

ROSA Architectuurscan/advies: Ontsluiting RIO-gegevens



Voor	Architectuurraad
Van	Bureau Edustandaard
Scan uitgevoerd door	Joeri van Es, Remco de Boer
Versie	2e concept
Datum	30 juni 2020
Versiehistorie	1e concept: opgesteld door BES 2e concept: afgestemd met de indiener en direct betrokkenen definitief: behandeld door Architectuurraad
Aanleiding	
Betreft	
Brondocument(en)	Project Start Architectuur RIO-Linked Open Data - Tweede fase, v1.i, 9 maart 2020 Technisch Infrastructuur Ontwerp Kadaster <-> DUO Koppeling, v1.2, 28 november 2019 Ontwerp Koppeling DUO - Upload API Kadaster DUO <-> Kadaster Koppeling, v0.5, 5 maart 2020
Begeleidende documenten	MEMO Verzoek ROSA Scan

Inleiding

Met de ROSA Architectuurscan worden op systematische wijze alle architectuuraspecten van een bij Edustandaard ingebracht onderwerp in kaart gebracht en worden knelpunten en kansen gesignaleerd. Niet alleen kan de indiener er zijn voordeel mee doen, ook kan ROSA ermee worden verrijkt. En tot slot stelt het andere ketenpartijen in staat om kennis te nemen van architectuurwijzigingen en het belang hiervan voor de eigen organisatie of achterban te bepalen (transparantie in de keten, informatiepositie).

Dit formulier bevat de uitkomst van een architectuurscan van de Ontsluiting van RIO-gegevens. Voor de indiener biedt de scan concrete handvatten voor toepassing van ROSA, en de mogelijkheid om lessen en ervaringen uit het project terug te koppelen aan ROSA. Een architectuurscan wordt in principe uitgevoerd met een hoge mate van betrokkenheid van vertegenwoordigers van de inbrenger. Deze wordt hierbij ondersteund door Bureau Edustandaard, de beheerder van ROSA. De inbrenger zou zich moeten herkennen in de uitkomsten.

Iedere architectuurscan begint met de vraag: welke onderdelen van ROSA zijn relevant voor het ingebrachte onderwerp, en indien relevant, op welke wijze? Vervolgens worden de vragen gesteld hoe het ingebrachte past op wat in ROSA is uitgewerkt, en of het project wellicht inzichten heeft die kunnen leiden tot verbetering of uitbreiding van ROSA. De antwoorden op deze vragen worden verwoord in termen van een advies richting zowel inbrenger, als richting ROSA zelf. De opzet van het advies is dat per onderdeel van ROSA uitspraken worden gedaan over:

1. Bevindingen uit project: *wat zegt het project zelf over het verband met ROSA van het ingebrachte onderwerp?*
2. Relatie met ROSA: *hoe verhoudt het ingebrachte zich tot ROSA¹?*
3. Voorgesteld advies van de Architectuurraad aan het project: *tips, verbeterpunten, en ook bekrachtiging dat er goed werk is geleverd vanuit het perspectief van ROSA²*

Adviezen in deze kolom zijn, gegroepeerd in 'PRODUCT' en 'CONTEXT'. De PRODUCT-adviezen bestrijken sec het ingediende 'product', d.w.z. de Ontsluiting van RIO-gegevens. Deze adviezen zijn direct gericht aan de project(deel)groep die zich met de totstandkoming van de Ontsluiting van RIO-gegevens bezighoudt. De CONTEXT-adviezen hebben betrekking op de context waarbinnen de Ontsluiting van RIO-gegevens toegepast gaat worden. Deze adviezen kunnen gericht zijn aan het project zelf, maar kunnen ook zijn gericht aan partijen die zich in die context bevinden, zoals de project(deel)groep die zich richt op de implementatie van de uiteindelijke Ontsluiting van RIO-gegevens, maar ook (sector)organisaties die met de uiteindelijke implementatie te maken gaan krijgen.

