Monitoring. In *18th International Conference on Semantics Systems (SEMANTiCS)*.

[27] Harald Sack, Eva Blomqvist, Mathieu d'Aquin, Chiara Ghidini, Simone Paolo Ponzetto, and Christoph Lange (Eds.). 2016. *RMLEditor: A Graph-Based Mapping Editor for Linked Data Mappings*. Springer International Publishing, Cham. DOI:https://doi.org/10.1007/978-3-319-34129-3

[28] Asunciónand Fernández-López Mariano Suárez-Figueroa Mari Carmenand Gómez-Pérez. 2012. The NeOn Methodology for Ontology Engineering. In *Ontology Engineering in a Networked World*, Asunciónand Motta Enricoand Gangemi Aldo Suárez-Figueroa Mari Carmenand Gómez-Pérez (ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 9–34. DOI:https://doi.org/10.1007/978-3-642-24794-1_2

[29] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Boris Villazón-Terrazas. 2009. How to write and use the ontology requirements specification document. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, 966–982.

[30] Protégé. Retrieved January 1, 2022 from https://protege.stanford.edu/