# Mapping kwaliteitsmodel DSO op standaard RDF vocabulaires

Dit document beschrijft de wijze waarop de mapping verloopt van het kwaliteitsmodel van de DSO naar standaard RDF vocabulaires.

## Het kwaliteitsraamwerk

Het DSO hanteert een algemeen kwaliteitsraamwerk, hieronder afgebeeld:

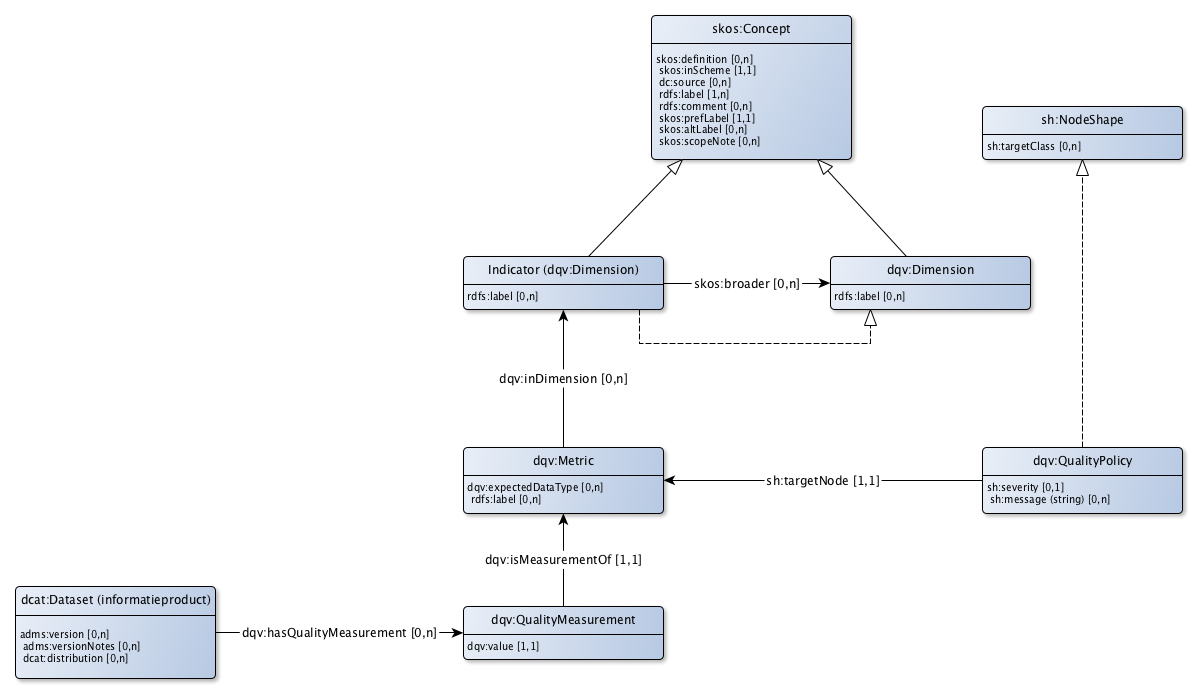


Dit kwaliteitsraamwerk laat zich goed mappen op de DQV standaard: <https://www.w3.org/TR/vocab-dqv>. Onderstaande tabel geeft de mapping.

|  |  |  |  |
| --- | --- | --- | --- |
| DSO term | DQV term | DQV definitie | Toelichting |
| Kwaliteitsdimensie | Dimension | “A dqv:Dimension is a quality-related characteristic of a dataset relevant to the consumer” | In de beschrijving van dqv zijn ook referenties gelegd naar de ISO standaard. Hier zie je overeenkomst met de kwaliteitsdimensies van DSO, dit mapt dus goed op elkaar. |
| Indicator | Dimension |  | Dqv heeft geen “tussenniveau” tussen Indicator en Meetfunctie. Indicators zijn specifieker dan kwaliteitsdimensies, maar nog niet concreet dat je ze kunt meten. Om dit op te vangen kunnen we Indicatoren ook als dqv:Dimensions zien, die d.m.v. skos:broader verbonden zijn (aangezien dqv:Dimension een subklasse is van skos:Concept zijn deze relaties hier goed bruikbaar en passend) |
| Meetfunctie | Metric | “A dqv:Metric gives a procedure for measuring a data quality dimension, which is abstract, by observing a concrete quality indicator” | De meetfunctie vertegenwoordigd de “definitie” van de te toetsen kwaliteitseis. |
| Kwaliteitsregel | QualityPolicy | “Represents a policy or agreement that is chiefly governed by data quality concerns.” | Hoewel DQV hier een invulling voor heeft, laat DQV vrij hoe dit verder wordt ingevuld. |
| Kwaliteitseis | QualityPolicy  shacl:PropertyShape | “Represents a policy or agreement that is chiefly governed by data quality concerns.” | Een kwaliteiteis geeft een concrete eis m.b.t. een meetresultaat. DQV biedt de mogelijkheid om QualityPolicy hier specifiek voor te maken, en geeft zelf als voorstel om daarvoor (in deze specifieke situatie) shacl voor te gebruiken als invulling. |
| Meetresultaat | QualityMeasurement | “Represents the evaluation of a given dataset (or dataset distribution) against a specific quality metric.” | Uitgangspunt is dat een kwaliteitseis in termen van de dataset feitelijk gaat over de kwaliteit die een dataset daadwerkelijk heeft (dus geen eis maar een beschrijving van de dataset) |
| Meetprocedure |  |  | In DQV is er geen logische tegenhanger aanwezig. De meetprocedure wordt niet in DQV beschreven. |

