

**Forum Standaardisatie**Wilhelmina van Pruisenweg 52
2595 AN Den HaagPostbus 96810
2509 JE Den Haagwww.forumstandaardisatie.nl

notitie

FORUM STANDAARDISATIE 13 juni 2018

Agendapunt:	3C		
Betreft:	Intake-advies voor SHACL		
Aan:	Forum Standaardisatie		
Van:	Stuurgroep Open Standaarden		
Datum:	30 mei 2018	Versie	1.0

Advies

Het Forum Standaardisatie wordt geadviseerd om de Linked Open Data standaard SHACL in procedure te nemen voor plaatsing op de lijst aanbevolen standaarden.

Korte toelichting:

Shapes Constraint Language (SHACL) is een recent gepubliceerde, opkomende standaard op het vlak van gegevenskwaliteit en de begrijpelijkheid van gegevensverzamelingen, vormgegeven volgens Linked Open Data principes. Daarmee heeft de standaard in potentie toegevoegde waarde voor interoperabiliteit. Steeds meer overheidsorganisaties tonen interesse voor Linked Open Data, tot heden is het Kadaster een voorloper in het gebruik.

Het beheer van SHACL wordt door W3C op een open wijze uitgevoerd. Het Platform Linked Open Data Nederland (PLDN) is gegeven de publiek / private samenstelling van deskundig partijen op het gebied van Linked Open Data uitstekend gepositioneerd voor het volgen van deze ontwikkelingen.

SHACL heeft een relatie met de standaarden SKOS (verplicht), RDF en OWL (beiden aanbevolen). Deze hebben geen invloed op de aanmelding en toetsing.

Het Kadaster ondersteunt de aanmelding van de standaard en gebruikt SHACL actief. Ook bij het Informatiehuis Ruimte wordt SHACL toegepast. Het is denkbaar dat bij bewezen toegevoegde waarde op termijn de inzet van SHACL uitbreidt naar de andere informatiehuizen in het digitaal stelsel

voor de omgevingswet. Ook gegevens van het Centraal Bureau voor de Statistiek (CBS) en de Rijksdienst voor Cultureel Erfgoed (RCE) worden als Linked Open Data ontsloten door het Kadaster, maar CBS en RCE passen SHACL zelf nog niet toe. Kadaster haalt de data bij het CBS en RCE op en converteert deze zelf tot Linked Open Data. De kennis van SHACL blijft daarmee tot nu toe beperkt tot het Kadaster.

Voorwaarde voor plaatsing van SHACL op de 'pas toe of leg uit' lijst als verplichte standaard is het gebruik van Linked Open Data door meerdere overheidspartijen die ook daadwerkelijk gegevens met elkaar uitwisselen. SHACL is dan voorwaardelijk voor het realiseren van interoperabiliteit. Dit is nog niet het geval. Door de recente publicatie (in juli 2017) is er nog nauwelijks ervaring met de implementatie en gebruik van de standaard, en is de marktondersteuning nog in ontwikkeling. Omdat bovendien er geen sprake is van buitengewone urgentie (bijvoorbeeld internet veiligheid), is plaatsing op de 'pas toe of leg uit' lijst nog prematuur.

Het advies is daarom om SHACL in procedure te nemen voor de lijst aanbevolen standaarden.

Toelichting

1. Aanmelding, intakegesprek en toetsingsprocedure

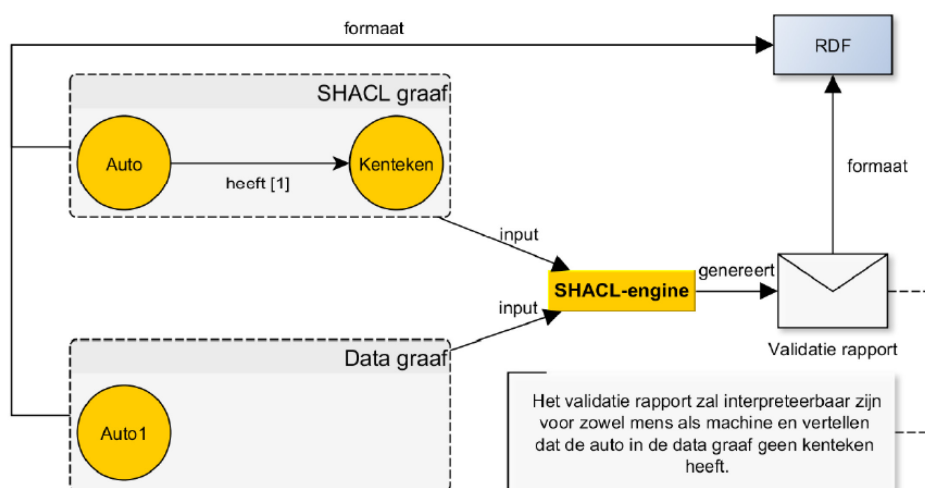
Op 30-04-2018 is door Jesse Bakker van Taxonic namens het Platform Linked Open Data Nederland (PLDN) een standaard aangemeld, betreffende *aanmelden* van de standaard Shapes Constraint Language (SHACL) voor de lijst met open standaarden. De aanmelder had bij indiening als doel de standaard verplicht ('pas-toe-of-leg-uit') te stellen.

