**Gegevenskwaliteit RIO**

**Controles**

Wiebe Buising, 30 augustus 2018

CONCEPT versie 0.3

1. **Inleiding**

Aanleiding

In de projectgroep RIO van 3 juli 2018 is gesproken over de wijze waarop de kwaliteit van de gegevens in het register RIO wordt geborgd, mede naar aanleiding van de ROSA architectuurscan naar het register RIO en het beheer van de informatiemodellen. Het is de bedoeling dat RIO een rol krijgt in vele ketenprocessen in het onderwijs, en dat leidt tot vragen over de kwaliteit van de gegevens in RIO. In het gesprek in de projectgroep kwam onder andere de mogelijke rol van controles bij het borgen van de gegevenskwaliteit aan de orde.

Kwaliteitsmodel PBLQ

Bij het opstellen van deze notitie is gebruik gemaakt van de notitie over de kwaliteit van RIO die PBLQ op verzoek van OCW heeft gemaakt. PBLQ hanteert een kwaliteitsmodel, dat in deze notitie wordt gebruikt als kader voor de uitwerking van voorstellen gericht op de gegevenskwaliteit van RIO.

ROSA zeggenschappen, gegevensverantwoordelijke en registerhouder

De projectgroep RIO heeft in 2015 aan de hand van het onderscheid tussen de gegevensverantwoordelijke en de registerhouder gesproken over de verschillende implementatiemogelijkheden van RIO. Deze notitie sluit aan bij de keuzes die hierover in 2015 zijn gemaakt.

Deze notitie is als volgt opgebouwd:

* In paragraaf 2 wordt het kwaliteitsmodel van PBLQ kort beschreven;
* In paragraaf 3 worden de eerder gemaakte keuzes inzake het beleggen van de ROSA zeggenschappen bij de gegevensverantwoordelijke en de registerhouder beschreven;
* In paragraaf 4 wordt ingegaan op de gegevenskwaliteit van de juridische werkelijkheid;
* In paragraaf 5 wordt nader ingegaan op de gegevens in de onderwijskundige werkelijkheid en worden drie types controles onderscheiden en toegelicht;
* In paragraaf 6 worden kort de controles in het kader van de koppeling van RIO en BRON beschreven.

1. **Kwaliteitsmodel PBLQ**

Om grip te krijgen op het begrip kwaliteit maakt PBLQ onderscheid tussen de kwaliteit van het informatiemodel en van de uitgevoerde processen: de bijhouding en het gebruik.

Modelkwaliteit

De kwaliteit van informatiemodellen voor registers kent drie aspecten:

* Beeldt het model de werkelijkheid goed af (*empirische modelkwaliteit*). De BRP is gericht op ingezetenen en kan slecht omgaan met personen zonder vaste woon- of verblijfplaats die in de werkelijkheid wel degelijk voorkomen. RIO is er gekomen omdat de bekostigingswerkelijkheid onvoldoende aansloot bij de maatschappelijke werkelijkheid.
* Sluit het model voldoende aan bij de juridische en organisatorische werkelijkheid van de partijen die betrokken zijn bij vastlegging en gebruik (*institutionele modelkwaliteit*).
* Is het model consistent, accuraat en volledig beschreven, en voldoet het aan relevante richtlijnen (*de technische kwaliteit van het model*).

Proceskwaliteit

De kwaliteit van de processen kan in twee delen worden gesplitst:

* Verloopt de bijhouding conform de criteria (*kwaliteit van registratieprocessen*)? Dit betreft zowel de feitelijke registratiehandelingen als het beheer op de verzameling van gegevens in het register (bijvoorbeeld het doorvoeren van modelaanpassingen in de data, of het waarborgen van de integriteit van opgeslagen gegevens).
* Worden gegevens uit de registratie op de juiste wijze gebruikt in de uitvoeringsprocessen van de afnemers (*kwaliteit van gebruiksprocessen*)?



