

# Wijzigingsverzoek NORA raamwerk gegevenskwaliteit

Danny Greefhorst, 26 april 2024

Deze notitie beschrijft een verzoek tot het aanpassen van het NORA raamwerk gegevenskwaliteit naar aanleiding van nader onderzoek in de context van het programma Federatief Data Stelsel. In de context van dit programma is het raamwerk omarmd en zijn stappen gezet om het raamwerk verder te valideren en verdiepen voor de basisregistraties. Daarbij is de focus gelegd op de dimensies juistheid, compleetheid en actualiteit. Dat heeft geleid tot extra inzichten, een aanpak waarmee op een meer uniforme manier gegevenskwaliteit kan worden uitgedrukt en betere voorbeelden.

## Aanscherping van het raamwerk

Bij het gebruik van het raamwerk ontstonden vragen over de precieze bedoeling van de verschillende dimensies en attributen. De volgende onderstreepte tekstuele toevoegingen in de toelichting van drie kwaliteitsattributen zouden een verbetering zijn:

**Toelichting bij thematische juistheid:** Dit gaat vooral over of gegevens inhoudelijk kloppen. Idealiter is dat of ze overeenkomen met de werkelijkheid, maar in de praktijk is dat of ze overeenkomen met een registratie die dichterbij de werkelijkheid ligt of meer betrouwbaar is. Het is daarom bij het definiëren van de kwaliteit deze registraties waarmee wordt vergeleken expliciet te benoemen. Daarnaast geldt dat een dergelijke vergelijking niet zinvol is voor een registratie die zelf een juridische werkelijkheid creëert, omdat de registratie die zelf bepaalt. Thematische juistheid kan zowel betrekking hebben op de reële werkelijkheid als de juridische werkelijkheid.

**Toelichting bij dataset compleetheid:** Dit gaat over of voor alle objecten in de werkelijkheid die in scope zijn van de dataset ook als gegevensobject aanwezig zijn in de dataset. Om dit te kunnen bepalen moet ook duidelijk zijn wat het geheel is. Het is daarom belangrijk om ook de populatie van de dataset expliciet te maken. Daarnaast kunnen er altijd bepaalde objecten in de werkelijkheid zijn die we nog niet kennen en die we dus ook niet beschouwen als onderdeel van het geheel.

**Toelichting bij versheid:** Dit heeft betrekking op hoe oud gegevens zijn. Gegevens moeten periodiek opnieuw worden gemeten of er moet worden gecontroleerd of ze nog correct zijn. Dat kun je uitdrukken in termen van het verschil tussen het moment waarop een gegeven is ingewonnen en het moment waarop dit gegeven zichtbaar is voor een eindgebruiker.

## Aanpak voor het formuleren van kwaliteit

Naast genoemde aanscherpingen van het raamwerk is er ook een aanpak gedefinieerd voor het definiëren van gegevenskwaliteit, die als standaard aanpak zou moeten worden toegevoegd bij de beschrijving van het raamwerk. Het kernidee daarbij is dat kwaliteit het best inzichtelijk kan worden gemaakt door deze uit te schrijven in zinnen, omdat het doel vooral is om kwaliteit op een begrijpelijke manier inzichtelijk te maken voor gebruikers. Daarnaast wordt voorgesteld om gebruik te maken van voorgedefinieerde zinsjablonen, omdat daarmee de consistentie en daarmee de begrijpelijkheid van de zinnen kan worden verhoogd. Een voorbeeld van een zinsjabloon is:

*voor ATTRIBUUT: PERCENTAGE van de gegevens komt overeen met de gegevens in BRONNEN*

Hierin is zichtbaar dat deze zinsjablonen gebruik maken van voorgedefinieerde parameters waarin de relevante parameters per registratie worden vastgelegd. Er zijn een aantal zinsjablonen gedefinieerd en getoetst op een aantal basisregistraties, waaruit bleek dat deze aanpak goed werkt. Dit heeft geleid tot een aantal expliciet getoetste zinsjablonen voor thematische juistheid, dataset compleetheid en versheid. Aanvullend zijn ook voor de andere kwaliteitsattributen in het raamwerk zinsjablonen geformuleerd, mede op basis van het eerder beschreven raamwerk gegevenskwaliteit Omgevingswet. Het voorstel is om deze zinsjablonen expliciet toe te voegen aan het raamwerk. Ze geven ook een beter beeld van hoe de kwaliteitsattributen geoperationaliseerd kunnen worden.