Op 16 mei 2018 heeft een intakegesprek plaatsgevonden met de aanmelder. In dit gesprek is geprobeerd zoveel mogelijk informatie te verkrijgen om te kunnen bepalen of de standaard aan de criteria voldoet om mogelijk als verplichte standaard op de 'pas toe of leg uit lijst' te worden geplaatst.

2. Korte beschrijving standaard

Waar gaat SHACL over?

SHACL is een op Resource Description Framework (RDF) gebaseerde taal om 'constraints' of condities op datasets uit te drukken. Bijvoorbeeld: "een auto moet een kenteken hebben". De standaard heeft betrekking op datasets die als Linked Open Data zijn ingericht. SHACL is niet alleen een beschrijvende taal maar implementeert ook instructies die door machines kunnen worden uitgevoerd. Bijvoorbeeld: "als een auto geen kenteken heeft geeft dan een foutmelding". Zo kan de kwaliteit van bestaande datasets (gegevensbeheer) en aangeboden data (bij gegevensuitwisseling) automatisch worden beheerd.



Figuur 1: illustratie van het gebruik van SHACL

SHACL beschrijft – naast de condities – hoe de informatie en semantiek van een dataset gestructureerd is. Deze beschrijving ("bijsluiters") is zowel voor mensen als machines leesbaar. Daardoor wordt het voor organisaties tevens makkelijker elkaars datasets te begrijpen en te integreren in de eigen gegevenshuishouding. SHACL heeft daarmee belangrijke toegevoegde waarde bij Linked Open Data.

SHACL werd in juli 2017 gepubliceerd door W3C en is dus nog een jonge standaard.

Welk probleem lost de standaard op?

Gebrek aan vertrouwen in gegevenskwaliteit is in de praktijk een grote drempel voor het hergebruik van gegevens tussen organisaties (interoperabiliteit). Dit is een algemeen probleem en niet beperkt tot een functioneel of organisatorisch toepassingsgebied. SHACL lost het probleem van gegevenskwaliteit beter op dan bestaande standaarden en werkmethoden.

Wie beheert de standaard?

W3C (<http://www.w3.org>) beheert de standaard.

Waarom is de standaard aangemeld voor pas toe of leg uit?

De aanmelder is ervan overtuigd dat SHACL de gegevensuitwisseling op basis van Linked Open Data ("triples") een enorme stap vooruit helpt. SHACL slecht de barrière van ontoereikende gegevenskwaliteit en de inefficiënte controleerbaarheid ervan. Door SHACL wordt het tevens voor organisaties makkelijker elkaars datasets te begrijpen en de juiste 'constraints' toe te passen bij gegevensuitwisseling. SHACL is van toepassing op iedere (semi-)overheidsorganisatie die gebruik maakt van Linked Open Datasets in welk inhoudelijk domein dan ook.

SHACL werkt in combinatie met RDF triple stores. Er is een open source implementatie van de SHACL tooling (Engine) beschikbaar. TopQuadrant heeft een open source tool ontwikkelt voor SHACL:

<https://github.com/TopQuadrant/shacl>

(zie ook: 7. Functionele use case)

3. Criteria voor inbehandelname

Om een standaard in behandeling te nemen moet de standaard vallen binnen de scope van de lijst. Hiervoor gelden vier criteria:

1. Is de standaard toepasbaar voor elektronische gegevensuitwisseling tussen (semi-)overheidsorganisaties en bedrijven, tussen (semi-)overheidsorganisaties en burgers of tussen (semi-)overheidsorganisaties onderling?

Ja, dat is het geval. Voorwaarde is dat de dataset als Linked Open Dataset is ingericht (triple store). De overheid heeft grote ambities als het gaat om het beschikbaar stellen van data voor hergebruik, zowel als gesloten data tussen overheden als open data voor de Nederlandse samenleving. De begrijpelijkheid van datasets neemt door SHACL toe, evenals de inhoudelijke kwaliteit van gegevens. Beiden dragen bij aan elektronische gegevensuitwisseling en de interoperabiliteit.