1. **ROSA zeggenschappen, gegevensverantwoordelijke en registerhouder**

In de ROSA worden de volgende zeggenschappen onderscheiden:

* Accorderen uitwisseling. Deze zeggenschap bepaalt welke partij toestemming moet geven voordat bepaalde gegevens in de keten mogen worden uitgewisseld.
* Beschikbaar stellen. Deze zeggenschap bepaalt welke partij in een ketenproces bepaalde gegevens beschikbaar stelt. Vanuit het principe “eenmalige registratie meervoudig gebruik” zou voor elke gegevenssoort steeds één en slechts één partij de bron van gegevens moeten zijn. Staande praktijk en wettelijke bepalingen maken echter dat gegevens nog vaak op verschillende plaatsen geregistreerd worden en (dus) door verschillende partijen beschikbaar worden gesteld.
* Corrigeren en bijwerken. Deze zeggenschap bepaalt welke partij de bevoegdheid heeft om gegevens aan te (laten) passen.
* Inwinnen. Deze zeggenschap bepaalt welke partijen bepaalde gegevens mogen inwinnen ter verdere verwerking.
* Inzien. Deze zeggenschap bepaalt welke partijen de gegevens mogen inzien.
* Registreren. Deze zeggenschap bepaalt welke partij bepaalde gegevens registreert ten behoeve van de keten. Zie ook de opmerking over meervoudige registratie bij Beschikbaar stellen
* Vernietigen. Deze zeggenschap bepaalt welke partij de bevoegdheid heeft om gegevens te vernietigen. Vaak ligt aan de invulling van deze zeggenschap wettelijke bepalingen met betrekking tot bewaartermijnen ten grondslag.

De gegevensverantwoordelijke

* is verantwoordelijk voor de juistheid van “zijn” (deelverzameling) van gegevens: *corrigeren en bijwerken*
* kan een eigen registratie voeren: *registreren*
* bepaalt wanneer gegevens vernietigd dienen te worden: *vernietigen*
* stelt “zijn” gegevens beschikbaar aan de registerhouder en of het publiek: *beschikbaar stellen*

De registerhouder

* wint gegevens in bij de gegevensverantwoordelijke(n): *inwinnen*
* voert de registratie van ingewonnen gegevens: *registreren*
* stelt deze gegevens beschikbaar aan afnemers c.q. het publiek: *beschikbaar* *stellen*

Omdat het bij RIO om open data gaat, mag iedereen de gegevens *inzien* en is het *accorderen van uitwisseling* van gegevens niet van toepassing.

Er zijn verschillende inrichtingsvarianten mogelijk, afhankelijk van de mate van (de)centralisatie van de registratie en het beschikbaar stellen van de gegevens.



Voor RIO is gekozen voor een centrale voorziening (het register RIO), waarbij DUO de registerhouder is. Een aantal RIO gegevens wordt centraal geregistreerd, deze gegevens worden centraal beschikbaar gesteld.

1. **Juridische werkelijkheid**

In de juridische werkelijkheid voert DUO een aantal processen uit, waarvan de uitkomsten door DUO vanaf 2019 in het register RIO worden vastgelegd (nu nog in het register BRIN). Deze processen starten over het algemeen met een aanvraag van een onderwijsinstelling. Het gaat om de volgende processen:

* Erkennen van bevoegd gezag (wie mag)
* Erkennen van instellingen (wie mag)
* Erkennen van vestigingen (wie mag waar) (alleen voor vo, niet voor mbo)
* Erkennen van opleidingen (wat mag)
* Toekennen van licenties (wie mag wat (waar) doen)

Gegevensverantwoordelijke en registerhouder

In het kader van de juridische werkelijkheid is DUO zowel gegevensverantwoordelijke als registerhouder. Als gegevensverantwoordelijke is DUO verantwoordelijk voor de juistheid van de gegevens en voor het vernietigen van de gegevens. Als gegevensverantwoordelijke en tevens registerhouder voert DUO een registratie van de gegevens en stelt DUO de gegevens beschikbaar.

