

# Naar een database van (startende) leraren en lerarenopleidingen

## Citation for published version (APA):

Belfi, B., Cörvers, F., & Mommers, A. (2017). Naar een database van (startende) leraren en lerarenopleidingen: verkennend onderzoek naar de potentie en de wenselijkheid. (ROA reports; No. 008). Research Centre for Education and the Labour Market.

## Document status and date:

Published: 01/01/2017

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Unspecified

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# Naar een database van (startende) leraren en lerarenopleidingen: verkennend onderzoek naar de potentie en de wenselijkheid

## ROA Rapport

ROA-R-2017/8

Researchcentrum voor Onderwijs en Arbeidsmarkt | ROA  
*Research Centre for Education and the Labour Market | ROA*

# NAAR EEN DATABASE VAN (STARTENDE) LERAREN EN LERARENOPLEIDINGEN:

VERKENNEND ONDERZOEK NAAR DE POTENTIE EN DE  
WENSELIJKHEID

ROA-R-2017/8

Barbara Belfi  
Frank Cörvers  
Ardi Mommers

## Colofon

© Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA). Niets uit deze uitgave mag op enige manier worden verveelvoudigd zonder voorafgaande schriftelijke toestemming van de directeur van het ROA.

### **Researchcentrum voor Onderwijs en Arbeidsmarkt**

School of Business and Economics

Maastricht University

email: [secretary-roa-sbe@maastrichtuniversity.nl](mailto:secretary-roa-sbe@maastrichtuniversity.nl)

website: [www.roa.nl](http://www.roa.nl)

### **Redactie en productie**

Researchcentrum voor onderwijs en arbeidsmarkt (ROA)

### **Vormgeving**

ROA secretariaat, Maastricht

ISBN: 978-90-5321-563-0

september 2017

# INHOUD

<b>Voorwoord</b>	<b>v</b>
<b>Samenvatting</b>	<b>vii</b>
<b>1. Inleiding</b>	<b>1</b>
1.1 Aanleiding en doelstelling	1
1.2 Vraagstelling en onderzoeksopzet	2
1.3 Leeswijzer	3
<b>2. Overzicht van relevante data en onderzoeken voor een database</b>	<b>5</b>
<b>3. CBS: beheer en microdata</b>	<b>11</b>
3.1 Inleiding	11
3.2 Visie op database	11
3.3 Aanwezige data	13
3.4 Databenutting: ontsluiting en koppelmogelijkheden	14
<b>4. DUO: registerdata onderwijs</b>	<b>19</b>
4.1 Inleiding	19
4.2 Visie op database	19
4.3 Aanwezige data	20
4.4 Databenutting: ontsluiting en koppelmogelijkheden	21
<b>5. Inspectie van het Onderwijs: kwaliteitsindicatoren</b>	<b>25</b>
5.1 Inleiding	25
5.2 Visie op database	25
5.3 Aanwezige data	26
5.4 Benutting: ontsluiting en koppelmogelijkheden	27
<b>6. Lerarenopleidingen: studentgegevens</b>	<b>29</b>
6.1 Inleiding	29
6.2 Visie op database	30
6.3 Aanwezige data	32

6.4	Databenutting: ontsluiting en koppelmogelijkheden	33
6.5	Case study 1: Hogeschool van Arnhem en Nijmegen	35
6.6	Case study 2: De Nieuwste Pabo / Zuyd Hogeschool	37
<b>7.</b>	<b>Lerarenregister: professionaliseringsgegevens</b>	<b>43</b>
7.1	Inleiding	43
7.2	Visie op database	44
7.3	Aanwezige data	44
<b>8.</b>	<b>Verdere onderzoeksbestanden</b>	<b>47</b>
8.1	Inleiding	47
8.2	Begeleiding Startende Leraren	47
8.3	Loopbaanmonitor	48
8.4	HBO-Monitor	50
<b>9.</b>	<b>OCW: kijk op database</b>	<b>53</b>
9.1	Inleiding	53
9.2	Visie op database	53
9.3	Uitdagingen	54
<b>10.</b>	<b>CPB: ervaring met gekoppelde leraargegevens</b>	<b>57</b>
10.1	Inleiding	57
10.2	'Paden naar leraarschap': de mogelijkheden	57
10.3	Kanttekeningen bij datakoppelingen	58
<b>11.</b>	<b>Onderwijsonderzoekers: mogelijkheden voor onderzoek</b>	<b>63</b>
11.1	Inleiding	63
11.2	Visie op database	63
11.3	Kritische noten bij de organisatie van de database	64
<b>12.</b>	<b>NCO: datakoppelingen en het veld</b>	<b>67</b>
12.1	Inleiding	67
12.2	Basisvoorwaarden: draagvlak, focus en realisme	68
<b>13.</b>	<b>Conclusies en discussie</b>	<b>71</b>
13.1	Inleiding	71
13.2	Wenselijkheid en meerwaarde	71
13.3	Onderliggende databestanden: haalbaarheid en ontwikkeling	74
13.4	Potentie en haalbaarheid verschillend per bestand	77
13.5	Beperkingen in het huidige onderzoek	78
13.6	Mogelijke vervolgstappen	79
	<b>Bijlage I: Overzicht van geïnterviewden</b>	<b>83</b>

# VOORWOORD

Dit onderzoek is door het Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA) uitgevoerd in opdracht van het Nationaal Regieorgaan Onderwijsonderzoek (NRO). Bij de totstandkoming van dit onderzoek zijn wij dank verschuldigd aan al degenen die hebben meegewerkt aan de interviews. Een overzicht van deze personen is opgenomen in de Bijlage.

Voor de begeleiding bij het onderzoek willen we graag Jelle Kaldewaij (NRO) en Gitta Snijders (NRO) bedanken. Bovendien hebben we het commentaar van Thijs Noordzij (OCW) en Marc van der Steeg (OCW) op een eerdere versie van dit rapport zeer gewaardeerd.

Maastricht, 19 september 2017

Prof. dr. Frank Cörvers,  
projectleider





# SAMENVATTING

## Aanleiding en onderzoekopzet

In de Lerarenagenda 2013-2020 'De leraar maakt het verschil' is een aantal belangrijke uitdagingen en verbeterpunten voor lerarenopleidingen opgenomen. De lerarenopleidingen hebben in reactie hierop reeds veel inspanningen verricht om hun opleidingen te verbeteren. Dit blijkt onder meer uit de invoering van de kennisbases en -toetsen, meer aandacht voor de kwaliteit van lerarenopleiders, en het opzetten van opleidingscholen. De vraag blijft echter in hoeverre het onderwijs dat startende leraren geven daadwerkelijk aan het verbeteren is. Hoewel verschillende studies zich reeds hebben gericht op de invloed van kenmerken van (beginnende) leraren op het onderwijs, kennen deze een aantal beperkingen. Zo zijn studies naar lerarenkenmerken zoals opleidingsachtergrond, ervaring en geslacht vaak gefragmenteerd of eenmalig. Daarnaast zijn ze moeilijk te relateren aan kenmerken van individuele lerarenopleidingen en is er vaak geen verband te leggen met specifieke maatregelen binnen de lerarenopleidingen.

Naar aanleiding van deze beperkingen, is de gedachte ontstaan om op zoek te gaan naar mogelijkheden om meer systematisch gegevens over de kenmerken van leraren en de door hen doorlopen opleidingen te verzamelen, wat moet leiden tot een algemene 'database' ten behoeve van langetermijnonderzoek. Het doel van de database is om meer inzicht te krijgen in de kenmerken van beginnende leraren en lerarenopleidingen die er werkelijk toe doen bij het lesgeven. De nadruk ligt bij een dergelijke database (in eerste instantie) op studenten aan de lerarenopleidingen, de lerarenopleidingen zelf en de startende leraren. De nadruk op studenten en beginnende leraren is te verklaren doordat zij het lerarenkorps van de nabije toekomst vormen en eventuele ontwikkelingen in de kwaliteit van leraren(opleidingen) binnen deze groepen het snelst zichtbaar zijn. Bovendien is de kans dat leraren hun baan in het onderwijs verlaten het grootst in de eerste jaren na hun afstuderen.

Aan de hand van een database bestaande uit gegevens van (startende) leraren en lerarenopleidingen kunnen op een meer systematische en structurele wijze dan nu het geval is gegevens worden verzameld en aan elkaar worden gerelateerd. Hierdoor is het mogelijk om meer verdiepend en verklarend onderzoek uit te voeren naar de kwaliteit(sbeïnvloedende factoren) van leraren, en is het bovendien beter mogelijk om

de rol van de lerarenopleidingen binnen dit geheel in kaart te brengen. Ook biedt een database de mogelijkheid om meer gedetailleerde arbeidsmarktinformatie te genereren over bijvoorbeeld de stromen van leraren binnen het onderwijs en daarbuiten.

Het Nationaal Regieorgaan Onderwijsonderzoek (NRO) heeft het Researchcentrum voor Onderwijs en de Arbeidsmarkt (ROA) verzocht een verkennend onderzoek uit te voeren naar de potentie en wenselijkheid van een dergelijke database. Meer specifiek dient te worden nagegaan a) welke kennisbehoefte er bestaat bij de mogelijke afnemers uit de wetenschap, het beleid en het onderwijsveld en in hoeverre een database hieraan kan bijdragen; b) welke onderzoeksbestanden over leraren (c.q. lerarenopleidingen) relevant zijn om op te nemen in een database; en c) in hoeverre bestaande gegevens – nu en in de toekomst – ontsloten en gekoppeld kunnen worden in een database en welke hindernissen daarbij overwonnen moeten worden. Om hier inzicht in te krijgen is er door de onderzoekers van het ROA een groot aantal bestaande databestanden geïnventariseerd en zijn er veel gesprekken gevoerd met diverse betrokkenen uit het veld. Het gaat daarbij om vertegenwoordigers of medewerkers van lerarenopleidingen, het ministerie van OCW, CBS, DUO, de Inspectie van het Onderwijs, de Onderwijscoöperatie en de wetenschap. Daarnaast is er aan de hand van bestaande documentatie een overzicht gemaakt van de beschikbare onderzoeksbestanden die voor een database potentieel relevant zijn. Het voorliggende rapport brengt verslag uit van deze verkenning.