4. **Voorgesteld advies voor de Architectuurraad voor plaatsing onderwerpen op de ROSA architectuur backlog:** *wat kan ROSA doen om in het vervolg een betere ondersteuning te bieden aan dit project, en andere?*

Samenhang met andere formulieren:



- **Pitch Architectuurscan:** Het doel van de architectuurpitch is om een eerste indruk te krijgen van een ketenafspraken . Op basis van de pitch en de aangeleverde documentatie voert Bureau Edustandaard een architectuurscan uit. Voor de leden van de Architectuurraad (en andere geïnteresseerden) verduidelijkt deze pitch de context van de afspraak en de resultaten uit de architectuurscan.
- **ROSA architectuurscan bevindingen:** aan het invullen van het adviesdeel van een architectuurscan (dit formulier) gaat het verzamelen van feitelijke informatie, en het analyseren daarvan, vooraf. Die informatie, en de analyses, worden vastgelegd in het bevindingendeel van de architectuurscan. De lezer van het adviesdeel kan die erop na slaan als hij wil weten hoe het advies tot stand is gekomen. Het lezen van het bevindingendeel is niet vereist om het adviesdeel te begrijpen. Waar van toepassingen verwijst het bevindingendeel naar specifieke locaties van de brondocumenten die als input dienden voor de architectuurscan. Ook het lezen van de brondocumenten is niet vereist om het adviesdeel te begrijpen.



¹ De verhouding tussen het ingediende en de ROSA wordt per onderdeel uitgedrukt in een 'level of conformance' ontleend aan TOGAF, zie de bijlage.


² Dit is een concept advies, de uitkomsten worden eerst door de Architectuurraad besproken.

ROSA Architectuurscan/advies: Ontsluiting van RIO-gegevens

ROSA-onderdeel	Bevindingen uit project: Ontsluiting van RIO-gegevens	Relatie met ROSA (blauw: ROSA, geel: Ontsluiting van RIO-gegevens)	Voorgesteld advies aan project	Voorgesteld advies aan AR voor plaatsing onderwerpen op de architectuurbacklog ROSA
Werkingsgebied	Het werkingsgebied van RIO-LOD omvat het gehele onderwijsdomein (§2.1, p.4 PSA, § 2.3 Scope PSA). Zie ook de eerder uitgevoerde scan op het RIO register en beheer (https://www.wikixl.nl/wiki/rosa/index.php/Architectuurscan_RIO_register_en_beheer).	 Compliant - Het werkingsgebied van RIO-LOD omvat (net als het RIO register) het gehele onderwijsdomein	PRODUCT: CONTEXT:	