Voor het daadwerkelijk beschrijven van de kwaliteitskenmerken van een dataset, zijn de elementen dqv:Dimension, dqv:Metric en dqv:QualityMeasurement voldoende. Voor het beschrijven van kwaliteitseisen kan hiervoor dqv:QualityPolicy in combinatie met SHACL worden gebruikt.

Onderstaand figuur geeft het model weer dat voor de catalogus wordt gebruikt.



Het figuur beschrijft zowel de eisen waaraan een meeting dient te voldoen (via dqv:QualityPolicy), als het werkelijke meetresultaat dat bij het informatieproduct wordt vastgelegd (via dqv:QualityMeasurement). De meeteis is concreet, dwz: het is ook een shacl NodeShape, waarmee de constraint wordt vastgelegd waar de meeteis aan moet voldoen.

## De dimensies en indicatoren zoals binnen het DSO gedefinieerd

Bovenstaand geeft slechts een raamwerk, maar nog geen invulling aan de dimensies en indicatoren die daadwerkelijk aanwezig zijn. Er zijn daarbij drie mogelijke keuzen:

1. Letterlijk overnemen van de dimensies en indicatoren zoals in het DSO framework gedefinieerd (dwz: een eigen lijst van dimensies en indicatoren definieren);
2. Gebruiken van ISO dimensies en indicatoren;
3. Gebruiken van ldqv (<https://www.w3.org/2016/05/ldqd>)

Optie (A) heeft de voorkeur, om de volgende redenen:

1. Het gebruik van ISO dimensies en indicatoren is niet mogelijk, aangezien ISO hiervoor geen URI’s heeft gedefinieerd. Het zelf munten van URI’s zou ertoe leiden dat we ons voordoen als ISO-organisatie, wat vanzelfsprekend niet zo is.
2. Het gebruik van ldqv dimensies en indicatoren is niet mogelijk: ldqv definieert alleen dimensies en geen indiatoren, en daarbij komt dat ldqv zich richt op Linked Data datasets, en niet datasets in het algemeen. Bij het gebruik van ldqv zou sowieso nog 1 van de andere opties nodig zijn.
3. Hoewel het zelf maken van een eigen lijst in tegenspraak is met de best practice om zoveel mogelijk gebruik te maken van bestaande vocabulaires is in dit geval geen geschikte vocabulaire online beschikbaar. Aangezien in dit geval sprake is van een lijst die alleen binnen de DSO gebruikt wordt, en omdat sprake is van skos:Concepts, is dit nadeel beperkt. Bovendien geeft DQV zelf ook aan dat het opnemen van eigen indicatoren en dimensies toegestaan is, en geeft DQV een manier om daarbij te linken met bestaande indicatoren en dimensies (via bv skos:broadMatch).

De volgende prefixes zullen worden gebruikt:

* http://standaard.omgevingswet.overheid.nl/id/kwaliteitsdimensie/{term}
* http://standaard.omgevingswet.overheid.nl/id/indicator/{term}

|  |  |
| --- | --- |
| **Kwaliteitsdimensie (dqv:Dimension)** | **Dimensie ISO/IEC 25012** |
| Nauwkeurigheid | Accuracy |
| Compleetheid | Completeness |
| Consistentie | Consistency |
| Actualiteit | Currentness |
| Precisie | Accuracy |
| Pausibiliteit | Credibility |
| Traceerbaarheid | Tracebility |
| Compliance | Compliance |
| Begrijpelijkheid | Understandability |

Voorbeeld van de uitwerking van een kwaliteitsdimensie:

@prefix kwaliteitsdimensie: <http://standaard.omgevingswet.overheid.nl/id/kwaliteitsdimensie/>

kwaliteitsdimensie:Compleetheid a dqv:Dimension;

skos:exactMatch iso:Consistency;

rdfs:label “Compleetheid”@nl;

skos:definition “De mate waarin gegevens gerelateerd aan een object waarden hebben voor alle verwachte attributen en gerelateerde object instanties in een specifieke gebruikscontext.”@nl;

.