### Beantwoording van de onderzoeksvragen

In deze samenvatting worden de centrale onderzoeksvragen uit de voorliggende verkenning puntsgewijs beantwoord. Daarna wordt kort gereflecteerd op de bevindingen en ingegaan op eventuele vervolgstappen. Voor een uitgebreidere beschrijving van de onderzoeksbevindingen en een meer beschouwende terugblik op het onderzoek, verwijzen we naar de conclusies en de discussie in hoofdstuk 13.

1. Welke bestaande data en onderzoeken zijn relevant voor een database over (startende) leraren en lerarenopleidingen, onder andere ten aanzien van de kwaliteit van leraren, de arbeidsmarktpositie, de curricula van de lerarenopleidingen en de kenmerken van de ontvangende scholen?

Een belangrijke aanleiding voor deze verkenning betreft de grote hoeveelheid gefragmenteerde en niet centraal toegankelijke databestanden over leraren en lerarenopleidingen. Het geven van een overzicht van alle databestanden uit eerdere onderzoeken (vaak ad hoc) is geen sinecure, en is in het kader van voorliggende verkenning ook niet zinvol. We bespreken in deze samenvatting daarom een aantal van de meest voor de hand liggende databestanden, die structureel van aard zijn, een groot bereik hebben en bij voorkeur records op individueel niveau bevatten. Een uitgebreid overzicht van potentieel interessante bestanden staat in hoofdstuk 2.

De belangrijkste databestanden die een groot bereik hebben, structureel verzameld worden en op individueel niveau beschikbaar zijn, worden beheerd door het CBS en DUO. Het CBS beschikt over een grote hoeveelheid data uit het zogenaamde 'Stelsel van Sociaal-statistische Bestanden' (SSB). Het gaat hier om (onderling koppelbare) bestanden met onder andere demografische, sociaal-culturele en sociaal-economische gegevens. Het gaat bijvoorbeeld om inschrijvings- en diplomagegevens van studenten (DUO-gegevens), baankenmerken (zoals loongegevens uit de Polisadministratie), achtergrondkenmerken (zoals etniciteit, geslacht) enzovoorts. De bestanden zijn grotendeels administratief van aard ('registerbestanden') of zijn gebaseerd op grootschalige enquêtes. De bestanden in SSB zijn zeer rijk en informatief, maar maken het niet mogelijk om specifieke beroepen en functies (zoals 'leraar') te onderscheiden, enkel arbeidsmarktsectoren.

Dit onderscheid is wél mogelijk binnen de onderwijspersoneelsgegevens van DUO. DUO beschikt over loon-, verlof- en baangegevens van leraren (type functie, aanstelling, inschaling). Met name voor het voortgezet onderwijs zijn de databestanden zeer rijk, omdat er vanuit de 'Integrale Personeelstelling Onderwijs' (IPTO) uitgebreide lesgegevens beschikbaar zijn. Hierdoor is bekend welke bevoegdheden leraren hebben en in welke vakken en hoeveel uren zij lesgeven. Ook beschikt DUO over veel studentgegevens, zoals inschrijvings- en diplomeringsgegevens. Deze gegevens worden voor een belangrijk deel echter al als microdata via het CBS beschikbaar gesteld. Tevens beschikt DUO over gegevens op schoolniveau, zoals jaarrekeningcijfers, gegevens over de leerlingpopulatie, denominatie, et cetera.

Een andere voor de hand liggende bron voor een lerarendatabase is de Inspectie van het Onderwijs. De Inspectie verzamelt gegevens om toezicht te houden op de onderwijskwaliteit in Nederland. Op het niveau van individuele leraren beschikt de Inspectie over (niet integrale) gegevens afkomstig uit lesobservaties en vragenlijstonderzoeken. Deze gegevens zijn echter niet tot individuen herleidbaar en enkel te benutten op het schoolniveau (BRIN-vestiging), waardoor ze relevant zijn voor gebruik als contextuele kenmerken.

De Nederlandse lerarenopleidingen zelf beschikken over veel relevante gegevens over de intrede, voortgang en uitstroom van studenten. Wat betreft intredegegevens gaat het bijvoorbeeld om jaarlijkse entreetoetsen voor Nederlandse taal en rekenen/wiskunde. Daarnaast beschikken de opleidingen over voortgangsgegevens tijdens de studie. Het betreft onder meer kennistoetsen en voortgangsgegevens over de ontwikkeling van competenties zoals algemene didactische vaardigheden, pedagogische vaardigheden en coaching. Uitstroomgegevens waarover opleidingen beschikken zijn bijvoorbeeld het wel of niet behalen van een diploma binnen de nominale studieduur. Daarnaast beschikken de opleidingen over alumnigegevens die worden verzameld via de algemene alumni-onderzoeken, maar ook via persoonlijke contacten met afgestudeerden.

Ook is in dit onderzoek het Lerarenregister verkend als mogelijke databron voor een database. In het (inmiddels voor 2018 verplicht gestelde) Lerarenregister houden bevoegde leraren hun professionaliseringsactiviteiten bij. Het register bevat zodoende in de toekomst veel interessante data over de professionalisering van leraren, maar de benutting als databestand is in de komende jaren technisch nog niet mogelijk door de datastructuur van het in ontwikkeling zijnde register.

Naast voornoemde bestandstypen zijn er nog vele andere databestanden die potentieel relevant zijn voor een database. Een drietal typen gegevens die verzameld zijn in het kader van onderzoek lichten we hier tot nog kort toe. Het betreffen onderzoeksbestanden die zich vooral richten op de eerste stappen van afgestudeerde leraren op de arbeidsmarkt. Allereerst is het project Begeleiding Startende Leraren (BSL) het vermelden waard. BSL betreft een landelijk onderzoek naar de effecten van inductiemanagement op de ontwikkeling van startende leraren in het voortgezet onderwijs. Deelnemende leraren worden gemonitord via lesobservaties (ICALT) en vragenlijsten. Ten tweede is de Loopbaanmonitor Onderwijs van belang. De Loopbaanmonitor heeft als doel het verkrijgen van inzicht in de arbeidsmarktsituatie van pas afgestudeerde leraren, de begeleiding die zij op de ontvangende school hebben ontvangen, en hun inschatting van het eigen kunnen. Tot slot noemen we de HBO-Monitor als interessant databestand. De HBO-Monitor geeft inzicht in de arbeidsmarktpositie van afgestudeerde hbo'ers anderhalf jaar na diplomering en toont in hoeverre opleidingen aansluiten op de eisen van de beroepspraktijk. Tevens geeft het inzicht in het oordeel van afgestudeerden over de opleiding.

## 2. In hoeverre zijn de beheerders/eigenaren van data bereid om hun onderzoeksresultaten beschikbaar te stellen voor een database?

De microdata van het CBS worden al beschikbaar gesteld voor onderzoeksdoeleinden aan geautoriseerde gebruikers. Toegang tot de CBS-microdata kan enkel plaatsvinden op het CBS zelf, of door geautoriseerde organisaties die beschikken over 'remote access' faciliteiten. De onderwijspersoneelsgegevens waarover DUO beschikt, worden op dit moment nog niet ontsloten via het CBS. Er zijn echter initiatieven ophanden om deze gegevens (deels) beschikbaar te gaan stellen als microdata van het CBS. Het is nog niet duidelijk om welke personeelsgegevens het dan precies zal gaan en of dit op termijn ook verder uitgebreid gaat worden. In de recente CPB-studie 'Paden naar leraarschap' is overigens al sprake van koppeling van CBS-gegevens aan onderwijspersoneelsgegevens van DUO. Ontsluiting van DUO-gegevens is dus in principe reeds mogelijk voor onderzoeksdoeleinden.

De Onderwijsinspectie staat open voor het ontsluiten van gegevens ten bate van onderzoek. Wel dient daarvoor aan bepaalde voorwaarden voldaan worden. Het gaat dan om afspraken over anonimisering en onthulling van scholen en over aan wie toegang tot de data wordt verleend.

De eventuele ontsluiting van studentgegevens vanuit de lerarenopleidingen voor een database zal een flinke uitdaging vormen. Zo bestaan er bij veel geïnterviewden, met name binnen de opleidingskoepels in het hbo, grote bedenkingen over de meerwaarde van een database en dus over de zinvolheid van het ontsluiten van studentgegevens. Ook bestaat er zorg over de privacygevoeligheid van de gegevens en de kwetsbaarheid van de opleidingen. Als er al sprake is van bereidheid tot ontsluiting van gegevens, dan zal dit in ieder geval onder zeer strikte voorwaarden moeten gebeuren. Mogelijke ontsluiting van gegevens via het CBS als een 'vertrouwde derde partij' wordt daarbij overigens niet per definitie als borging van de privacy gezien. De levering van studentgegevens zonder een absolute waarborg voor de privacybescherming is dan ook uitgesloten. Veel geïnterviewden geven aan dat zij zonder expliciete toestemming van de studenten geen gegevens aan zouden willen en/of mogen leveren. Vanuit de ervaringen uit het Nationaal Cohortonderzoek Onderwijs (NCO) lijkt het niet 'mogen' leveren overigens niet direct een probleem (hierover later meer). Een mogelijkheid voor ontsluiting van opleidingsgegevens aan een database zou een 'opt in / opt out'-principe kunnen zijn. Hierbij moeten studenten expliciet aangeven of ze hun gegevens beschikbaar willen stellen en kunnen ze op ieder moment aangeven dat zij hun gegevens verwijderd willen zien.