<p>Toepassings-gebied</p>	<p>Het ontsluiten van RIO-gegevens als Linked Open Data.</p> <p>Doel is tweeledig:</p> <ol style="list-style-type: none"> 1. Min of meer gestandaardiseerde halfproducten bieden zodat willekeurige applicatiebouwers effectief en efficiënt eindproducten kunnen samenstellen voor een veelheid van stakeholders als burgers, onderwijsaanbieders en anderen. (§2.1 PSA) -> <i>developersperspectief</i> 2. Optimale ontsluiting voor externe afnemers, op 1 manier aangeboden voor iedereen met een ontsluiting zo dicht mogelijk bij de bron (memo). -> <i>afnemersperspectief</i> <p>Er wordt wel gehint richting het afnemersperspectief in bijvoorbeeld de beheerarchitectuur (p.28), maar er worden geen specifieke afnemers benoemd en welke interactie die hebben met de linked data.</p> <p>De vorm is ontsluiting van RIO als Linked Open Data (LOD), waarmee RIO voor iedereen toegankelijk is.</p>	 <p>Compliant -</p> <p>RIO-LOD heeft direct betrekking op de ketenfuncties Informatieverwerking en Informatiestandaardisatie, en gedeeltelijk op Identificatie en toegang en Informatiebeveiliging en privacy Vanuit afnemersperspectief zijn alle overige ketenfuncties ook relevant, immers op alle plaatsen waar instellings- en opleidingsinformatie gebruikt wordt is RIO in beeld als de bron van die data.</p>	<p>PRODUCT:</p> <p>Werk in de PSA het afnemersperspectief verder uit: Breng afnemers en hun mogelijke interactie met de RIO LOD goed in kaart. Wat moet de afnemer 'kunnen' om hier gebruik van te maken? (betreft o.a. De structurering/interpretatie van LOD gerelateerd aan het informatiemodel dat ook in LOD staat)</p> <p>Dit geeft ook een beeld van eventuele haalbaarheidsrisico's bij grootschalige implementatie.</p> <p>CONTEXT:</p> <p>-></p>	
<p>Ontwerpergebied</p> <p>Bovensectorale samenwerking</p>	<p>§ 2.2 PSA: REST API Design Rules (aka Nederlandse API Strategie); Forum Standaardisatie. Daarmee werkt dit project toe naar het publiceren van zogenaamde sectorale registers in lijn met de stelselcatalogus van de overheid, naast de basisregistraties als BPR en BAG.</p> <p>Externe gebruikers kunnen realtime details ophalen op basis van een vastgehouden (gecachte) webidentiteit/URI (§ 2.4 PSA) -></p>	 <p>Consistent -</p> <p>Hoewel niet expliciet benoemd, is de relatie met de ROSA-principes "Een gezamenlijke basisinfrastructuur" en "Eenheid in verscheidenheid" af te leiden uit de PSA .</p> <p>RIO-LOD ondersteunt eenmalige registratie, meervoudig gebruik van</p>	<p>PRODUCT:</p> <ul style="list-style-type: none"> • Verwijs expliciet naar de relevante ROSA-principes. • Onderzoek mogelijke effecten op (interne) processen van afnemers. <p>CONTEXT:</p> <ul style="list-style-type: none"> • Bepaal de mogelijke impact van RIO-LOD op de eigen (interne) processen. 	<ul style="list-style-type: none"> • Bewaak de relatie tussen ontwikkelingen op het vlak van REST (w.o. Edukoppeling REST-profiel) en de Nederlandse API Strategie en/of andere relevante design rules