Modelkwaliteit

* *Empirische modelkwaliteit*. In het kader van de juridische werkelijkheid dient het model deze juridische werkelijkheid goed weer te geven. Dat is het geval. (Juist omdat de juridische werkelijkheid maar een deel van de werkelijkheid omvat, is RIO er gekomen. Naast de juridische werkelijkheid is in RIO ook de onderwijskundige (maatschappelijke) werkelijkheid gemodelleerd).
* *Institutionele modelkwaliteit*. Het model sluit voldoende aan bij de juridische en organisatorische werkelijkheid van de partij (DUO) die betrokken is bij vastlegging en gebruik.
* *Technische kwaliteit van het model.* Is het model consistent, accuraat en volledig beschreven, en voldoet het aan relevante richtlijnen?

Proceskwaliteit

* *Kwaliteit van registratieprocessen.* De registratiehandelingen en het beheer op de verzamelingen van gegevens worden uitgevoerd door DUO. Nadere uitwerking door DUO.
* *Kwaliteit van gebruiksprocessen.* Een zeer belangrijke afnemer van de gegevens uit de registratie is de afdeling bekostiging van DUO. De gegevens worden op de juiste wijze gebruikt voor de bekostiging van de onderwijsinstellingen. Nadere uitwerking door DUO.

1. **Onderwijskundige werkelijkheid**

**Instellingen**

De volgende objecten (met betrekking tot de onderwijsinstellingen) uit het informatiemodel RIO worden in het register RIO geregistreerd:

* Onderwijsbestuur/Bevoegd Gezag (wie)
* Onderwijsaanbieder (wie)
* Onderwijslocatie (waar)
* Aangeboden opleidingen (wie wat waar)
* Communicatiecontexten (wie)

Gegevensverantwoordelijke

De onderwijsinstellingen zijn de gegevensverantwoordelijke voor het onderwijsbestuur/bevoegd gezag, de onderwijsaanbieders, onderwijslocaties, aangeboden opleidingen en communicatiecontexten. Als gegevensverantwoordelijke zijn de onderwijsinstellingen verantwoordelijk voor de juistheid van de gegevens en voor het vernietigen van de gegevens. Als gegevensverantwoordelijke voeren de onderwijsinstellingen een registratie van de gegevens en stellen ze de gegevens beschikbaar aan de registerhouder.

Registerhouder

DUO is de registerhouder voor het onderwijsbestuur/bevoegd gezag, de onderwijsaanbieders, onderwijslocaties, aangeboden opleidingen en communicatiecontexten. Als registerhouder voert DUO een registratie van deze gegevens en stelt DUO de gegevens beschikbaar.

Modelkwaliteit

* *Empirische modelkwaliteit*. In het kader van de onderwijskundige werkelijkheid dient het model deze werkelijkheid goed weer te geven. Of dit het geval is, zal moeten blijken na de implementatie van RIO. Bij de modellering van de onderwijskundige werkelijkheid en de voorbereiding van de implementatie van RIO is veel aandacht besteed aan het aansluiten bij de werkelijkheid van de onderwijsinstellingen. Onder andere door validatiebijeenkomsten met onderwijsinstellingen in alle sectoren van het onderwijs, uitwerking en toetsing van het informatiemodel in de werkgroepen RIO vo en mbo en de modelleersessies met de mbo-instellingen.
* *Institutionele modelkwaliteit*. Sluit het model voldoende aan bij de juridische en organisatorische werkelijkheid van de partijen die betrokken zijn bij vastlegging en gebruik? Nadere juridische borging van het vastleggen en gebruik van de gegevens uit de onderwijskundige werkelijkheid dient nog plaats te vinden.
* *Technische kwaliteit van het model.* Is het model consistent, accuraat en volledig beschreven, en voldoet het aan relevante richtlijnen?