De gegevens uit het Lerarenregister zullen in de komende jaren zeker niet benut kunnen worden voor onderzoek, aangezien dit technisch niet haalbaar wordt geacht. De Onderwijscoöperatie staat echter niet onwelwillend tegenover een toekomstige verkenning van ontsluitingsmogelijkheden. Mochten data uit het Lerarenregister op den duur ontsloten worden voor onderzoek, dan zal in ieder geval geëist worden dat hier op een zorgvuldige, van tevoren gereguleerde manier mee omgegaan wordt waarmee de organisaties van de Onderwijscoöperatie en de achterban akkoord gaan. De Onderwijscoöperatie geeft aan dat de leraar zelf altijd eigenaar dient te zijn en te blijven van zijn eigen gegevens, en toestemming zal moeten geven voor het benutten van de gegevens.

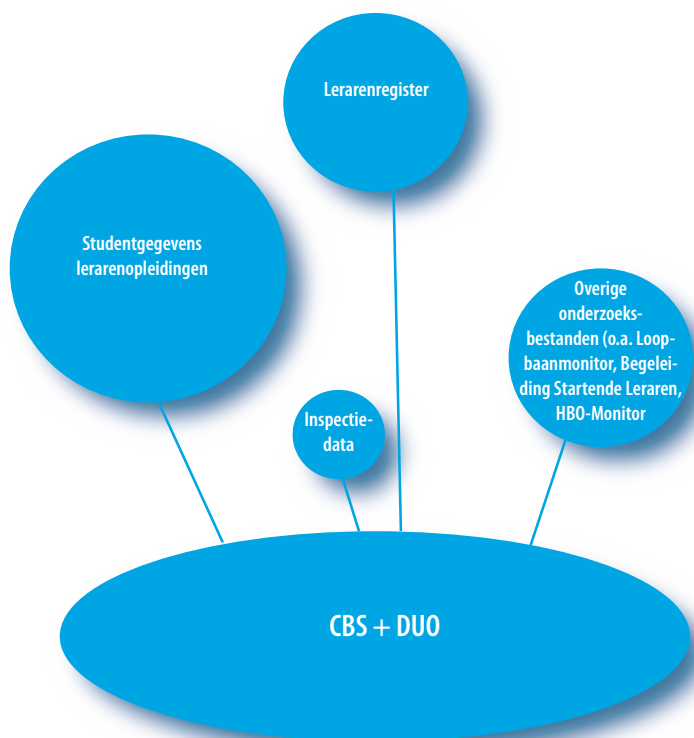
De bereidheid tot ontsluiting van andere bestanden die relevant kunnen zijn voor een database van (startende) leraren en lerarenopleidingen, zoals de onderzoeksdata in de Loopbaanmonitor, de HBO-Monitor of de Nationale Studenten Enquête, zal in toekomstige vervolgstappen nader verkend moeten worden.

### 3. Welke van deze data zijn het meest geschikt om in de database op te nemen?

In onderstaande figuur is de relevantie en haalbaarheid van de in dit rapport verkende bestandstypen grafisch uitgedrukt. De omvang van de cirkel geeft hierbij de ingeschatte relevantie van de bestandstypen weer: een grote cirkel correspondeert met een hoog ingeschatte relevantie. Gezien de hoge mate van integraliteit en de inhoudelijke relevantie van de CBS- en DUO-bestanden, de goede technische mogelijkheden om deze bestanden onderling te koppelen, en de goede ontsluitingsmogelijkheden van deze bestanden, staan deze bestanden in deze figuur aan de 'basis' van een potentiële data-

base. We veronderstellen daarnaast een grote inhoudelijke relevantie van de studentgegevens van lerarenopleidingen. De relevantie van Inspectie-data beschouwen we op dit moment als minder groot, aangezien het hier in principe niet om gegevens op individueel niveau gaat, maar om schoolgeaggreerde gegevens. De cirkel met Inspectie-data is derhalve relatief klein.

FIGUUR 1 Ingeschatte relevantie en haalbaarheid van bestandstypen voor database



De lengte van de lijnen vanaf de 'basis' (CBS- en DUO-data) richting de andere bronnen indiceert de ingeschatte haalbaarheid van ontsluiting van de betreffende bron. Een lange lijn correspondeert met een naar verwachting lage praktische haalbaarheid en een grote benodigde inspanning om de bron te kunnen benutten. CBS- en DUO-bestanden lijken naast inhoudelijk relevant ook relatief eenvoudig te benutten en vormen derhalve de 'basis'. Inspectiedata lijken eveneens praktisch haalbaar om te benutten voor een database, mits hierover goede afspraken worden gemaakt. Het benutten van studentgegevens van lerarenopleidingen of gegevens vanuit andere onderzoeksbestanden (zoals Loopbaanmonitor, Begeleiding Startende Leraren) zal vermoedelijk meer inspanning

vergen. De lijn naar het Lerarenregister is het langst: volgens de Onderwijscoöperatie zullen deze gegevens gezien de structuur van het register in de komende jaren niet bruikbaar zijn om onderzoek op te verrichten.

#### 4. Wat zijn mogelijke juridische en technische hindernissen rondom een database?

Op basis van de gesprekken concluderen we dat er in beginsel geen onoverkomelijke juridische en technische beperkingen zijn voor een database. Het inrichten van een database binnen het CBS lijkt op dit moment het meest voor de hand liggend. Het CBS verzamelt, bewerkt en publiceert statistische gegevens over de Nederlandse samenleving (personen, bedrijven en instellingen) en mag deze ook koppelen aan andere persoonsgegevens in het kader van statistiek. Dit is wettelijk verankerd in de Wet op het Centraal bureau voor de statistiek.

Het CBS beschikt reeds over een grote hoeveelheid demografische, sociaal-culturele en sociaal-economische gegevens op individueel niveau ('microdata'), afkomstig vanuit administraties en surveys, en stelt deze voor een belangrijk deel beschikbaar voor onderzoek. Deze gegevens zijn onderling koppelbaar waardoor zeer rijke databases ontwikkeld kunnen worden. Echter, in de administratieve CBS-bestanden is het beroep van leraar niet als dusdanig herkenbaar. Om een gerichte database te ontwikkelen zijn derhalve aanvullende, beroeps- en functiespecifieke gegevens van DUO nodig. Het CBS biedt de mogelijkheid om externe bestanden (mits de personen in deze bestanden voldoende identificeerbaar zijn om te koppelen) te matchen aan de eigen microdata. Deze koppeling is het eenvoudigst en minst foutgevoelig wanneer het Burgerservicenummer (BSN) benut wordt, aangezien het BSN (vrijwel helemaal) integraal beschikbaar is en uniek voor individuen. Het BSN mag echter alleen gebruikt worden door overheidsorganisaties om hun taak uit te voeren en door niet-overheidsorganisaties indien dit wettelijk bepaald is. In veel externe bestanden, zoals extern vragenlijstonderzoek, zal het BSN dus niet bekend zijn.

Een alternatieve manier om externe bestanden te koppelen aan CBS-bestanden is op basis van persoonsgegevens. Het gaat dan in principe om de combinatie van geboortedatum, geslacht en het volledige woonadres. Overigens is het ook mogelijk om nieuwe 'externe' data te laten verzamelen door het CBS middels aanvullende vragenlijsten. Doordat er vanuit het CBS aangeschreven wordt, kunnen de resultaten eenvoudig gekoppeld worden aan andere CBS-bestanden. Wel dient het onderzoek aan te sluiten bij de doelen en taken van het CBS en zijn er vrij strenge voorwaarden aan verbonden. Zo heeft het CBS als volwaardige partner in het proces onder meer invloed op de inhoud van de vragenlijst en de procedure van steekproeftrekking, dient het CBS met de onderzoeksbevindingen naar buiten te mogen treden, en dient de samenwerkingspartner een goede reputatie te hebben.

Naast CBS- en DUO-data, zijn studentgegevens vanuit lerarenopleidingen inhoudelijk potentieel zeer relevant voor een database. Een belangrijk hindernis bij de benut-

ting van deze gegevens is echter de welwillendheid van opleidingen/instellingen om dergelijke gegevens te ontsluiten. De lerarenopleidingen, met name de opleidingen in het hbo, staan sceptisch tegenover de benutting van studentgegevens voor een database. Juridisch gezien lijken er echter geen onoverkomelijkheden te bestaan. Vanuit de ervaringen uit het Nationaal Cohortonderzoek Onderwijs (NCO) blijkt het benutten van voortgangsgegevens van studenten voor onderzoek mogelijk. In het NCO worden leerlinggegevens in de beveiligde omgeving van het CBS (die daar als 'entrusted third party' optreedt) opgeslagen en bewerkt. Voortgangsgegevens over scholieren, afkomstig uit het leerlingvolgsysteem, kunnen benut worden als schoolbesturen daarvoor toestemming geven. Wel wordt aanbevolen om de (ouders van) scholieren hierover te informeren. Gezien deze ervaringen lijken er dus geen grote juridische struikelblokken voor het benutten van studentgegevens bij lerarenopleidingen te zijn, mits de instellingen daarvoor toestemming geven. Wat wel een mogelijke hindernis kan zijn, is de export van gegevens uit een informatie-/studentvolgsysteem naar een onderzoeksbestand. Of dit al dan niet mogelijk is, is afhankelijk van het systeem dat door een instelling gebruikt wordt. Deze exportemogelijkheid lijkt dus op basis van de gehouden interviews bij lerarenopleidingen te variëren van 'een druk op een knop' tot 'een arbeidsintensieve, complexe klus'. De mogelijkheden om studentgegevens te koppelen aan andere data lijken wel goed, aangezien instellingen beschikken over het onderwijsnummer (BSN) van de studenten en NAW-gegevens.