	<p>Enmalige registratie, meervoudig gebruik; data wordt niet rondgepompt</p> <p>In de PSA (§2.2) wordt de hoofddoelstelling van RIO LOD benoemd. Deze kan worden samengevat als het verbeteren van de “Developer Experience (DX)”.</p> <p>“Scholen beschrijven zelf hun eigen wereld” (PSA §2.2) -> eenheid in verscheidenheid</p> <p>Effecten op (interne) processen van de afnemers van de gegevens zijn niet omschreven in de PSA. (-> koppelen niet kantelen)</p>	<p>instellings- en opleidingsaanbodgegevens.</p> <p>“Compliance” met NORA en ROSA wordt genoemd als een “business driver” in §2.2 van de PSA. Dit sluit aan bij de ROSA basisprincipes: “Aansluiten bij NORA” en “Digitaal doen we het zo”.</p> <p>DUO zorgt voor aansluiting op Basisregistraties (PSA §3.1.1)</p>		
<p>Ontwerpegebied</p> <p>Informatie-beveiliging en privacy (IBP)</p>	<p>Privacygevoelige of anderszins vertrouwelijke data zijn buiten scope (PSA §2.3)</p> <p>Data moeten ‘van goede kwaliteit’ zijn. (PSA §2.2). Daaronder valt in ieder geval de integriteit van de gegevens.</p> <p>Een eis voor de LOD API is een hoge standaard beschikbaarheid. (PSA §2.4) Dit sluit aan bij het ontwerpkader “Continuïteit van de dienstverlening”.</p>	 <p>Compliant -</p> <p>Privacygevoelige informatie wordt buiten Scope verklaard (PSA §2.3) voor RIO LOD.</p> <p>Er wordt aangegeven dat data van goede kwaliteit moet zijn. (PSA §2.2) Hiervoor is ROSA ontwerpkader “Voorkom aantasting van integriteit” belangrijk, maar er wordt niet direct naar gerefereerd.</p> <p>Een eis voor de LOD API is een hoge standaard beschikbaarheid, maar deze wordt nog niet gehaald. Wat een hoge beschikbaarheid inhoudt is nog niet genoeg omschreven.</p>	<p>PRODUCT:</p> <ul style="list-style-type: none"> • Verduidelijk de kwaliteitseisen (ihb voor de aspecten B en I) en de bijbehorende maatregelen. • Geef aan welke ‘Quality of Service’ afnemers standaard mogen verwachten (buiten evt. aanvullende afspraken in SLA). Omschrijf daarbij wat een “hoge standaard beschikbaarheid van de API” inhoudt. <p>CONTEXT:</p>	
<p>Ontwerpegebied</p> <p>IAA</p>	<p>IAA speelt niet voor raadplegen, wel voor beheer van RIO-data.</p> <p>In §3.3 van de PSA wordt gesteld dat data zonder beperkingen door iedereen mag</p>	 <p>Irrelevant -</p> <p>Irrelevant - Het wijzigen van RIO data is niet in scope van deze scan, aangezien het hier alleen gaat over de ontsluiting van RIO-LOD.</p>	<p>PRODUCT:</p> <p>CONTEXT:</p>	

	worden gebruikt, maar dat wijzigingen niet gedaan mogen worden door onbevoegden.			
<p>Ontwerpgebied</p> <p>Gegevens-uitwisseling in de keten</p>	<p>Gegeven uit RIO worden via 3 routes ontsloten:</p> <ol style="list-style-type: none"> 1. REST-API / LOD (fair use bevragingen; niet voor synchronisatie van gehele register) 2. SPARQL-endpoint (directe bevragingen op LOD; kan tijdelijk worden afgekoppeld bij te grote bevragingen) 3. Webservice direct op RIO (niet via LOD; specifiek voor één use case met near-realtime operationele eisen) <p>PSA §2.2: Niet alleen de data, ook het logisch model van RIO wordt, in samenhang, gepubliceerd op het web. Allereerst betekent dit dat de data zelfverklarend is. Herinterpretaties en misverstanden van de data worden tot een minimum beperkt. Verder zijn ook de integriteitsregels voor de betreffende data voor iedereen waarneembaar en bruikbaar om de juistheid, volledigheid en actualiteit van de data geautomatiseerd te valideren. Scholen beschrijven zelf hun eigen wereld (herkenbare namen, daadwerkelijke locaties, opleidingsprofielen e.d.) en leveren dat aan bij RIO. Door middel van deze ontsluiting wordt dat direct zichtbaar op het web of middels een RIO raadplegende applicatie. Hierdoor kan de school zijn eigen werk controleren. En dat kan in feite iedereen. De datakwaliteit verbetert met een feedbackloop.</p> <p>paragraaf 3.3: "Gebruikers mogen verwachten dat ze webapplicaties kunnen bouwen" -> alleen webapplicaties? Geen gebruik van RIO-data in andere systemen / datalandschap van school?</p>	 <p>Consistent -</p> <p>Ontsluiting van RIO gegevens via Linked Open Data geeft invulling aan het ROSA ontwerp kader "Openbare registergegevens worden ontsloten als Linked Open Data". Voor één use case is een aparte webservice rechtstreeks op RIO ingericht buiten LOD om.</p> <p>De LOD-URIs kennen een eigen levenscyclus. Afnemers moeten daarop voorbereid zijn (ihb wanneer URIs vervallen).</p> <p>Semantiek uit RIO-modellen heeft raakvlakken met semantiek in andere uitwisselingen. Linked data kan een manier zijn om die semantiek ook over uitwisselingen heen traceerbaar te maken. Hier ligt een relatie met bijvoorbeeld AMIGO ('virtuele modellenrepository').</p>	<p>PRODUCT:</p> <ul style="list-style-type: none"> - Geef inzicht of en op welke termijn en wijze near-realtime actualiteitsniveaus kunnen worden gerealiseerd, inclusief de implicaties voor partijen die nu nog via de workaround (Webservice) gegevens uit RIO betrekken. - Maak expliciet welke datakwaliteitsaspecten, zoals juistheid volledigheid en actualiteit, belangrijk zijn en hoe het beoogde kwaliteitsniveau wordt gerealiseerd, zodat voor afnemers duidelijk is welke verwachtingen zij mogen hebben t.a.v. de ontsloten data. <p>CONTEXT:</p> <ul style="list-style-type: none"> - Zorg dat je als partij die nu nog gebruik maakt van statisch gepubliceerde BRIN/RIO-data, zo veel als mogelijk (waar relevant) voorbereid op aansluiting op de dynamische LOD. 	<p>Bepaal hoe mbv Linked Data traceerbaarheid van semantiek tussen uitwisselingen kan worden gefaciliteerd. Hiervoor loopt inmiddels een verkenning, op initiatief van DUO icm AMIGO / Kennisnet.</p> <p>Doe, in de context van het ontwerp kader "Openbare registergegevens worden ontsloten als Linked Open Data", een uitspraak over welke verwachtingen ketenpartijen redelijkerwijs zouden mogen hebben t.a.v. actualiteit, persistentie en inhoud van openbare registergegevens die als Linked Open Data worden ontsloten. In het bijzonder waar het processen betreft met near-realtime operationele eisen en/of processen die gericht zijn op (zeer) lange termijn (e.g., archivering).</p>