2. Is het beoogde functioneel toepassingsgebied en het organisatorisch werkingsgebied van de standaard, voldoende breed om substantieel bij te dragen aan de interoperabiliteit van de (semi-)overheid?

Ja, dat is het geval. De standaard heeft betrekking op begrijpelijkheid datasets en de kwaliteit van gegevens. Veel (semi-) overheidsorganisaties

ontsluiten datasets voor burgers, bedrijven en andere overheidsorganisaties. SHACL is niet gebonden aan een specifieke sector. De enige voorwaarde voor de toepassing van SHACL is dat data als Linked Open Data is ingericht. Dat is echter een inrichtingskeuze van een dataset door de eigenaar en niet verplicht. In het project Linked Open Data Overheid wordt gewerkt aan de bredere inzet van Linked Open Data bij de overheid, zie <https://data.overheid.nl/linked-data-overheid>.

3. Is het zinvol de standaard op te nemen, gezien het feit dat deze niet al wettelijk verplicht is voor het beoogde functioneel toepassingsgebied en organisatorisch werkingsgebied?

Ja, SHACL is niet wettelijk verplicht.

4. Draagt de standaard bij aan de oplossing van een bestaand, relevant (interoperabiliteits)probleem en het voorkomen van leveranciersafhankelijkheid?

Ja, dat is aannemelijk. In potentie wel (technisch), maar in de praktijk momenteel nog niet.

Technisch gezien draagt SHACL bij aan het oplossen van (onder meer) een interoperabiliteitsprobleem bij de toepassing van Linked Open Data. De indieners betogen dat Linked Open Data een boost zal krijgen als gevolg van SHACL en dat de potentie van Linked Open Data hiermee beter tot zijn recht komt. Linked Open Data (triple stores) zou dan de voorkeursimplementatie voor datasets kunnen worden.

Conclusie

SHACL voldoet aan de basiscriteria om in behandeling genomen te worden voor plaatsing op de lijst open standaarden. Voor een 'pas toe of leg uit' verplichting lijkt het echter nog vroeg, een procedure voor plaatsing op de lijst aanbevolen standaarden ligt meer voor de hand. In het volgende hoofdstuk wordt dit verder beargumenteerd.

4. Toetsing kansrijkheid procedure

Het Forum Standaardisatie wil voorkomen dat er standaarden in procedure worden genomen, waarvan bij voorbaat al bekend is dat deze in de expertronde of consultatieronde zullen stranden op één van de inhoudelijke criteria. Daarom heeft de procedurebegeleider de beantwoording van de criteriavragen nagelopen, waar mogelijk zelf aangevuld en vervolgens besproken met de indiener.

1. Open standaardisatieproces

De ontwikkeling en het beheer van de standaard moeten op een open, onafhankelijke, toegankelijke, inzichtelijke, zorgvuldige en duurzame wijze zijn ingericht.

W3C beheert meerdere standaarden die op de lijst van het Forum Standaardisatie staan. Voor SHACL hanteert W3C hetzelfde proces voor

ontwikkeling en beheer van de standaard. Deze voldoet aan de eisen. Er is geen wens om het predicaat uitstekend beheer toe te wijzen voor SHACL.

2. Toegevoegde waarde

De interoperabiliteitswinst en andere voordelen van adoptie van de standaard wegen overheidsbreed en maatschappelijk op tegen de kosten, de risico's en nadelen. Voor elk van de te onderscheiden stakeholders (overheid, bedrijven en burgers) afzonderlijk zouden de baten voor de informatievoorziening en de bedrijfsvoering op moeten wegen tegen de kosten. Verder moeten de risico's aan overheidsbrede adoptie van de standaard (beveiliging, privacy) acceptabel zijn.

Voor organisaties die Linked Open Data ontsluiten of gebruiken heeft SHACL *potentieel* toegevoegde waarde omdat de standaard de kwaliteit van de data verhoogt. Er is pas sprake van *werkelijk* toegevoegde waarde als meerdere organisaties SHACL toepassen, zodat de interoperabiliteit aantoonbaar wordt. Dit is nu nog niet het geval.

Er is geen details bekend over de implementatiekosten voor SHACL. Er is open source tooling beschikbaar die zonder kosten te verkrijgen is. Daarnaast zijn de inrichtings-/implementatiekosten niet gespecificeerd maar worden als gering ingeschat mits de dataset als Linked Open Data is ingericht. In veel gevallen is dat niet zo en zal de bestaande dataset eerst als Linked Open Data moeten worden geïmplementeerd. De kosten hiervan zijn niet bekend.