Het Lerarenregister is inhoudelijk erg interessant voor een Lerarendatabase, maar kan in ieder geval in de komende jaren niet benut worden voor onderzoek. De datastructuur van het register leent zich hier nog niet voor en is hier in principe ook niet voor bedoeld. Bovendien zou eventuele ontsluiting alleen onder strenge voorwaarden mogelijk zijn, hetgeen in de toekomst nader verkend zou moeten worden. Zo zal onder meer vereist worden dat leraren toestemming geven voor de ontsluiting van hun gegevens, en dienen ze volgens de Onderwijscoöperatie altijd eigenaar te blijven van deze gegevens. Wel lijken de mogelijkheden om de data in de toekomst te koppelen met andere bestanden goed, aangezien leraren zich via DigID zullen inschrijven. Dit maakt koppeling via BSN mogelijk.

Het benutten van data afkomstig van de Inspectie van het Onderwijs lijken dan weer veel haalbaarder, de Inspectie staat namelijk in principe open voor voorwaardelijke ontsluiting. Deze voorwaarden zullen in ieder geval betrekking hebben op anonimisering en het voorkomen van onthullingsrisico's. Het gaat dan om schoolgeaggregeerde gegevens die op basis van de unieke schoolcode 'BRIN' eenvoudig extern koppelbaar zijn.

Andere typen bestanden zoals vragenlijstdata uit bijvoorbeeld de Loopbaanmonitor, HBO-Monitor of Nationale Studenten Enquête kunnen zeker relevant zijn, maar zullen niet zonder meer te benutten zijn voor een database. Hiervoor zal in principe 'informed consent' noodzakelijk zijn: deelnemers dienen erover geïnformeerd te worden dat de gegevens benut kunnen worden voor ander onderzoek en dienen hiermee akkoord te



gaan. Op voorhand is de benutting van dergelijke bestanden zeker niet uit te sluiten, maar de specifieke mogelijkheden en meerwaarde van dergelijke bestanden dient nader uitgezocht te worden.

#### 5. Is er draagvlak onder de vertegenwoordigers van de lerarenopleidingen voor een database?

De reacties op een database vanuit de lerarenopleidingen zijn divers, maar voornamelijk sceptisch. Met name de koepelverenigingen van lerarenopleidingen in het hbo zijn kritisch over een database. Zo geven de gesproken representanten van het LOBO (Landelijk Overleg Lerarenopleidingen Basisonderwijs) en het ADEF (Algemeen Directeurenoverleg Educatieve Faculteiten) aan geen duidelijke meerwaarde te zien in een database. De kwaliteit van de opleidingen wordt reeds geborgd door het accreditatiestelsel, en ook in de eigen kennisbehoeften wordt voor een belangrijk deel reeds voorzien. Zo wordt de voortgang van studenten al doorlopend gemonitord en is er al veel kennis uit alumnistudies en contacten met het veld. De aanwezige kennis wordt daarbij al volop gedeeld door opleidingen onderling. Daar waar opleidingen wél aanvullende kennisvragen hebben, lijkt er vooral interesse te bestaan voor contextspecifiek (bijvoorbeeld lokaal/regionaal) en themaspecifiek onderzoek: een generieke database wordt derhalve als weinig nuttig beschouwd. Onderzoek naar leraarkwaliteit via een database wordt bovendien als uitermate complex of zelfs vrijwel onmogelijk beschouwd. Met name bestaat er de zorg dat er al snel naar de lerarenopleidingen zal worden gewezen als schuldige voor de eventuele mindere 'kwaliteit' van bepaalde leraren.

Representanten van de universitaire lerarenopleidingen (ICL) zijn positiever over het idee van een database. Een dergelijk onderzoeksbestand zou volgens hen bijvoorbeeld meer inzicht kunnen bieden in de mogelijke effecten van verschillen in (het curriculum van) lerarenopleidingen op de kwaliteit van leraren tijdens hun loopbaan. Ook wordt aangegeven dat een database meer inzicht zou kunnen bieden in studenten- en arbeidsmarktstromen. De geïnterviewden binnen de ICL staan dan ook niet onwelwillend tegenover een database en eventuele toekomstige ontsluiting van studentgegevens. Wel zullen in dat geval goede afspraken moeten worden gemaakt over de wijze waarop dat zou gebeuren (bijvoorbeeld door actief toestemming te vragen aan studenten) en over privacybescherming.

#### 6. Wat zijn de mogelijke opbrengsten, voordelen en nadelen van een database volgens beleidsmakers, wetenschappers en mensen uit het onderwijsveld?

De reacties op een database vanuit het veld zijn zeer divers. Het spectrum loopt van 'onnodig, niet zinvol en niet aansluitend op de kennisbehoefte' tot 'zeer interessant, bruikbaar en zinvol'. De lerarenopleidingen, in het bijzonder de lerarenopleidingen in het hbo, zitten zoals gezegd vaak aan de sceptische zijde van het spectrum. De Onderwijscoöperatie ziet een database in principe als potentieel nuttig instrument, bijvoorbeeld om onderzoek te verrichten naar de effecten van professionaliserings-

activiteiten op het onderwijs. De doelmatigheid van het scholingsaanbod is dan ook een belangrijk thema voor de Onderwijscoöperatie, en een database zou mogelijk bij kunnen dragen aan het aanbieden van een selectie van 'bewezen effectieve' professionaliseringsactiviteiten. Over onderzoek naar de kwaliteit van leraren is men wel sceptisch, omdat kwaliteit lastig te definiëren is en het complex is om aan te tonen waar kwaliteit door beïnvloed wordt. De Onderwijscoöperatie staat echter niet per definitie onwelwillend tegenover een database. Wel is het voor de Onderwijscoöperatie essentieel dat er goede, heldere onderzoeksvragen ten grondslag liggen aan de eventuele ontwikkeling van een database. Een database enkel 'omdat het kan' wordt als onwenselijk en weinig zinvol geacht.

De eerste reacties over een database vanuit de kant van landelijk onderwijsbeleid, zoals OCW en de Inspectie van het Onderwijs, zijn positief. Ook vanuit onderzoekskant (zoals het CPB en onderwijsonderzoekers aan universiteiten) zijn er vooral positieve geluiden. Er worden weliswaar de nodige kanttekeningen gemaakt, bijvoorbeeld over het gebruik van registerbestanden vanwege de vertragingfactor (dergelijke bestanden zijn niet actueel) en over de soms moeizame ervaringen met het CBS, maar in algemene zin wordt een database als nuttig gezien. Een database kan zowel praktische als wetenschappelijke doelen dienen. Een praktisch doel is bijvoorbeeld het verminderen van de enquêtedruk in het veld. Zo kan vragenlijstonderzoek onder leraren bondiger worden doordat veel achtergrondvragen overbodig worden (zoals persoonskenmerken of schoolkenmerken). Tevens kan, indien aangeschreven zou worden vanuit een database, veel meer focus aangebracht worden: zo kunnen specifieke doelgroepen geselecteerd worden waardoor de 'bijvangst' geminimaliseerd kan worden. Verder biedt een database met gekoppelde registerbestanden de mogelijkheid om grootschalig en verdiepend onderzoek te verrichten zonder het veld te belasten. De CPB-studie 'Paden naar leraarschap' is hier een voorbeeld van.

Op meer onderzoeksinhoudelijk gebied kan een database nieuwe mogelijkheden bieden voor verdiepend onderzoek, waarbij er voor vele factoren 'gecorrigeerd' kan worden (achtergrondkenmerken, schoolkenmerken enzovoorts). Er valt dan bijvoorbeeld te denken aan onderzoek naar de invloed van lerarenopleidingen op de kwaliteit of arbeidsmarktbevingen van leraren, de invloed van leraren op leerlingprestaties, enzovoorts. Ook biedt het mogelijkheden om de effectiviteit van beleidsmaatregelen beter in kaart te brengen, bijvoorbeeld onderzoek naar de lerarenbeurs, subsidies voor zij-instroom, opleiden in de school, aanscherping toelatingseisen pabo's, subsidieregeling begeleiding startende leraren, tegemoetkoming studiekosten lerarenopleidingen tekortvakken. Tevens kan een database zorgen voor minder 'versnippering' van lerarengegevens. Op dit moment worden gegevens over leraren (en opleidingen) door veel verschillende instanties en onderzoeksinstellingen verzameld en worden deze data vaak niet centraal gedocumenteerd. Hierdoor is het bijvoorbeeld niet altijd duidelijk of en waar dergelijke data gevonden kunnen worden. Een centrale verzamelplek kan de toegankelijkheid van (leraar)onderzoek sterk vergroten. Een ander voordeel van een database kan zijn dat er gestandaardiseerde afbakening en definities voor onderzoek

naar leraren en (lerarenopleidingen) opgesteld kunnen worden. In het huidige onderzoek, onder meer op basis van registerdata, worden selecties en definities vaak 'ad hoc' bepaald binnen het betreffende onderzoek. De vergelijkbaarheid van (leraar)onderzoek kan vergroot worden als een database wordt opgebouwd rondom bepaalde standaard-definities. Welke partijen betrokken dienen te worden bij de afstemming van deze selecties en definities zal in de toekomst nader verkend moeten worden. Overigens wordt door diverse geïnterviewden benadrukt dat er altijd wel ruimte moet blijven bestaan om in specifieke gevallen beredeneerd af te kunnen wijken van standaardafbakening en -definities.