Voor specifieke toepassingen die hoge eisen stellen aan actualiteit (met name voorziening VVA) is op dit moment een (tijdelijke?) alternatieve workaround via onderwijsdata.duo.nl.

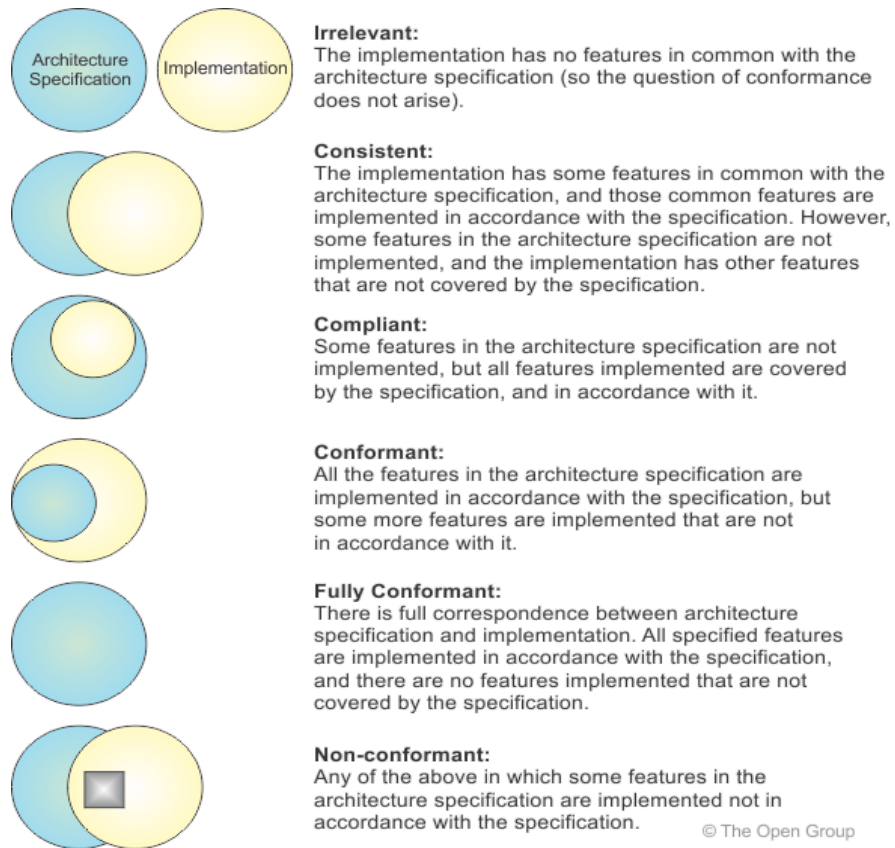
In §2.4 van de PSA wordt genoemd dat het doel is dat ontsloten data hoogst actueel is. Naamswijzigingen, het sluiten van een fixus opleiding en dergelijke in RIO zijn dan op secondeschaal toegankelijk. Uit toelichtende gesprekken blijkt dit in praktijk lastig te realiseren. Een actualiteit op minuten- in plaats van secondeschaal is realistischer.