De volgende aandachtspunten gelden bij deze aanpak:

- Het is niet per definitie mogelijk om in een specifieke context een voorgesteld zinsjabloon te gebruiken, bijvoorbeeld omdat een net iets anders en/of meer genuanceerd perspectief relevant is. De zinsjablonen moeten daarom niet als verplicht worden beschouwd (wel als sterk aanbevolen). Het is wel waardevol als partijen aanvullende zinsjablonen aandragen ter aanvulling (of verbetering) van de voorgestelde zinsjablonen. Zo kunnen er dus meerdere zinsjablonen per kwaliteitsattribuut ontstaan.
- Het is relevant om zowel de kwaliteitseis (norm) als het resultaat van kwaliteitsmetingen (realisatie) te beschrijven. Daarbij kunnen dezelfde zinsjablonen worden gebruikt. Het voorstel is om niet in de zinsjabloon op te nemen of het over een norm of de realisatie gaat, omdat de zinnen zo eenvoudig mogelijk te houden en ook los over normen en realisaties te kunnen spreken. Het voorstel is om dit onderscheid op te lossen op de plaats waar de zinnen worden gebruikt. Ze zouden onder aparte koppen in een document kunnen worden opgenomen, of op andere wijze onderscheidend gemarkeerd.
- De zinnen zijn niet per definitie zelfverklarend. Het is daarom belangrijk om ruimte te kunnen bieden aan extra toelichting, waarin relevante context en nuances kunnen worden beschreven. Het is nader te onderzoeken in hoeverre dit soort toelichtingen ook passen in de DQV standaard en dus kunnen worden omgezet in gestructureerde vorm.
- Er worden in de zinsjablonen de standaard termen OBJECTTYPE en ATTRIBUUT als parameter gebruikt. Deze termen zijn gebaseerd op de standaard Metamodel Informatie Modelling. Een objecttype is een logische groepering van gelijksoortige data. Een attribuut is een kenmerk van een objecttype. Praktisch is er bij het gebruik van de zinsjablonen de precieze naam van het objecttype of attribuut(soort) zoals ze zijn gedefinieerd in een informatie- of gegevensmodel niet altijd het meest logisch. Er is een behoefte om ook iets meer vrij te kunnen spreken over soorten gegevens. Een dergelijke pragmatiek wordt daarom aanbevolen.
- In een aantal zinsjablonen wordt ATTRIBUUT als ingangspunt gebruikt voor het zinsjabloon. Het kan zijn dat een attribuut in meerdere objecttypes wordt gebruikt. In dat geval kan in de plaats van de parameter ATTRIBUUT ook een combinatie van een objecttype en een attribuut worden ingevuld.

## Voorbeelden

Er zijn in het proces op basis van de zinsjablonen een aantal nieuwe voorbeelden gegenereerd voor een aantal basisregistraties die een veel beter beeld geven van een waardevolle manier om gegevenskwaliteit uit te drukken dan de huidige voorbeelden die worden gegeven bij het raamwerk. Die huidige voorbeelden zijn eigenlijk geen goed uitgewerkte voorbeelden, maar geven slechts hints van aspecten die relevant zijn om te beschouwen. Dit is duidelijk niet hetzelfde als Het voorstel is dan ook om:

- De uitgewerkte voorbeelden voor de basisregistraties toe te voegen.

- De huidige voorbeelden te verwijderen en om te vormen naar een extra categorie van aanvullende toelichting (naast de voorgestelde extra categorie richtlijnen voor formulering) bij de beschrijving van de kwaliteitsattributen.
- De categorie van aanvullende toelichting verder te verrijken met informatie zoals die eerder was opgenomen in de documenten die hoorden bij het raamwerk gegevenskwaliteit Omgevingswet, die een belangrijke bron was voor het NORA raamwerk gegevenskwaliteit.

Het verdient verder sterke aanbeveling om in de verdere toekomst ook voorbeelden te genereren voor andere kwaliteitsattributen.