SHACL introduceert geen nieuwe risico's op het vlak van privacy en informatiebeveiliging.

3. Draagvlak

Aanbieders en gebruikers moeten voldoende ervaring hebben met de implementatie en het gebruik van de standaard.

Het is nog te vroeg om te stellen dat er draagvlak is op basis van ervaring met de implementatie en het gebruik van de standaard bij gegevensuitwisseling. SHACL werd in juli 2017 gepubliceerd en er is nog weinig praktijkervaring met de standaard.

Ervaring met het gebruik van SHACL binnen de Nederlandse overheid beperkt zich tot het Kadaster. Het Kadaster maakt ook gebruik van gegevens van het CBS en RCE. Datasets zijn verkregen door het Kadaster en worden verrijkt tot Linked Open Data door het Kadaster en hierbij wordt gebruik gemaakt van SHACL.

Gezien de ontwikkelingen in het ruimtelijk domein, voornamelijk rondom de Omgevingswet, is de verwachting dat de adoptie en de toepassing van Linked Open Data zal toenemen. Het toepassen van SHACL kan dan essentieel worden voor het bevorderen van de interoperabiliteit.

Verder overweegt de Nationale Politie een pilot/experiment met SHACL. Als deze pilot zich ook uitstrekt tot uitwisseling van gegevens kan deze pilot eveneens belangrijk ervaring opleveren.

De experts aangesloten bij Platform Linked Open Data Nederland (PLDN) (<https://www.pilod.nl>) omarmen de meerwaarde van SHACL en adviseren deze op de 'pas toe of leg uit' lijst te plaatsen.

Aanbieders van ICT componenten gebaseerd op SHACL zijn niet bekend. Er is open source tooling beschikbaar vanuit TopQuadrant. Aanbieder van adviesdiensten voor de toepassing van SHACL zijn Taxonic, Ordina, Netage en Semaku en mogelijk andere private deelnemers (marktpartijen) in PLDN.

4. Opname bevordert adoptie

De opname op de lijst moet een geschikt middel zijn om de adoptie van de standaard te bevorderen.

De standaard is aangemeld voor de lijst van verplichte standaarden. Gezien de geringe ervaring met de standaard, de zich nog ontwikkelende marktondersteuning en de beperkte toepassing van de standaard binnen de Nederlandse (semi) overheidspartijen is niet zonder meer te stellen dat een 'pas toe of leg uit' status nu het passende middel is om het gebruik van SHACL te stimuleren.

SHACL kan echter wel worden gekarakteriseerd als een opkomende standaard met potentieel sterke toegevoegde waarde in het Linked Open Data domein. Opname op de lijst aanbevolen standaarden zou in de huidige situatie wel passend kunnen zijn.

Conclusie

De kansrijkheid van de toetsingsprocedure voor de 'pas toe of leg uit' lijst is onvoldoende. Het ontbreekt op dit moment aan ervaring met de standaard, aan solide marktondersteuning en aan gebruik door meerdere overheidspartijen.

SHACL heeft meer kans om met positief getoetst te worden voor opname op de lijst aanbevolen standaarden.

Door de standaard op de lijst aanbevolen standaarden op te nemen kan er draagvlak ontstaan onder een bredere groep belanghebbenden. Tevens wordt ook inzicht verkregen in de noodzaak van bevordering van adoptie van de standaard wat in de toekomst kan leiden tot opname als verplichte standaard. Het PLDN is gegeven de publiek / private samenstelling van deskundig partijen op het gebied van Linked Open Data uitstekend gepositioneerd voor het volgen van deze ontwikkelingen.

5. Samenhang

Het Forum Standaardisatie wil weten of de aangemelde standaard samenhangt met standaarden die reeds op de lijst zijn opgenomen, of standaarden die voor toetsing in aanmerking komen. Uit de intake moet duidelijk worden of dit gevolgen heeft voor de toetsing en eventuele opname van de aangemelde standaard.

1. *Bestaat er samenhang tussen de aangemelde standaard en de verplichte ('pas-toe-of-leg-uit') standaarden die reeds op de lijst zijn opgenomen en wat betekent dit voor de toetsing en eventuele opname van de standaard?*

De indiener van SHACL verwijst naar de volgende standaard die opgenomen is als verplichte standaard:

- Simple Knowledge Organization System (SKOS)

SHACL en SKOS sluiten bij elkaar aan, beide zijn standaarden van belang voor de relaties tussen waardenlijsten. Er is verder geen afhankelijkheid tussen SHACL en SKOS.