De PSA benoemt datakwaliteit als speerpunt (p. 5). De PSA duidt het begrip 'datakwaliteit' slechts beperkt, refererend aan een 'feedback loop' (p.6) waarmee de datakwaliteit verbetert. Verder wordt daarover opgemerkt dat de integriteitsregels voor de betreffende data voor iedereen waarneembaar en bruikbaar zijn om de juistheid, volledigheid en actualiteit van de data geautomatiseerd te valideren.

<p>Keten-processen</p>	<p>Met het ontsluiten van RIO data als LOD wil DUO de “developer experience” verbeteren voor ontwikkelaars. (PSA §2.2) Aangezien RIO data een ondersteunende rol speelt bij vrijwel alle onderwijsprocessen (PSA §2.1) zal dit vrijwel alle ketenprocessen raken.</p> <p>In de memo wordt het doel genoemd “optimale ontsluiting voor externe afnemers”, dit doel wordt niet direct genoemd in de PSA. Het wordt niet duidelijk uit de PSA of uit de memo hoe deze twee doelen in relatie met elkaar staan.</p> <p>Developers zullen niet de enige afnemers zijn, want er is een impact op alle ketenprocessen. Alleen in strikt technische zin zijn de developers de ‘afnemers’, want zij ‘praten’ tegen de API aan. De eindgebruikers zien het resultaat daarvan, maar niet de technische implementatie.</p>	 <p>Onbepaald - Met het ontsluiten van RIO data als LOD wil DUO de “developer experience” verbeteren voor ontwikkelaars. (PSA §2.2) Aangezien RIO data een ondersteunende rol speelt bij vrijwel alle onderwijsprocessen (PSA §2.1) zal dit vrijwel alle ketenprocessen raken.</p> <p>Het project wordt vooral gepositioneerd als technologische verandering, effecten op (interne) processen van de afnemers van de gegevens worden buiten beschouwing gelaten.</p>	<p>PRODUCT: Beschrijf hoe de doelen “optimale ontsluiting voor externe afnemers” en “developer experience” met elkaar in relatie staan.</p> <p>CONTEXT:</p>	
<p>Zeggenschappen en gegevenssoorten</p>	<p>Zeggenschap over gegevens ligt bij onderwijsinstellingen zelf (PSA §2.2).</p>	 <p>Compliant - Zeggenschap over gegevens ligt bij onderwijsinstellingen zelf (PSA §2.2). Dit ligt aan de kant van het RIO-register, niet bij RIO-LOD,</p> <p>RIO LOD heeft verder betrekking op de zeggenschappen “Beschikbaar stellen” en “Inzien”. DUO stelt RIO data beschikbaar via Linked data endpoints aan iedere afnemer die gebruik wil maken van deze data.</p>	<p>PRODUCT:</p> <p>CONTEXT:</p>	