In het DSO kwaliteitsraamwerk zijn bovengenoemde kwaliteitsdimensies uitgewerkt in indicatoren. Het gaat te ver om al deze indicatoren hier nogmaals op te noemen. De vertaling zal als volgt gaan:

* Elke indicator wordt opzichzelf een dqv:Dimension;
* Daarbij krijgt elke indicator een skos:broader relatie naar de bijbehorende kwaliteitsdimensie (ook weer een dqv:Dimension).

Voorbeeld van de uitwerking van een indicatie (De indicatie “metadata compleetheid” van de kwaliteitsdimensie “compleetheid”):

@prefix kwaliteitsdimensie: <http://standaard.omgevingswet.overheid.nl/id/kwaliteitsdimensie/>

@prefix indicator: <http://standaard.omgevingswet.overheid.nl/id/indicator>

indicator:MetadataCompleetheid a dqv:Dimension;

skos:broader kwaliteitsdimensie:Compleetheid;

rdfs:label “Compleetheid”@nl;

skos:definition “De mate waarin metadata attributen aanwezig zijn.”@nl;

.

## Meetfuncties en meetresultaten

Het DSO raamwerk geeft geen invulling van de meetfuncties. Daarmee kan feitelijk nog niet worden afgemaakt hoe het DSO raamwerk ingevuld moet worden. Praktisch gezien betekent dit dat de meetfuncties (en meetwaarden) onderdeel zijn van de metadata van een dataset (dwz: een informatiehuis kan zelf de meetfuncties bepalen). Voor de controleerbaarheid van de kwaliteit en voor het gemak van de informatiehuizen zou het goed zijn als er in de loop van de implementatie een lijst ontstaat. Mogelijk kan een minimale lijst opgenomen worden in de acceptatiecriteria voor informatiehuizen. Een dergelijke lijst kan dan worden beheerd in de DSO catalogus. Hieronder een voorbeeld de invulling van een meetfunctie inclusief meetresultaat er uit komt te zien. Het meetresultaat is dan de concrete waarde voor de kwaliteiteis zoals gemeten c.q. vastgesteld voor het betreffende informatieproduct.

@prefix meetfunctie: <standaard.omgevingswet.overheid.nl/id/meetfunctie/>

meetfunctie:PercentageGedefinieerdeBegrippen a dqv:Metric;

dqv:inDimension indicator:MetadataCompleetheid;

dqv:expectedDatatype xsd:integer;

rdfs:label “Percentage gedefinieerde begrippen”;

skos:definition “Aantal gedefinieerde begrippen t.o.v. het aantal klassen en attributen in de dataset”@nl;

skos:scopeNote “Van elke klasse of attribuut wordt een relatie verwacht naar een begrip. Wel kunnen twee klassen of attributen verwijzen naar hetzelfde begrip. Het aantal begrippen hoeft dus niet gelijk te zijn aan het aantal klassen en attributen om tot een 100% score te komen op deze meetfunctie”@nl.

.

meetresultaat:PercentageGedefinieerdeBegrippen a dqv:QualityMeasurement;

dqv:isMeasurementOf meetfunctie:PercentageGedefinieerdeBegrippen;

dqv:value 95;

.

## Kwaliteitseis

Om vast te stellen of voldaan wordt aan een kwaliteitseis, wordt gebruik gemaakt van dqv:QualityPolicy in combinatie met SHACL (<https://www.w3.org/TR/shacl>). Dit garandeert dat het mogelijk is om automatisch te valideren of een combinatie van meetresultaten voldoet aan de kwaliteitseisen die zijn vastgesteld voor het betreffende informatieproduct.

Onderstaand voorbeeld geeft aan hoe een dergelijke kwaliteitseis eruit zou zien met betrekking tot de meetfunctie “percentage gedefinieerde begrippen”, waarbij de waarde minimaal 95 moet zijn.

kwaliteitseis a dqv:QualityPolicy, sh:NodeShape;

sh:targetNode meetfunctie:PercentageGedefinieerdeBegrippen;

sh:message “Het percentage gedefinieerde begrippen dient minimaal 95 te zijn”;

sh:severity sh:Violation;

sh:property [

sh:path [ sh:inversePath sh:isMeasurementOf ];

sh:node [

sh:property [

sh:path dqv:value;

sh:minInclusive 95

]

]

]

.