### *Kanttekeningen en vervolgstappen*

Het centrale doel van het huidige onderzoek was om de mogelijkheden te verkennen om gegevens over de kenmerken van leraren en de door hen doorlopen opleidingen op een meer systematische wijze te verzamelen dan nu het geval is. De mate waarin gegevens/databestanden relevant zouden kunnen zijn voor een dergelijk database is onder meer beoordeeld op het regelmatig beschikbaar komen van updates, de gerichtheid op de onderwijssectoren, en de grootschaligheid en reputatie van de databestanden. Bij de selectie van relevante databestanden hebben we vrijwel uitsluitend gekeken naar individuele data over leraren en studenten, en veel minder naar de kenmerken van lerarenopleidingen en van ontvangende scholen. Over systematische gegevensverzameling met betrekking tot deze 'contextuele' factoren is niet veel bekend en deze dient dan ook nader onderzocht te worden.

De interviews zijn gehouden met vertegenwoordigers van de leveranciers en gebruikers van de databestanden die wij als meest relevant hebben beoordeeld. De standpunten die zijn geformuleerd kunnen persoonlijke inkleuringen bevatten, maar ook belangen van de betreffende organisaties die niet per se het algemeen belang vertegenwoordigen. Voorts hebben we nu vooral gesproken met koepels van lerarenopleidingen. Er is slechts in beperkte mate met leraren zelf en de directeuren van lerarenopleidingen gesproken. Het is de moeite waard om in de toekomst verder met vertegenwoordigers van individuele lerarenopleidingen te praten over de mogelijkheden en haalbaarheid van een centrale database.

Een andere kanttekening bij ons onderzoek is dat we in onze gesprekken meer in algemene zin hebben gesproken over het belang van een veelomvattende database van onderzoeksgegevens die inzicht bieden in de kwaliteit van startende leraren en lerarenopleidingen. Veel gesprekspartners vonden dit uitgangspunt te globaal en vonden het opbouwen van een database voor (startende) leraren en lerarenopleidingen zonder specifiekere onderzoeksvragen niet gewenst. Voordat verder wordt gegaan met de opbouw van een database over (startende) leraren en lerarenopleidingen dient daarom duidelijk te zijn wat vanuit het perspectief van beleid, onderwijsinstellingen en wetenschap de functie is van een dergelijke database. Een belangrijk aandachtspunt, gereleateerd aan de functie van de database, is wellicht ook de wijze waarop de database

van startende leraren en lerarenopleidingen wordt ingericht. In plaats van een complete database waarin alle mogelijke databestanden aan elkaar worden gekoppeld zou men ook kunnen denken aan een 'romp' van databestanden van CBS-, DUO- en eventueel IPTO-gegevens.

In samenhang met het bespreken van de functie van de database, het formuleren van de onderzoeksvragen en het vaststellen van de informatiebehoefte, kan er draagvlak gecreëerd worden onder de relevante partijen. Dit is vooral nodig bij de lerarenopleidingen. Belangrijk hierbij is het aanbrengen van focus, het formuleren van realistische doelstellingen, een goede planning en het doorlopend evalueren van opbrengsten. Het gaat dus om een stapsgewijze opbouw, waarbij wordt begonnen met 'laaghangend fruit' (data die gemakkelijk te verkrijgen zijn) en stap voor stap wordt uitgebreid met 'moeilijkere' typen gegevens. Een eerste traject met voorlopers van lerarenopleidingen zou bijvoorbeeld kunnen eindigen in een aantal vergelijkbare instellingsrapportages met indicatoren die tegemoet komen aan de informatiebehoefte van onderwijsinstellingen.

Van belang is verder dat duidelijk wordt welke kosten er verbonden zijn aan de opzet van een database voor startende leraren en lerarenopleidingen. Om het gebruik van een door het CBS beheerde systematische database te stimuleren zou er enige naamsbekendheid aan moeten worden gegeven. Wellicht kan tevens enige subsidie worden verleend aan het gebruik en/of kan als voorwaarde worden gesteld dat onderzoeksbureaus de database benutten bij de toekenning van onderzoeksopdrachten door OCW. Door beheer door het CBS zou ook bespaard kunnen worden op de kosten van eenvoudige monitoring van dezelfde indicatoren en respondenten in verschillende onderzoeken. Tot slot zou op termijn onderzocht moeten worden wat de mogelijkheden en kosten zijn van het koppelen van de database van startende leraren en lerarenopleidingen aan de NCO-database van leerlingen. Op die manier kan de kwaliteit van startende leraren en lerarenopleidingen worden gerelateerd aan leerlingresultaten.

# 01

## INLEIDING

### 1.1 Aanleiding en doelstelling

In de Lerarenagenda 2013-2020 'De leraar maakt het verschil' is een aantal belangrijke uitdagingen en verbeterpunten voor lerarenopleidingen opgenomen. De lerarenopleidingen hebben in reactie hierop reeds veel inspanningen verricht om hun opleidingen te verbeteren. Zo hebben anno 2016 bijna alle opleidingen een onderzoekslijn ingevoerd, flexibele leerroutes geïmplementeerd, beoordelingscriteria aangescherpt, studieloopbaanbegeleiding versterkt, en meer internationale oriëntatie aangebracht. Ook besteden lerarenopleidingen steeds meer aandacht aan intakeprocedures en strengere selectie (pabo's hanteren nu verplichte entreetoetsen) om zo het instroomniveau van de studenten te verhogen. Daarnaast organiseren alle lerarenopleidingen zogenaamde 'kennistoetsen' tijdens de studietrajecten om het kennisniveau van hun studenten op peil te houden en stemmen diverse lerarenopleidingen de kennisvereisten met elkaar af (Nederlands-Vlaamse Accreditatieorganisatie [NVAO], 2016).

Een belangrijke vraag die tot nog toe onvoldoende beantwoord kan worden, is in hoeverre men met deze maatregelen er daadwerkelijk in is geslaagd om de kwaliteit van het onderwijs van de nieuwe cohorten leraren te verbeteren. Hoewel verschillende onderzoeken zich reeds hebben gericht op de kenmerken van (startende) leraren, zoals opleiding, ervaring en geslacht, kennen deze een aantal beperkingen. Zo zijn studies naar lerarenkenmerken en daarmee samenhangende factoren vaak gefragmenteerd of eenmalig. Daarnaast zijn ze moeilijk te relateren aan bepaalde lerarenopleidingen en is er vaak geen relatie te leggen met de genomen maatregelen binnen deze lerarenopleidingen. Naar aanleiding van deze beperkingen, is de gedachte ontstaan om op zoek te gaan naar mogelijkheden om meer systematisch structurele gegevens over leraren (en lerarenopleidingen) te verzamelen ten behoeve van langetermijnonderzoek, leidend tot een algemene 'database met gegevens over studenten in lerarenopleidingen, de lerarenopleidingen zelf en de startende leraren' (vanaf hier: 'database'). De nadruk ligt bij een dergelijke database voornamelijk op afgestudeerden van lerarenopleidingen en startende leraren om snel inzicht te krijgen in de veranderingen van de kenmerken en

de kwaliteiten van het lerarenkorps van de nabije toekomst. Bovendien is de kans dat leraren het onderwijs verlaten het grootst in de eerste jaren na hun afstuderen.<sup>1</sup>

Met de database wordt zoals gezegd beoogd om op een meer systematische wijze structurele gegevens over (startende) leraren en lerarenopleidingen te verzamelen en aan elkaar te relateren. Hierdoor wordt het mogelijk om meer verdiepend en verklaarend onderzoek uit te voeren naar de kwaliteit(sbeïnvloedende factoren) van leraren en om de rol van de lerarenopleidingen binnen dit geheel beter in kaart te brengen. Er kan bijvoorbeeld meer inzicht verkregen worden in de gevolgen c.q. de effectiviteit van (potentieel kwaliteitsbevorderende) maatregelen binnen lerarenopleidingen, zoals het aanscherpen van selectie-/instroomcriteria. Ook valt te denken aan onderzoek naar verschillen in curricula tussen lerarenopleidingen en de mogelijke effecten hiervan op afgestudeerden: leiden verschillen in de curricula ook tot 'verschillende' typen leraren en waarin verschillen zij dan? Daarnaast biedt een database de mogelijkheid om meer gedetailleerde arbeidsmarktinformatie te genereren. De database zou veel gemakkelijker en eerder nieuwe inzichten en onderzoeksmogelijkheden op moeten leveren voor wetenschappers, beleidsmakers en onderwijsinstellingen dan op basis van de gefragmenteerde data die nu her en der voorhanden zijn.

Het Nationaal Regieorgaan Onderwijsonderzoek (NRO) heeft het Researchcentrum voor Onderwijs en de Arbeidsmarkt (ROA) verzocht een verkennend onderzoek uit te voeren naar de potentie en wenselijkheid van een dergelijke database. Meer specifiek dient te worden nagegaan a) welke kennisbehoefte er bestaat bij de mogelijke afnemers uit de wetenschap, het beleid en het onderwijsveld en in hoeverre een database hieraan kan bijdragen; b) welke (onderzoeks)bestanden over leraren (c.q. lerarenopleidingen) relevant zijn om op te nemen in een database; en c) in hoeverre bestaande gegevens (nu en in de toekomst) ontsloten en gekoppeld kunnen worden in een database en welke hindernissen daarbij overwonnen moeten worden.

