

# Verslag Onderwijsdata webinar 18 juni 2020 data.overheid.nl

*Op donderdag 18 juni 2020 organiseerden we weer een bijeenkomst voor gebruikers van data.overheid.nl in de vorm van een webinar. Tijdens de vorige online gebruikersbijeenkomst kregen we verzoeken om iets meer de diepte in te gaan. Daarom stond deze webinar in het teken van Onderwijsdata en de Onderwijs community! Hoe kan data.overheid.nl deze communities het beste ondersteunen? En wat wil de (her)gebruiker?*

**Alle presentaties zijn hier terug te vinden:**

<https://data.overheid.nl/actueel/bijeenkomsten/presentaties-gehouden-bijeenkomsten>

Het programma van de webinar zag er als volgt uit:

1. Jelle Verburg (data.overheid.nl)  
Introductie data.overheid.nl en data communities. Wat is het?
2. Pearl Dijkstra (KNAW) – Onderzoeksrapport
3. Jelle Verburg (data.overheid.nl) & Erik Smits (Min. OCW) – Onderwijsdata community
4. Marcel Dekker (Smoelt) – Toepassing met onderwijsdata
5. Ronald de Jong (CBS) – Toepassingen in het domein onderwijs met CBS data
6. Jelle Verburg (data.overheid.nl) – dataverzoeken en interactie.  
Welke datasets wil je hebben in het domein onderwijs?

---

## **1. Jelle Verburg (data.overheid.nl): Introductie data.overheid.nl en data communities**

Jelle is product owner van data.overheid.nl en tevens de presentator van deze webinar. Hij werd hierbij ondersteund door Hayo Schreijer en Victorine Nillesen van team data.overheid.nl.

Data.overheid.nl werkt als vindplaats, monitor, makelaar en ondersteuning van open (overheids)data. Om deze rollen te versterken zijn in mei 2020 de data communities opgericht: online en offline netwerken van mensen om het aanbieden, gebruiken en delen van data te ondersteunen in een specifiek domein. De communities geven daarnaast meer inzicht in de impact van open data.

## **2. Pearl Dijkstra (KNAW): Adviesrapport over hergebruik van publieke data**

Pearl Dijkstra is wetenschappelijk directeur van ODISSEI: Open Data Infrastructure for Social Science and Economic Innovations (<https://odissei-data.nl/nl/>). Het adviesrapport over hergebruik van publieke data is opgesteld in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW). Het is belangrijk dat het overheidsbeleid van een stevige wetenschappelijke kennisbasis wordt voorzien. Daarom is toegang en gebruik van publieke data zo belangrijk. Data.overheid.nl is niet altijd bruikbaar voor de wetenschap omdat data eerst aan elkaar gekoppeld moet worden. Door het aan elkaar koppelen van datasets krijg je interessante vraagstukken. Helaas zijn er nog veel obstakels voor wetenschappers bij het vinden van data.

In hoeverre zijn publieke data FAIR? = Findable, Accessible, Interoperable, Reusable  
- Zijn ze vindbaar? Er wordt nog weinig gebruik gemaakt van “persistent identifiers” om data altijd vindbaar te maken.

- Zijn publieke data toegankelijk? Er is een terughoudendheid om data te delen, dit heet *data hogging*. Daarbij zijn er zorgen om privacy, datakwaliteit, kosten/menskracht en claims vanuit de samenleving.
- Zijn publieke data interoperable? Vaak zijn data niet machine-leesbaar door bijvoorbeeld incompatibele formaten.
- Zijn publieke data herbruikbaar?

Cruciale observaties van het rapport: formele barrières om publieke data aan wetenschappers beschikbaar te stellen zijn er niet. Moet data dan altijd gratis zijn? Nee, zeker niet. Voorbeelden van modellen die navolging verdienen: INSPIRE (data milieugegevens), en CBS microgegevens/microdata.

Data.overheid.nl is een klankbordfunctie, deze is nodig bij het FAIR maken van publieke data, ook voor wetenschap. Hoe kunnen we wetenschappers helpen om deze data te krijgen? Er moeten interdepartementale afspraken over procedures gemaakt worden waarmee publieke data actief beschikbaar worden gesteld. Overheidsorganisaties moet worden gevraagd hun data open te stellen.

### **Vraag gesteld aan Pearl: Wat wordt verstaan onder publieke data?**

Publieke data zijn data die de overheid verzamelt bij de uitvoering van publieke taken. Dit zijn statistieken die bijgehouden worden op de ministeries/gemeenten. Publieke data zijn vaak niet open omdat ze niet altijd publiek beschikbaar kunnen worden gesteld. Er kan ook een commercieel belang bij zitten, of privacy, daarom zijn ze niet open.

Het hele rapport van het KNAW is hier terug te lezen:

<https://www.knaw.nl/nl/actueel/publicaties/hergebruik-van-publieke-data>

### **1. Jelle Verburg (data.overheid.nl) & Erik Smits (ministerie van OCW) – Onderwijsdata community**

Waarom hecht het ministerie van Onderwijs, Cultuur en Wetenschap waarde aan data communities, en specifiek aan de Onderwijsdata community?