## Bijlage: voorgestelde zinsjablonen

Dimensie	Kwaliteitsattribuut	Zinsjabloon
Juistheid	Thematische juistheid	voor ATTRIBUUT: PERCENTAGE van de gegevens komt overeen met de gegevens in BRONNEN
Juistheid	Classificatie juistheid	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten in de registratie zijn hebben het juiste objecttype
Juistheid	Positionele juistheid	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten in de registratie waarvan de geometrie maximaal AFSTAND afwijkt van de vorm en ligging van de objecten in de werkelijkheid
Juistheid	Kwantitatieve juistheid	voor ATTRIBUUT: PERCENTAGE van de gegevens waarvan de waarde maximaal AFWIJKING afwijkt van de juiste waarde
Compleetheid	Dataset compleetheid	voor OBJECTTYPE: PERCENTAGE van de objecten zoals bekend in BRONNEN zijn opgenomen in de registratie
Compleetheid	Attribuut compleetheid	voor ATTRIBUUT: PERCENTAGE van de gegevens in de registratie hebben een waarde
Compleetheid	Overcompleetheid	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten in de registratie zijn hier niet onterecht in opgenomen
Compleetheid	Historie compleetheid	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten in de registratie bevat gegevens over de volledige materiële historie van het object in de werkelijkheid
Compleetheid	Metadata compleetheid	PERCENTAGE van de modelementen in het gegevensmodel zijn voorzien van een definitie
Compleetheid	Ruimtelijke dekking	voor PERCENTAGE van het geografische gebied zijn gegevens beschikbaar
Compleetheid	Identificeerbaarheid	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten in de registratie is voorzien van een unieke identificatie
Validiteit	Formaatvaliditeit	PERCENTAGE van de gegevensobjecten in de registratie voldoet aan de formaateisen
Validiteit	Domeinvaliditeit	voor ATTRIBUUT: PERCENTAGE van de gegevens hebben een toegestane waarde
Validiteit	Geometrische validiteit	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten in de registratie hebben een valide geometrie
Consistentie	Logische consistentie	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten in de registratie voldoen aan de gedefinieerde beperkingsregels
Consistentie	Referentiële integriteit	voor ATTRIBUUT: PERCENTAGE van de gegevens in de registratie die verwijzen naar een bestaand gegevensobject
Consistentie	Homogeniteit	voor ATTRIBUUT: PERCENTAGE van de gegevens in de registratie is beschreven met DETAILNIVEAU
Consistentie	Topologische consistentie	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten in de registratie hebben een topologisch consistente geometrie

Actualiteit	Updatefrequentie	voor OBJECTTYPE: PERCENTAGE van de gegensubjecten in de registratie is geactualiseerd binnen PERIODE
Actualiteit	Versheid	voor OBJECTTYPE: PERCENTAGE van de objecten zijn binnen NORMTIJD opgenomen of gewijzigd in de registratie
Precisie	Opslagprecisie	voor ATTRIBUUT: is ingewonnen met een precisie van AANTAL EENHEID
Precisie	Geometrische precisie	voor ATTRIBUUT: is ingewonnen met een resolutie van AANTAL EENHEID
Precisie	Statistische precisie	voor ATTRIBUUT: is ingewonnen met een precisie van AANTAL EENHEID
Plausibiliteit	Authenticiteit	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten is aantoonbaar afkomstig van de aanleverende partij
Plausibiliteit	Reputatie	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten is aangeleverd door een vertrouwde partij
Plausibiliteit	Bewijsbaarheid	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten is voorzien van BEWIJS
Plausibiliteit	Representativiteit	de registratie is gebaseerd op een steekproefgrootte van AANTAL EENHEID op een populatie van AANTAL EENHEID
Plausibiliteit	Waarschijnlijkheid	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten in de registratie voldoen aan de gedefinieerde plausibiliteitsregels
Traceerbaarheid	Herleidbaarheid	voor OBJECTTYPE: PERCENTAGE van de gegevensobjecten dat herleid kan worden tot BRON
Traceerbaarheid	Reproduceerbaarheid	PERCENTAGE van de afgeleide attributen zijn gerelateerd aan de daarbij gebruikte transformatieregels
Traceerbaarheid	Gebruiksinzicht	PERCENTAGE van de attributen in de registratie zijn nadrukkelijk zijn expliciet gerelateerd aan een gebruiksdoel
Begrijpelijkheid	Duidelijkheid	voor ATTRIBUUT: is geschreven op taalniveau TAALNIVEAU