## 1.2 Vraagstelling en onderzoeksopzet

In dit verkennende onderzoek wordt onder meer nagegaan welke kennisbehoeften er in het veld bestaan, hoe een database hierop van meerwaarde kan zijn, en welke inspanningen nodig zijn om een database op te zetten. De bedoeling is om vast te stellen op welke wijze het een en ander het beste georganiseerd en geregeld kan worden, en welke hordes daarbij genomen moeten worden. Voor het welslagen van de database is het van belang om al in een vroeg stadium meer inzicht te krijgen in de volgende vragen:

1. Welke bestaande data en onderzoeken zijn relevant voor een database over (startende) leraren en lerarenopleidingen, onder andere ten aanzien van de kwaliteit van leraren, de arbeidsmarktpositie, de curricula van de lerarenopleidingen en de kenmerken van de ontvangende scholen?

<sup>1</sup> Fontein et al. (2016), *Loopbaanmonitor onderwijs*, Tilburg: CentERdata.

1. In hoeverre zijn de beheerders/eigenaren van data bereid om hun onderzoeksresultaten beschikbaar te stellen voor een database?
2. Welke van deze data zijn het meest geschikt om in de database op te nemen?
3. Wat zijn mogelijke juridische en technische hindernissen rondom een database?
4. Is er draagvlak onder de vertegenwoordigers van de lerarenopleidingen voor een database?
5. Wat zijn de mogelijke opbrengsten, voordelen en nadelen van een database volgens beleidsmakers, wetenschappers en mensen uit het onderwijsveld?

Om antwoord te verkrijgen op bovenstaande vragen, is in de afgelopen maanden door de onderzoekers van het ROA een groot aantal bestaande databestanden geïnventariseerd en zijn er vele gesprekken gevoerd met diverse betrokkenen uit het veld. Het gaat daarbij om vertegenwoordigers van de lerarenopleidingen zoals verenigd in de koepelorganisaties ICL, LOBO en ADEF, om dataspecialisten (c.q. beleidsadviseurs) van lerarenopleidingen, alsmede om representanten van het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW), van het CBS, van het CPB en van DUO. Tot slot is gesproken met onderwijsonderzoekers: Prof. Lex Borghans, Prof. Paulien Meijer, Prof. Rolf van der Velden, Prof. Dinand Webbink, Prof. Klaas van Veen.<sup>2</sup> Het voorliggende rapport is gebaseerd op deze gesprekken, aangevuld met deskresearch naar bestaande, potentieel interessante databestanden.

### 1.3 Leeswijzer

Het rapport laat zich als volgt lezen. In hoofdstuk 2 geven we eerst een overzicht van de bestaande data over de kenmerken van leraren en gaan we na welke van deze het meest geschikt lijken om in de database op te nemen. In de daaropvolgende hoofdstukken komen de gesprekken met de diverse beheerders/verzamelaars van belangrijke leraar-/onderwijsdata aan bod. Daar gaan we in op de visie van de gesprekspartners op de database en de potentiële meerwaarde ervan, maar vooral ook op mogelijkheden van de databestanden waarover zij beschikken (i.e. de inhoud, ontsluiting en koppelbaarheid van de bestanden). Achtereenvolgens geven we een weergave van de gesprekken met het CBS (hoofdstuk 3), DUO (hoofdstuk 4), de Onderwijsinspectie (hoofdstuk 5), de lerarenopleidingen (hoofdstuk 6) en de Onderwijscoöperatie (over het Lerarenregister, hoofdstuk 7). In hoofdstuk 8 worden enkele databestanden geïntroduceerd die eveneens relevant kunnen zijn voor een database, zonder in te gaan op alle details van een

---

<sup>2</sup> De verkennende gesprekken met Prof. Meijer en Prof. Van Veen zijn vooral gevoerd vanuit hun rollen als respectievelijk vice-voorzitter en voorzitter van het ICL (zie hoofdstuk 6). Daarnaast is vanzelfsprekend ook in algemene zin kort gesproken over de wetenschappelijke mogelijkheden van een database. Met Prof. Van Veen is tevens kort gesproken over het onderzoeksproject Begeleiding Startende Leraren (voor een korte introductie in BSL zie hoofdstuk 8). In het gesprek met Prof. Van der Velden is vooral ingegaan op zijn rol binnen het Nationaal Cohortonderzoek Onderwijs (NCO), zie hoofdstuk 12.

---

opname van deze databestanden in de database. In de daaropvolgende hoofdstukken komen de gesprekken met diverse andere sleutelpartijen aan bod, te weten met het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW, hoofdstuk 9), het CPB (hoofdstuk 10) en onderwijsonderzoekers (hoofdstuk 11). In hoofdstuk 12 staan we stil bij de lessen die uit de aanpak van het Nationaal Cohortonderzoek Onderwijs (NCO) kunnen worden geleerd met betrekking tot de juridische en technische hindernissen bij het opzetten van een nationale database. Tot slot volgen in hoofdstuk 13 de belangrijkste conclusies van het onderzoek en bespreken we de eventuele vervolgstappen op basis van deze verkenning en de voornaamste hindernissen die daarbij kunnen spelen.



# 02

## OVERZICHT VAN RELEVANTE DATA EN ONDERZOEKEN VOOR EEN DATABASE

Als startpunt van dit rapport geven we een overzicht van de data en onderzoeken die relevant kunnen zijn voor een database over (startende) leraren en lerarenopleidingen (Tabel 2.1). De data hebben betrekking op verschillende typen bestanden met microdata (i.e. over individuele personen) over uiteenlopende onderwerpen. Het gaat bijvoorbeeld om gegevens over leraren betreffende hun opleiding, de overgang van opleiding naar werk, de baankenmerken, de professionalisering gedurende de loopbaan, enzovoorts. Hoewel de tabel niet beschouwd mag worden als een uitputtend overzicht van gegevens over leraren (inclusief studenten en afgestudeerden aan lerarenopleidingen), gaat het wel over de meest relevante bestanden die de afgelopen jaren verzameld zijn over deze groep. We hebben er voor gekozen om hoofdzakelijk databestanden op te nemen die structureel van aard zijn (i.e. doorlopend) en die betrekking hebben op meerdere jaren. Incidentele leraargegevens, bijvoorbeeld vanuit eenmalige thema-onderzoeken, kunnen weliswaar voor specifieke onderzoeks- of beleidsvragen relevant zijn, maar bieden niet de mogelijkheid om trends over de tijd, dan wel ontwikkelingen op individueel niveau in kaart te brengen. Bovendien voert het voor dit verkennende onderzoek te ver om een overzicht te geven van alle incidentele databestanden over leraren van onderzoeken uit het verleden.

We kunnen de verschillende bestanden met microdata ruwweg indelen in enerzijds 'registerdata' (gebaseerd op administratieve gegevens) en anderzijds 'enquêtedata' (verkregen door vragenlijsten of lesobservaties). Bij registerdata worden er in principe gegevens uit bestaande administraties van instellingen verzameld. Zo worden bijvoorbeeld voor de 'Integrale Personeelstelling Onderwijs' (IPTO) lesgegevens van leraren verzameld vanuit een onderzoeksopdracht van OCW. Ze zijn grotendeels afkomstig uit schooladministraties (roosterpakketten) die door een onderzoeksinstituut verzameld en vervolgens gekoppeld worden aan DUO-gegevens (waarna ze weer 'verrijkt' worden teruggestuurd aan DUO). Ook de via het CBS beschikbaar gestelde studentgegevens (zoals inschrijvingen, diploma's en dergelijke) zijn een duidelijk voorbeeld van registerdata (op basis van onderwijsadministraties van DUO), terwijl de gegevens uit de HBO-Monitor een duidelijk voorbeeld zijn van enquêtedata (vragenlijstonderzoek naar de positie van hbo-alumni).

De volgorde waarin de databestanden zijn weergegeven in Tabel 2.1 is gebaseerd op de ingeschatte relevantie van de betreffende bestanden voor een 'database', en is gebaseerd op een aantal factoren. Deze factoren zijn: (1) de veronderstelde inhoudelijke meerwaarde voor een database, (2) de omvang / het bereik van het bestand, (3) de mogelijkheid tot afbakening van de doelgroep (bij voorkeur data over startende leraren) en (4) de mate waarin de dataverzameling structureel van aard is. De ingeschatte relevantie van de verschillende databestanden voor een database wordt geïndiceerd door het aantal stippen in de rechterkolom. Deze voorlopige, richtinggevende 'rangschikking' heeft voor een belangrijk deel input geleverd voor de selectie van de gesprekspartners en de inhoud van de gesprekken in dit verkennende onderzoek. We hebben ons bij de gesprekken namelijk vooral gericht op de beheerders en de opdrachtgevers van de databestanden met de grootste ingeschatte relevantie voor een database. De daadwerkelijke inhoudelijke relevantie van databestanden en de haalbaarheid om de bestanden te kunnen benutten voor een database kan pas na individuele gesprekken met de beheerders en de opdrachtgevers van de databestanden worden vastgesteld. In de volgende hoofdstukken wordt over de inhoud van deze individuele gesprekken verslag gedaan.