Proceskwaliteit

* *Kwaliteit van registratieprocessen.* De onderwijsaanbieders, onderwijslocaties en aangeboden opleidingen worden door de instellingen beheerd in het eigen administratiesysteem en via de geautomatiseerde koppeling met het register RIO worden deze gegevens geregistreerd in het register RIO. Wijzigingen van deze gegevens worden in het eigen administratiesysteem doorgevoerd en via de koppeling in RIO verwerkt. De communicatiecontexten en het onderwijsbestuur/bevoegd gezag worden door de instellingen direct in het register RIO beheerd, door middel van de selfservice schermen.
* *Kwaliteit van gebruiksprocessen.* Het is de bedoeling dat de gegevens uit de onderwijskundige werkelijkheid door vele partijen in vele processen gebruikt gaan worden. Over de kwaliteit van dit gebruik kan nu nog geen oordeel worden gegeven.

**Opleidingen**

Een ander onderdeel van de onderwijskundige werkelijkheid betreft het wat: de opleidingseenheden (opleidingen, opleidingsonderdelen, opleidingsgroepen, opleidingsfasen), evenals de waardenlijst van opleidingsprogrammakenmerken.

Per sector moet bepaald worden wie verantwoordelijk is voor de inhoud van deze gegevens. In het mbo is dat bijvoorbeeld de SBB.

**Controles**

De volgende soorten controles worden in deze notitie onderscheiden:

* Technische controles
* Controles op de juistheid van de gegevens
* Controles aan de hand van wet- en regelgeving

Technische controles

In het register RIO, de administratiesystemen en de koppeling daartussen kunnen (en zullen) een aantal technische controles worden ingebouwd. Deze controles zullen worden beschreven in het op te stellen PvE-RIO. Voorbeelden van deze controles:

* Voldoet een gegeven aan het technische format.
* Als er een verplichte relatie is, moet deze aanwezig zijn bij de aanlevering aan het register RIO. Bijvoorbeeld de relatie tussen een onderwijsaanbieder en het bevoegd gezag (mbo) / onderwijsbestuur (vo)
* Het bestaan van een adres (DUO koppelt met de BAG).

Controles op de juistheid van gegevens

Met de juistheid van de gegevens wordt bedoeld of een instelling inderdaad de onderwijsaanbieders en onderwijslocaties heeft en de opleidingen aanbiedt, zoals ze zijn geregistreerd in het register RIO (en of de attributen bij deze objecten kloppen).

Als gegevensverantwoordelijke zijn de onderwijsinstellingen verantwoordelijk voor de juistheid van de gegevens. Er worden door de registerhouder DUO geen controles uitgevoerd op de juistheid van deze gegevens.

De controle op de juistheid van de gegevens in RIO ligt besloten in het gebruik van deze gegevens in allerlei processen en systemen in het onderwijs. Hoe breder dit gebruik, hoe sterker het zelfregulerend mechanisme dat leidt tot borging van de kwaliteit van deze gegevens. Bijvoorbeeld:

* De aangeboden opleidingen in RIO mbo zullen door de te ontwikkelen applicatie Centraal aanmelden mbo worden gebruikt. Als een instelling de aangeboden opleidingen niet goed heeft geregistreerd in RIO, kunnen aankomende studenten die zich willen aanmelden voor een opleiding bij die instelling, de betreffende opleiding mogelijk niet vinden, of een opleiding zien die niet wordt aangeboden. De instellingen zullen dit risico niet willen lopen, en er voor zorgen dat de aangeboden opleidingen correct in RIO zijn geregistreerd.
* Scholen in het vo willen RIO gebruiken voor de presentatie van hun gegevens op vensters voor verantwoording en scholen op de kaart. Scholen zullen er voor zorgen dat hun gegevens goed in RIO staan, zodat hun verantwoording naar de buitenwereld op basis van de juiste gegevens plaats vindt.