Erik Smits: de interactie tussen gebruiker, afnemer en aanbieder, iedereen probeert dit te organiseren maar dat gaat vrij moeizaam. Je hoopt dat wat je aanbiedt als onderwijsdata voorziet in de behoefte. Als gebruiker kan je doorvragen naar verdere gegevens. Daarom is het fijn dat we de community opzetten. Maar daar horen vragen bij: hoe groot wordt de community en blijft het beheersbaar? Erik is heel benieuwd naar het gebruik van de Onderwijsdata community. Je ziet vaker startende communities die snel doodbloeden.

Welke organisaties doen er mee? Er is een breed veld en veel kennis binnen onderwijs. Onder andere het ministerie van OCW, Inspectie van het Onderwijs, het CBS, UWV, CINOP, SBB, SLO, DUO, ECBO, Kennisnet, NWO en VNG zijn aangesloten.

Jelle: Via de Onderwijsdata community kunnen gebruikers zoeken naar data in het domein onderwijs die worden gepubliceerd. We vragen iedereen om verbeteringen aan ons door te geven! Wij proberen de community zo goed mogelijk in te stellen, maar we zitten nog in de opstartfase.

We spelen ook in op actuele en/of maatschappelijke vraagstukken, zoals bijvoorbeeld kansengelijkheid of financiering. Zo halen we datasets bij elkaar. Dit gaan we verder vormgeven en finetunen. Daarbij is een discussieforum opgericht voor vragen en topics. Tenslotte hebben we een Gitbook met technische documentatie zodat je alles kan nalezen. Hierop plaatsen we ook informatie over releases en een roadmap met toekomstige ontwikkelingen. Dit is allemaal ter ondersteuning van de opbouw van de community. Neem een kijkje en stel je vragen over onderwijsdata!

#### **4. Marcel Dekker (Smoelt) – Toepassing met onderwijsdata**

Marcel Dekker van Smoelt heeft een mooie toepassing gemaakt met onderwijsdata. Deze toepassing gaat over fietsafstanden naar scholen in Nederland, met name in de Achterhoek. Hierbij is veel contact geweest met AchterhoekVO. Welke data is hierbij gebruikt? O.a. de locaties van scholen, dit is beschikbaar via DUO. Dit vraagt natuurlijk nog wel veel uitpluiswerk. Er was bijvoorbeeld veel kennis van data nodig, dus heeft Marcel veel contact gehad met het bestuur van AchterhoekVO. De kaart is gemaakt met QGIS.

De presentatie van Marcel is hier terug te vinden:

<https://www.linkedin.com/feed/update/urn:li:activity:6679359941838483457/>

#### **5. Ronald de Jong (CBS) – Toepassingen in het domein onderwijs met CBS data**

Ronald de Jong: we zien veel nieuwe interessante vraagstukken. We plaatsen visualisaties op de website van het CBS over actuele maatschappelijke onderwerpen. Daarbij hebben we natuurlijk heel veel data en statistieken.

Statline draait op onze API en onze open data draait er ook op. Bijvoorbeeld onder het thema Onderwijs staat veel downloadbare data.

Interessante toepassingen in het domein onderwijs, onder andere gemaakt met data van het CBS: Open Data Jeugdmonitor en Onderwijs in Cijfers, de laatste is van het ministerie van Onderwijs, Cultuur en Wetenschap. Ook OCW in cijfers is interessant – deze wordt waarschijnlijk samengevoegd met Onderwijs in Cijfers.

Ga vooral naar de Onderwijsdata community als je data zoekt! Hier staan bijvoorbeeld al 124 datasets van het CBS. Daarbij is de community ook de ingang om naar data te vragen. Mis je iets? Geef een gil en dan kunnen wij als organisaties leren wat nodig is en een oplossing bieden.

#### **6. Jelle Verburg (data.overheid.nl) – dataverzoeken en interactie. Welke datasets wil je hebben in het domein onderwijs?**

Welke data wil je hebben? Vertel het ons! Er zijn heel veel dataverzoeken gesteld in het domein onderwijs.

#### **Afsluiting**

De communities zijn een plek om in een specifieke context inzicht in data, toepassingen, gebruik en impact te krijgen. Hier gaan we komende maanden verder mee aan het werk. Contact ons via [data@koop.overheid.nl](mailto:data@koop.overheid.nl)

### **Vragen, antwoorden en informatie uit de chat tijdens de webinar:**

**Vraag:** Kunnen we een link krijgen naar het adviesrapport en Odissei?

**Antwoord:** <https://odissei-data.nl/nl/>

**Opmerking:** Publieke data zijn dus niet per se openbaar (vertrouwelijke gegevens waaronder persoonsgegevens). Die alleen koppelen met specifieke doelen!

**Opmerking:** Meta data geeft zicht op betrouwbaarheid data.

**Vraag:** doen de ziekenhuizen/ wetenschappers ook aan data openstellen (of semi open)?

**Opmerking:** Interoperability is ook een problematisch thema bij veel wetenschappelijke datasets die afkomstig zijn. Uit 'the long tail of science'

**Vraag:** Waarom is onderzoeksdata vanuit de wetenschap zelf niet beschikbaar op data.overheid.nl?

**Opmerking:** De vindbaarheid is iig erg goed bij de overheid, mede door het gebruik van DCAT-AP en gecontroleerde vocabulaires!