Databestanden die minder dan drie stippen krijgen komen in het vervolg nauwelijks of niet aan de orde. De redenen voor het geringe aantal stippen van deze bestanden zijn uiteenlopend. Meerdere databestanden richten zich niet specifiek op de onderwijssectoren, en bevatten daardoor relatief weinig variabelen die relevant zijn voor het onderzoek naar het beleid gericht op de leraren en de lerarenopleidingen. Voorbeelden hiervan zijn de Enquête Beroepsbevolking (EBB), de Nationale Enquête Arbeidsomstandigheden (NEA) en de Nationale Studenten Enquête (NSE). Dat wil overigens niet zeggen dat deze enquêtes helemaal geen interessante gegevens bevatten over leraren en lerarenopleidingen. De variabelen over bijvoorbeeld het ziekteverzuim en het oordeel over de opleidingen door studenten kunnen wel degelijk dienen als input van stuurinformatie voor beleid, onder andere omdat er vergelijkingen met sectoren buiten het onderwijs mogelijk zijn. Toch schatten wij op voorhand de relevantie voor onderzoeks- en beleidsdoeleinden in het kader van de (startende) leraren en de lerarenopleidingen niet heel hoog in. Andere databestanden met maximaal twee stippen (Talis, Studie & Werk, POMO) komen niet jaarlijks uit, bevatten relatief weinig respondenten of zijn om diverse redenen moeilijk te koppelen zijn aan de belangrijke registerbestanden van DUO en het CBS.

**TABEL 2.1** Overzicht van relevante data en onderzoeken voor een database voor (startende) leraren en lerarenopleidingen

Type bestand c.q. naam onderzoek	Databeheerder (tussen haakjes eventuele opdrachtgever)	Doelgroep en bereik	Methode <sup>1</sup>	Periode en frequentie	Inhoud	Ingeschatte relevantie
1 Studentgegevens (inschrijvingen, diploma's etc.)	CBS	Studenten in bekostigd onderwijs: vrijwel integraal	Administratie	1999- nu; jaarlijks	Studentgegevens, zoals opleidingsnummer, (versleutelde) schoolcode, onderwijssoort, details over de betreffende onderwijssoort	*****
2 Registerdata sociaal-statistische bestanden (achtergrond, baangegevens etc.)	CBS	Nederlandse bevolking: vrijwel integraal	Administratie	1995- nu; jaarlijks	Grote variatie aan demografische, sociaal-culturele en sociaal-economische gegevens	*****
3 Formatie- en verzuimbestanden onderwijspersoneel	DUO	Onderwijspersoneel: vrijwel integraal (voor zover in dienst bij scholen)	Administratie	1994- nu; jaarlijks	Informatie aanstellingen, zoals brin, leeftijd, geslacht, functie, schaal etc. Tevens informatie over verzuim.	*****
4 Integrale Personeelstelling Onderwijs (IPTO)	DUO Centerdata/DUO (OCW)	Lesgegenden in vo: vrijwel integraal	Administratie (afkomstig van roosterpakketten aangevuld met schriftelijke data)	1994- nu; jaarlijks	Leraar- en lesgegevens ( o.a. functie, bevoegdheid; gegeven vak, onderwijstype, leerjaar, lesduur)	*****
5 Lerarenopleidingen aan studentgegevens	Lerarenopleidingen aan hbo- en wo-instellingen	Studenten aan hbo- en wo-lerarenopleidingen: vermoedelijk integraal	Toetsing en monitoring	doorlopend	Entree-toetsen, voortgangstoetsen, stagebeoordelingen etc.	*****
6 Inspectiedata <sup>3</sup>	Inspectie van het Onderwijs	Onderwijs in geheel (leraren, leerlingen, scholen etc.). Bereik verschillend per type data	Vershillende methodes voor verschillende data, o.a. lesobservaties, vragenlijsten	verschillend	Zeer divers. Wat betreft leraren o.a. steekproefsgewijs: lesobservaties, opvattingen etc.	****
7 Lerarenregister	CIBG (Onderwijs-coöperatie)	Leraren: tot nu vrijwillige aanmeldingen, wordt omgezet naar verplicht register (dus ruwweg	Administratie	2012- nu; doorlopend	Beroepsregister bijgehouden door leraren. Informatie over bevoegdheden en bekwaamheden, zoals	****

**TABEL 2.1** Overzicht van relevante data en onderzoeken voor een database voor (startende) leraren en lerarenopleidingen

Type bestand c.q. naam onderzoek	Databeheerder (tussen haakjes eventuele opdrachtgever)	Doelgroep en bereik	Methode <sup>1</sup>	Periode en frequentie	Inhoud	Ingeschatte relevantie
8 Begeleiding Startende Leraren	Verschillende uo's, hoofduitvoering Rijksuniversiteit Groningen (OCW)	Startende leraren in vo: aanmelding door scholen, per cohort 800-1500 starters	Meervoudig: Lesobservatie; vragenlijsten onder startende leraren en leerlingen	2014-nu; voorsnog doorlopend	O.a. competenties van starters, perceptie over begeleiding, kijk van leerling op startende leraar	***
9 HBO-Monitor	ROA (Vereniging Hogescholen)	Afgestudeerden aan hbo: respons ong. 23.000, ca. 3.000 van lerarenopleidingen	Vragenlijst hboers 1,5 jaar na afstuderen	1996-nu; jaarlijks	Arbeidsmarktpositie afgestudeerden en opvattingen over voorbereiding op arbeidsmarkt	***
10 Loopbaanmonitor <sup>4</sup>	CentERdata & MOOZ (OCW)	Afgestudeerde leraren: respons vragenlijst ong. 2500 leraren; microdata integraal	Vragenlijsten (2x) afgestudeerden lerarenopleidingen; analyses CBS-microdata	2004-nu; jaarlijks	Arbeidsmarktpositie beginnende leraren en ontwikkelingen in begeleiding / ondersteuning	***
11 Enquête-Beroepsbevolking (EBB)	CBS	Nederlandse beroepsbevolking: respons > 100.000	Vragenlijst onder personen > 15 jaar via roterend panel (5 metingen)	1987-nu; doorlopend	Arbeidsmarktpositie. O.a. over hebben van betaald werk, arbeidsduur, beroep en bedrijf, opleiding	**
12 Nationale Enquête Arbeidsomstandigheden (NEA)	TNO & CBS (SZW)	Werknemers NL; respons ong. 42.000 waarvan 3.000 onderwijs	Vragenlijst onder werknemers	2003-nu; jaarlijks	Arbeidsomstandigheden, o.a. werktijden, overwerk, werkdruk, agressie, pesten en discriminatie	**
13 Nationale Studenten Enquête (NSE)	GfK (Studiekeuze123)	Studenten hbo en wo: respons ong. 300.000 studenten	Vragenlijst onder huidige studenten	2009-nu; jaarlijks	Mening over opleiding, o.a. inhoud opleiding, vaardigheden, voorbereiding loopbaan	**
14 Personeelonderzoek / Mobiliteitsonderzoek (POMO)	ICTU & GfK (BZK)	Medewerkers publieke sector: respons ong. 24.000 waarvan 6.000 po, vo, mbo	Vragenlijsten onder medewerkers publieke sector	1999-nu; tweesjaarlijks	Arbeidsmarktpositie ambtenaren. O.a. mobiliteit, inzetbaarheid, werkbeleving	**

**TABEL 2.1** Overzicht van relevante data en onderzoeken voor een database voor (startende) leraren en lerarenopleidingen

Type bestand c.q. naam onderzoek	Databeheerder (tussen haakjes eventuele opdrachtgever)	Doelgroep en bereik	Methode <sup>1</sup>	Periode en frequentie	Inhoud	Ingeschatte relevantie
15	Studentenmonitor	Studenten hbo en wo: respons ong. 20.000 studenten	Vragenlijst onder huidige studenten	2000-nu; jaarlijks	Opvattingen / gedragingen over studie (o.a. keuzes, motivatie)	**
16	Teaching and Learning International Survey (Talis)	Leraren en schoolleiders vo onderbouw: respons ong. 2.200 leraren	Vragenlijst leraren en schoolleiders	2008-nu; vijfjaarlijks	Landenvergelijkend onderzoek naar leer- en werkomstandigheden van leraren	**
17	Studie & Werk	Afgestudeerden aan hbo en wo: respons ong. 2.700 hbo'ers en 2.800 wo'ers.	Vragenlijst onder afgestudeerden hbo en wo	2009-nu; jaarlijks	Arbeidsmarktpositie van afgestudeerden (o.a. zoekgedrag, arbeidsvoorwaarden)	**
18	Nationale Alumni Enquête (NAE)	Afgestudeerden wo: respons ong. 8.000, waarvan ong. 160 wo-onderwijs	Vragenlijst onder wo'ers, 1 tot 1,5 jaar na afstuderen	2009-nu; tweesjaarlijks	Arbeidsmarktpositie afgestudeerden en opvattingen over voorbereiding arbeidsmarkt	*
19	Onderwijs werkt!	Leraren en schoolleiders po, vo, mbo, hbo: respons ong. 6000 leraren en 600 schoolleiders	Vragenlijst onder werkende leraren en schoolleiders	2013-2015 (beëindigd); jaarlijks	O.a. activiteiten en opvattingen tevredenheid scholing, personeelsbeleid en begeleiding (starters)	*
20	Staat van de Leraar	Leraren po, vo, mbo, so: respons ong. 750	Vragenlijst onder werkende leraren	2015-nu; jaarlijks	Opvattingen en activiteiten op gebied van professionalisering	*
21	Startmonitor	Eerstejaars hbo en wo: ong. 9.400 eerstejaars (2008)	Vragenlijst (3x) onder eerstejaars hbo/wo	2008-nu; jaarlijks	O.a. over studiekeuze, voorlichtinggebruik, aansluiting met de vooropleiding, studiegedrag	*

**Noten:**

- 1) In de loop der tijd kan de methodiek veranderd zijn, we baseren ons in dit overzicht op de meest actuele methodiek (voor zover bekend).
- 2) De periode en updatefrequentie van data die via het CBS beschikbaar worden gesteld, verschilt van bestand tot bestand.
- 3) Het gaat hier om een grote variatie aan verschillende typen data die ten bate van de overzichtelijkheid hier onder één noemer 'inspectedata' worden geschaard. Nadere contacten met Inspectie moeten duidelijk maken wat dit zoal behelst en in hoeverre data beschikbaar kunnen worden gesteld. Naast data over onderwijskwaliteit beschikt de Inspectie ook over onderzoeksgegevens. In dit kader is het onderzoek 'Beginnende leraren kijken terug' inhoudelijk zeer relevant, dit betreft echter een (vooral) eenmalig onderzoek. Wanneer dit onderzoek een structureel karakter zou krijgen, zou het een uiterst relevante bron zijn.
- 4) De Loopbaanmonitor bestaat voor een deel uit analyse op CBS-microdata. Aangezien we CBS-microdata (stelsel van sociaal-statistische bestanden) al eerder in dit overzicht afzonderlijk benoemen, is hier vooral het vragenlijst-gedeelte van de Loopbaanmonitor interessant als aanvullende bron voor een database.