Referentiecomponenten en applicaties	<p>§1.2 p.3 PSA Het vervangen van bestaande ontsluitingen op basis van BRIN.</p> <p>Vanwege de zin “Gebruikers mogen verwachten dat ze webapplicaties kunnen bouwen” (paragraaf 3.3 van de PSA) lijkt de ontsluiting van RIO-gegevens zich te beperken tot webapplicaties. Er zijn echter ook andersoortige systemen binnen het applicatielandschap van een onderwijsinstelling waarvoor RIO-data relevant kan zijn.</p>	 <p>Compliant - Op alle plaatsen waar nu nog (informatie uit) BRIN wordt gebruikt, zal RIO gebruikt moeten gaan worden.</p>	<p>PRODUCT: Geef aan voor welke toepassingen (naast webapplicaties) RIO LOD bedoeld is. Zijn er toepassingen waar het expliciet NIET voor bedoeld is? Hoe komen die toepassingen aan hun RIO-gegevens?</p> <p>CONTEXT: Alle systemen die nu gekoppeld zijn aan BRIN (of op een andere manier BRIN-data ontvangen) zullen moeten veranderen. Is LOD de enige ontsluitingsvorm?</p>	<p>Werk de veranderingen aan het (referentie)architectuurlandschap uit, die de vervanging van BRIN door RIO en de ontsluiting van RIO als LOD veroorzaakt.</p>
Architecturele randvoorwaarden			<p>PRODUCT:</p> <p>CONTEXT:</p>	
Governance	<p>Afneemers en hun belangen zijn nog niet in kaart gebracht. In de memo staat wel als één van de vervolgstappen genoemd “Stakeholders en hun producten en verwachtingen leren kennen”.</p>	 <p>Nonconformant - Uitgangspunt voor het ROSA ontwerpgebied governance is dat alle belangen worden afgewogen; daarvoor moeten alle belangen en belanghebbenden in kaart zijn gebracht.</p>	<p>PRODUCT: Maak afneemers en hun belangen expliciet. Verduidelijk de rol van belanghebbenden in de (door)ontwikkeling van de ontsluiting van RIO-gegevens..</p> <p>CONTEXT:</p>	
Implementatie	<p>Er worden pilots met scholen uitgevoerd. De uitvoering, status en resultaten van die pilots zijn in de PSA niet verder beschreven.</p>		<p>PRODUCT: Maak expliciet wanneer implementatie van RIO-LOD ‘succesvol’ is. (Technische werking vs. gebruik, en hoe / waar / door wie?)</p> <p>CONTEXT:</p>	

Bijlage 1: ARCHITECTURE COMPLIANCE (TOGAF)



Een Nederlandse vertaling van de beschrijving van de TOGAF-categorieën:

- a. **irrelevant** = er is geen relatie tussen het ingebrachte en ROSA
- b. **consistent** = er is overlap tussen het ingebrachte en ROSA, en binnen die overlap is het ingebrachte conform ROSA gerealiseerd, de overlap is echter niet **volledig** = sommige specificaties van ROSA zijn niet overgenomen, en het ingebrachte heeft onderdelen die niet door ROSA worden gedekt.
- c. **compliant** = het ingebrachte valt volledig binnen ROSA (subset) en is conform ROSA gerealiseerd
- d. **conformant** = ROSA dekt alleen een deel van het ingebrachte, maar dat deel is wel conform ROSA gerealiseerd
- e. **fully conformant** = ROSA dekt het geheel van het ingebrachte, en niets van het ingebrachte valt buiten ROSA
- f. **non-conformant** = er is overlap tussen het ingebrachte en ROSA, en binnen die overlap is er iets van het ingebrachte *niet* conform ROSA gerealiseerd

Bron: http://pubs.opengroup.org/architecture/togaf9-doc/arch/Figures/48_conformance.png