#### **Opmerkingen over INSPIRE:**

- INSPIRE = Infrastructure for Spatial Information in Europe
- INSPIRE gaat over environmental data. Dt is op zich wel wat breder dan wat we in Nederland milieugegevens noemen...
- INSPIRE is een grote belangrijke standaard, maar het biedt niet altijd voldoende granulariteit
- Niet voldoende granulariteit, maar sluit ook niet altijd aan bij beleving van eindgebruiker.
- "granulariteit" is iets wat je bij de data aanbieder onder de aandacht kunt brengen, inspire doen we ook voor onszelf, niet alleen voor europa
- Daarnaast is cross country de interpretatie van de semantiek in de modellen van INSPIRE soms anders waardoor harmonisatie maar deels slaagt.
- Maar INSPIRE is wel een grote aanjager voor toegang tot en hergebruik van open geodata via PDOK

**Opmerking:** Onderzoeksdata wordt veelal via DANS gedeeld: <https://dans.knaw.nl/en>

**Vraag:** Waarom niet ook op data.overheid.nl?

**Antwoord:** We zijn met DANS in gesprek. Metadata standaarden lopen ver uiteen.

**Reactie:** met DCAT kom je een heel eind

**Vraag:** Wat doet KNAW om haar data beschikbaar te maken als open data?

**Antwoord:** Financiering loopt langs NWO en ZonMW. De universiteiten onderschrijven het belang van data delen—maar ook daar is nog winst te behalen.

**Opmerking:** 'Open Data' zijn vanuit de definitie rechtenvrij, openbaar en vrij te gebruiken. Dus specifieke gegevens voor wetenschap en onderzoek die niet openbaar zijn kun je niet betitelen als 'Open Data'

**Vraag:** Wat is/doet een data steward?

**Antwoord:** dat is contactpersoon binnen organisatie die alles van open data weet

**Opmerking:** Open Data hebben doorgaans een gemaximaliseerde embargoperiode.

**Vraag:** Is er geen aansluiting bij de SURF Open Onderwijs API community?

**Antwoord:** Via Kennisnet denk ik

**Vraag:** Gaat het hier om data 'over onderwijs', of juist ook om data 'voor onderwijs'? In het laatste geval denk ik aan het delen van onderwijsmateriaal, iets wat bv. soms al gebeurt via FigShare.

**Antwoord:** Beide. Het mooie is dat we thema's steeds meer in context kunnen benaderen.

**Opmerking:** Via 4TU Data Repository <https://researchdata.4tu.nl> wordt data van technische universiteiten gedeeld

**Vraag:** Waarom is deze niet aangesloten op data.overheid.nl?

**Antwoord:** 'aansluiten' betekent dat de catalogus metadata kan harvesten volgens een vast schema. 4TU maakt geen gebruik van DCAT-AP, maar van andere standaarden (DataCite geloof ik)

**Antwoord:** Ik denk dat een tussenliggende DCAT adapter snel gemaakt is

Mee eens, DCAT is prima te mappen op DataCite en vv

**Opmerking:** Voor degenen die met R werken, er is een package (cbsdataR) om data eenvoudig rechtstreeks binnen te halen vanuit CBS data.

**Vraag:** ander punt over zowel 4tu als dans, is dat de data niet vindbaar is in google dataset search, is dit bewust?

<https://datasetsearch.research.google.com/search?query=site%3Aresearchdata.4tu.nl>

**Antwoord:** wordt wel gevonden wanneer er van schema.org gebruik wordt gemaakt voor Google is van belang dat de metadata standaard van schema.org wordt gebruikt. Data.overheid.nl heeft DCAT naar schema.org vertaald.

**Vraag:** hebben meer studenten zich aangemeld voor 2020/2021 ? c.q. is er effect van Covid-19

**Antwoord:** Hier wordt al veel onderzoek naar gedaan door de verschillende instellingen. Vooral de universiteiten (bijvoorbeeld de VU) delen hun cijfers en prognoses

**Opmerking:** er is in het kader van het versnellingsplan ICT in het onderwijs van SURF ook een werkgroep studiedata waarmee verbinding waardevol is; die kunnen beter aangeven wat voor hen relevante data is

**Vraag:** Data over niet bekostigde opleidingen (inschrijvingen, landen van herkomst van internationale studenten, rendementen)

**Opmerking:** Op het eerste gezicht zie ik verschil tussen de geo-plaatjes in de verschillende presentaties over afstanden tot onderwijsinstellingen. Impact zou kunnen zijn dat er discussie ontstaat over de manier waarop data geïnterpreteerd zijn.

**Vraag:** Zijn er data voor handen over het afronden van post Hbo-opleidingen en/of masters door leraren? In relatie tot de kwaliteit van het onderwijs?

Link: <https://versnellingsplan.nl/zones/studiedata/>