Het feit dat SKOS verplicht is heeft geen gevolgen voor de toetsing en opname van SHACL op de lijst met standaarden.

2. *Bestaat er samenhang tussen de aangemelde standaard en de aanbevolen standaarden die reeds op de lijst zijn opgenomen en wat betekent dit voor de toetsing en eventuele opname van de standaard?*

Het specificatiedocument van SHACL verwijst naar de volgende standaarden die opgenomen zijn als aanbevolen standaard:

1. Resource Description Framework (RDF)
2. Web Ontology Language (OWL)

SHACL is compatible met deze standaarden (en alle andere Linked Open Data standaarden).

Over OWL stelt de indiener dat deze omwille van twee inhoudelijke eigenschappen (waaronder de 'open world assumption') ongeschikt is voor het voor het uitdrukken van 'constraints' op datasets. Naar verwachting van de indiener zal OWL vooral wetenschappelijke toepassingen kennen, SHACL praktische toepassingen.

Het feit dat bovengenoemde standaarden aanbevolen zijn heeft geen gevolgen voor de toetsing en opname van SHACL op de lijst met standaarden.

3. *Bestaat er samenhang tussen de aangemelde standaard en standaarden die in aanmerking komen voor opname op de lijst en wat betekent dit voor de toetsing van de standaard(en)? (Denk bijvoorbeeld ook aan een gezamenlijke toetsing met (een deel van) deze aanvullende standaarden).*

Er zijn geen standaarden bekend die in aanmerking komen voor opname op de lijst, die een relatie hebben met SHACL en derhalve voor gemeenschappelijke toetsing in aanmerking komen.

6. Sponsorschap

De aanmelding van standaarden voor de lijst van het Forum en het Nationaal Beraad dient ondersteund of gesponsord te worden door overheids- en/of (semi)publieke organisaties die de standaard reeds in gebruik hebben (of voornemens zijn dit te doen) en die de beoogde

opname op de lijsten ondersteunen. Dit draagt bij aan het draagvlak voor de standaard, geeft zicht op de functionele use case voor de overheid en helpt bovendien om tijdens de toetsing de juiste experts te benaderen.

1. Welke overheden en/of (semi)publieke organisaties ondersteunen de aanmelding van de standaard?

Het Kadaster ondersteunt de aanmelding van de standaard en is concreet aan de slag met de implementatie ervan. Het Kadaster is tevens medeoprichter van PLDN.

2. Hebben deze organisaties de standaard geïmplementeerd? (zie ook punt 7 voor een uitwerking).

Het Kadaster is begonnen met het implementeren van de standaard en ziet toepassingsmogelijkheden bij het beschikbaar stellen van geo informatie. Daarbij wordt zowel het open data domein verkend (onder meer PDOK) als de tarief gebaseerde dienstverlening van het Kadaster. Aangevuld met eerder genoemde ontwikkelingen bij het CBS, RCE en Nationale Politie.

7. Functionele use case

Voor de standaard dient een duidelijke use case beschikbaar te zijn op basis waarvan overheden en/of instellingen uit de (semi) publieke sector kunnen bepalen of de aangemelde standaard voor hen relevant is en wie eventueel moet deelnemen aan de experttoetsing van de standaard.

De enige bekende use case betreft het Kadaster die SHACL inzet voor kwaliteitsborging en dienstverlening. De indiener stelt hierover het volgende:

“SHACL wordt gebruikt binnen het Kadaster, en daar is ook een business case voor. Zoals veel organisaties, beschikt het Kadaster over meerdere datasets die op verschillende plekken in de organisatie gebruikt worden. Deze datasets ondergaan vaak verschillende ETL (Extract Transform Load) processen. Om te borgen dat de gegevens kwaliteit zo hoog mogelijk blijft, wordt SHACL ingezet. Dankzij SHACL is het Kadaster verzekerd dat de data wat binnen komt, aan dezelfde regels voldoet als wat er uit gaat. Bovendien wordt het SHACL model, als een ‘bijsluiter’ (metadata) bij de data geleverd. Dit, omdat het helpt bij het begrijpen van de informatie. Het SHACL model kan eenvoudig gevisualiseerd worden en is daarmee interpreteerbaar voor eenieder, zonder voorkennis van Linked Open Data. Op deze manier voegt het Kadaster waarde toe aan haar diensten.”

Het is denkbaar dat bij bewezen toegevoegde waarde op termijn de inzet van SHACL uitbreidt naar de andere informatiehuizen in het digitaal stelsel voor de omgevingswet, CBS, RCE en nationale politie.