

Scope Onderwijsgegevensplatform

Voor: Architectuurraad Edustandaard
Van: Remco de Boer, namens Bureau Edustandaard
Datum: 11 april 2022
Betreft: Scope Onderwijsgegevensplatform

Doel

Te komen tot een afbakening van de scope van het Onderwijsgegevensplatform, als opmaat voor een gedragen ketenarchitectuur.

Aanleiding

DUO heeft op verzoek van de Datakamer een eerste versie van een ketenarchitectuur Onderwijsgegevensplatform uitgewerkt. De Datakamer wil graag dat op basis hiervan architecten uit de onderwijsketen gezamenlijk in een werkgroep komen tot een gedragen ketenarchitectuur. In deze context heeft Bureau Edustandaard een ROSA-architectuurscan van deze versie van de ketenarchitectuur uitgevoerd.

De belangrijkste bevinding uit de architectuurscan is dat een scherpere scope nodig is zodat helder is voor welk soort gegevens zo'n platform bedoeld is, en daarmee ook voor welk soort gegevens het platform juist niet is bedoeld.

DUO heeft aangegeven graag het gesprek hierover aan te gaan met de Architectuurraad. Het advies over de scope en de vervolgstappen kan daarna worden teruggekoppeld aan de Datakamer

Scopeafbakening – een aanzet

Voor de eerste use case wordt de scope met name bepaald door het in de pitch aangehaalde artikel 23 WRO. Het is de bedoeling dat een onderwijsgegevensplatform uiteindelijk breder kan worden toegepast voor “het zoeken, vinden, begrijpen, vergelijken, kiezen en integreren van onderwijsgegevens om ze te kunnen gebruiken voor onderzoek, ondersteuning en evaluatie doeleinden”. Dit vraagt om een nadere duiding van voor het platform relevante ‘onderwijsgegevens’. Het gaat dan achtereenvolgens om:

1. Aanbod van gegevens die DUO worden verzameld voor de WRO
2. Aanbod van gegevens die DUO inwint (ook bijvoorbeeld WSF, onderwijspersoneel, bekostigingsstatus)
3. Breed aanbod van gegevens voor onderzoeksbehoefte, bijvoorbeeld ten aanzien van “leerresultaten”.

Een limitatieve opsomming van gegevens die al dan niet in scope zijn van zo'n platform is lastig, omdat de aard van onderzoek en evaluatie juist is dat de te beantwoorden vragen niet per se al op voorhand (d.w.z. op het moment dat gegevens beschikbaar komen) bekend zijn. De scope laat zich daarom beter uitdrukken in termen van karakteristieken van de desbetreffende gegevens. Voor gegevens die in scope zijn van een onderwijsgegevensplatform kan worden gedacht aan:

- Geaggregeerde gegevens, en
- Gegevens op individueel niveau, waarbij het irrelevant is wie het individu is (geanonimiseerd, gepseudonimiseerd), en
- Datasets (intern coherent) waarbij hetzelfde individu weliswaar in meerdere datasets kan voorkomen maar dat individu niet tussen verschillende datasets traceerbaar is, en/of

- Gekoppelde datasets, waarbij niettemin de privacy van individuen in de datasets is gewaarborgd, en
- Publieke (register)gegevens, en
- ...?

Voor gegevens die qua onderwerp in scope zijn kan dan worden gedacht aan gegevens die – aan de hand van de hierboven genoemde karakteristieken - informatie geven over:

- Onderwijsaanbieders
- Onderwijsvolgers
- Medewerkers / Onderwijsprofessionals
- Leerresultaten
- ...?

Nota bene, dit is mogelijk ruimer dan de scope die de Datakamer voor ogen heeft, en vereist hoe dan ook goed ingerichte Privacy by design en goede juridische AVG-begeleiding en/of -toetsing. Het is daarbij van essentieel belang dat de inhoudelijke betekenis van de gegevens op het platform eenduidig is gedefinieerd, en dat adequate metagegevens beschikbaar zijn om (onderzoeks)vraag en aanbod te kunnen matchen.

Een vervolgvraag is wat de functionele scope van het platform zou moeten zijn. Is het

- Een plaats waar datasets zijn geregistreerd, gemetadateerd en vindbaar zijn;
- Een plaats waar alle data zelf vindbaar en doorzoekbaar is
- Een omgeving waarbinnen onderzoek kan worden gedaan (zgn. VRE / Virtual Research Environment)

Op basis van die keuze, moeten de benodigde *functies* van het platform verder worden uitgewerkt.

Vraag

De leden van de Architectuurraad wordt gevraagd:

- te reflecteren op de aanzet tot scopeafbakening en hun aanvullingen en ideeën hierover te delen tijdens de bijeenkomst van de Architectuurraad op 14 april 2022:
 - De aard van de gegevens
 - Het onderwerp van de gegevens
 - De functionele scope
- Aansluitend ligt de vraag voor hoe een vervolg te geven aan de wens van de Datakamer om “gezamenlijk in een werkgroep [te] komen tot een gedragen ketenarchitectuur”. DUO heeft aangegeven bij de bijeenkomst op 14 april aanwezig te zijn.