Het gaat hier om open data, die voor iedereen beschikbaar zijn. Stel dat er een instelling is die één onderwijsaanbieder registreert in RIO, terwijl uit de website van deze instelling duidelijk blijkt dat er meerdere onderwijsaanbieders zijn. Dit is dus voor iedereen zichtbaar. Er is niemand die de instelling kan dwingen om de registratie in RIO aan te passen. Tegelijkertijd kan verwacht worden dat er een proces van druk ontstaat, waardoor de instelling toch de registratie aanpast. Bijvoorbeeld omdat gebruikers van de gegevens uit het register RIO deze onderwijsinstelling gaan aanspreken.

Controles aan de hand van wet- en regelgeving

In RIO worden gegevens geregistreerd uit de juridische werkelijkheid en de onderwijskundige werkelijkheid en worden de relaties tussen deze beide werkelijkheden geregistreerd. Registraties in de juridische werkelijkheid kunnen eisen stellen aan registraties in de onderwijskundige werkelijkheid. Bijvoorbeeld: een instelling heeft een licentie om bepaalde opleidingen aan te bieden (toegekende licentie in de juridische werkelijkheid) en in RIO is te zien dat deze instelling ook een andere opleiding aanbiedt (aangeboden opleidingen in de onderwijskundige werkelijkheid).

Het ligt voor de hand dat dit soort controles door DUO worden uitgevoerd. Voorstel bij dit soort controles is dat het om controles achteraf gaat. Dit betekent dat niet al bij registratie van onderwijsaanbieders, onderwijslocaties en aangeboden opleidingen in RIO de controles (“aan de poort”) worden uitgevoerd. Deze controles worden periodiek, achteraf, uitgevoerd, waarbij in geval van geconstateerde inconsistenties DUO contact opneemt met de betreffende instelling. Als blijkt dat de instelling inderdaad iets doet wat niet mag, past de instelling de registratie in RIO aan.

**Kwaliteitsmetingen**

PBLQ adviseert om een palet van kwaliteitsmetingen in te richten en uit te voeren, om te kunnen sturen op de afgesproken kwaliteit. PBLQ noemt:

* *Procesaudits* uitvoeren om informatie over de kwaliteit van registratieprocessen te verzamelen.
* Controles aan de hand van *referentiegegevens*. Kent RIO evenveel bevoegde gezagen in de VO sector als de VO-raad bijvoorbeeld? Is er werkelijk gedurende zeer lange tijd geen enkele mutatie bij een gegeven opgetreden? Controles op mogelijke maar onwaarschijnlijke combinaties van gegevens in de registratie passen ook in deze categorie.
* *Steekproefsgewijs onderzoek* doen of de werkelijkheid correct is weergegeven. Dat is bewerkelijk en duur, maar wel een waardevolle manier om iets over de kwaliteit van de registratie te kunnen zeggen.
* Niet echt een meting, maar wel belangrijke input over de kwaliteit van de registratie is een *terugkoppeling* van de geregistreerde instellingen zelf. Om ze daartoe in staat te stellen dient geregeld te zijn dat ze de geregistreerde gegevens kunnen inzien en om correctie kunnen verzoeken wanneer de gegevens naar hun oordeel onjuist zijn. Bij het Digitaal Klantdossier was de ervaring dat de gegevenskwaliteit met sprongen steeg toen burgers hun eigen gegevens konden inzien.

Aan de projectgroep RIO de vraag hoe om te gaan met dit advies.

1. **Koppeling RIO-BRON**

De onderwijsaanbieder en de onderwijslocatie worden toegevoegd aan de inschrijving in BRON. BRON zal een aantal controles uitvoeren op deze gegevens, die er samengevat op neerkomen dat een onderwijsaanbieder en onderwijslocaties moeten zijn aangeleverd, dat deze moeten bestaan (zijn geregistreerd in RIO) en dat ze actief zijn (aan de hand van de inbedrijfDatum en uitbedrijfDatum).