# 03

## CBS: BEHEER EN MICRODATA

### 3.1 Inleiding

Het Centraal Bureau voor de Statistiek (CBS) verzamelt, bewerkt en publiceert statistische gegevens over de Nederlandse samenleving (personen, bedrijven en instellingen) en mag deze ook koppelen aan andere persoonsgegevens in het kader van statistiek. Dit is wettelijk verankerd in de *Wet op het Centraal bureau voor de statistiek*.<sup>3</sup> Gezien de belangrijke rol van het CBS bij de verzameling van persoonsgegevens, de grote hoeveelheid microdata die het CBS ontsluit, en de prominente rol die het speelt in het Nationaal Cohortonderzoek Onderwijs<sup>4</sup> (NCO, zie Hoofdstuk 12) is het de meest voor de hand liggende partij die een eventuele database zou kunnen (en mogen) beheren. Het gesprek met diverse representanten van het CBS<sup>5</sup> vormt daarom het startpunt van dit verkennende onderzoek.

### 3.2 Visie op database

#### *Enthousiasme over idee van database*

Het CBS is enthousiast over het idee van een database. Voor het NCO zijn reeds slagen gemaakt op het gebied van bestandskoppelingen op leerlingniveau, maar tot op heden ontbreekt de rol van de leraar in het NCO. De leraar heeft echter een centrale rol in de ontwikkeling van leerlingen. Het CBS acht het daarom nuttig om meer te weten te komen over de leraar. Door een brede database op te bouwen kan verdiepend onderzoek uitgevoerd worden en kan er voor veel variabelen 'gecontroleerd' worden, waardoor onderzoeksresultaten sterker onderbouwd worden.

<sup>3</sup> Het ten behoeve van statistische doeleinden gebruiken van het Burgerservicenummer (BSN) en verwerken van persoonsgegevens als bedoeld in artikel 16 van de *Wet bescherming persoonsgegevens* (WbP) is geregeld in de artikelen 34 en 35 van de CBS-wet.

<sup>4</sup> Het NCO betreft een nieuw type leerlingcohort-onderzoek waarbij (zoveel) mogelijk gebruik gemaakt wordt van bestaande gegevens (registerdata, toetsgegevens et cetera), zie hoofdstuk 12 voor meer informatie.

<sup>5</sup> De hele lijst met gesprekspartners in dit verkennende onderzoek is te vinden in de bijlage.

Het CBS (h)erkent dat het huidige 'datalandschap' rondom leraren erg gefragmenteerd is. Het zou daarom goed zijn om een meer geïntegreerd bestand te ambiëren waarin verschillende onderzoeksbestanden samenkomen. Ook kennen onderzoeken vaak hun eigen beslisregels en afbakeningen: enerzijds is dat zinvol naar gelang de onderzoeksvraag, maar het lijkt volgens het CBS zinvol om meer vanuit een zekere gedeelde basis naar bestanden te kijken, waarbij standaarddefinities worden gehanteerd. Wel moet daarnaast altijd flexibiliteit blijven bestaan voor specifieke vragen.

Bij het ontwikkelen van een database is het belangrijk om het veld hier goed bij te betrekken. Belangrijke spelers zijn daarin onder meer de Inspectie, OCW, DUO, onderzoeksinstituten, enzovoorts. Het is belangrijk om bewust te zijn van mogelijke scepsis in het veld, met name bij onderwijsinstellingen en lerarenopleidingen. Om het draagvlak in het onderwijs te vergroten is het volgens het CBS belangrijk ze iets 'terug te geven' waar ze echt wat aan hebben, bijvoorbeeld benchmarkrapportages (zoals in het NCO).

### *Praktische overwegingen*

De ambitie om een database te ontwikkelen is volgens het CBS goed maar hoog gegrepen. Het CBS adviseert daarom om dit stapsgewijs aan te pakken. Bijvoorbeeld door eerst op enkele basisbestanden vernieuwende analyses te doen en hierover te rapporteren, om vervolgens te kijken waar witte vlekken liggen die met aanvullende data weggenomen zouden kunnen worden. Het is belangrijk om regelmatig de 'wins' te benoemen, daarmee toon je aan dat je op de goede weg bent en laat je de vernieuwing en meerwaarde zien, maar laat je ook zien wat nog niet mogelijk is.

Bij het ontwikkelen van een database moet verder goed nagedacht worden over welke bestanden de basis dienen te vormen. Het is immers niet opportuun om alle mogelijke databestanden over leraren aan elkaar te koppelen in één onderzoeksbestand. Vaak is immers slechts een deel van de bestaande gegevens nodig voor een specifieke onderzoeksvraag. Zo kunnen gegevens over de herkomst van leraren (bijvoorbeeld informatie over hun ouders) vaak niet, maar soms ook wél interessant zijn. Een kerndatabase die naar gelang de onderzoeksvraag aangevuld kan worden met 'modules' is dan het meest verstandig en praktisch. Achter de schermen bij het CBS kunnen de gegevens dan natuurlijk wel al 'klaar' staan.

Wanneer er meer duidelijkheid bestaat over welke bestanden de kerndatabase vormen, dient vervolgens nagedacht te worden over de frequentie van updaten en aanvullen. Afhankelijk van de insteek kan het uitgangspunt bijvoorbeeld zijn om de nieuwste gegevens te gebruiken (oftewel vervanging van bestanden) of juist om gegevens over meerdere momenten te benutten om individuele ontwikkelingen in kaart te brengen (oftewel aanvulling van bestanden). Om een eenvoudige verwerking van verbeterde of nieuwe gegevens mogelijk te maken kan volgens het CBS het beste gekozen worden voor losse, koppelbare bestanden per onderwerp. Als de insteek echter vooral gericht



is op lerarenloopbanen in hun geheel, dan adviseert het CBS om een geïntegreerd bestand te maken (vergelijkbaar met de onderwijsloopbanen in het NCO).

### 3.3 Aanwezige data

- Het CBS beschikt over een grote hoeveelheid databestanden die grotendeels beschikbaar gesteld worden aan geautoriseerde onderzoekers als zogenaamde 'microdata'. Het aantal databestanden (en het aantal overkoepelende 'thema's') is zeer groot. Zo zijn er microdatabestanden op het gebied van arbeid en sociale zekerheid; gezondheid en welzijn; onderwijs; veiligheid en recht, enzovoorts. In het kader van de database zijn in het bijzonder de diverse onderwijsbestanden relevant, maar ook de grote hoeveelheid andere data uit het zogenaamde 'stelsel van sociaal-statistische bestanden'. Dit stelsel wordt gevormd door (onderling koppelbare) bestanden met onder andere demografische, sociaal-culturele en sociaal-economische gegevens. Zo zou een eventuele database in de basis kunnen bestaan uit de (reeds toegankelijke) onderwijsbestanden, zoals ingeschrevenen en gediplomeerden aan lerarenopleidingen.<sup>6</sup> Deze onderwijsbestanden zouden volgens het CBS bijvoorbeeld aangevuld kunnen met demografische bestanden (leeftijd, geslacht, herkomst, et cetera)
- Arbeidsbestanden, zoals POLIS-bestanden (sector, inkomen, aanstelling, et cetera)
- Enquêtedata die als CBS-microdata beschikbaar worden gesteld, zoals de Enquête Beroepsbevolking (EBB) over de arbeidsmarktpositie van Nederlanders, en de Nationale Enquête Arbeidsomstandigheden (NEA) naar de werksituatie van werkenden (zoals werkdruk, opleiding, en ontwikkeling et cetera).

Gezien de enorme hoeveelheid databestanden die benut kunnen worden als microdata, voert het hier te ver om deze uitgebreid te bespreken. Het is in ieder geval duidelijk dat dit type bestanden inhoudelijk (en gezien de onderlinge koppelbaarheid) zeer bruikbaar is voor onderzoek naar leraren. Een groot nadeel van de bestaande microdata is echter dat er in principe geen specifieke functies in onderscheiden kunnen worden;<sup>7</sup> wel kunnen cao-sectoren herkend worden. Met de huidige data kan dus wel worden gezien of iemand in het onderwijs werkt en met welke diploma, maar niet of dat deze persoon ook daadwerkelijk als leraar werkt.<sup>8</sup> Dit is een grote belemmering, maar er lijkt goed nieuws op komst. Het CBS geeft aan dat er in de toekomst vermoedelijk DUO-onderwijspersoneelsgegevens ontsloten zullen worden. Het CBS spreekt momenteel met DUO over het beschikbaar stellen van dergelijke bestanden. Op het moment

6 Een overzicht van de onderwijsbestanden die het CBS beschikbaar stelt is te vinden op <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata/onderwijs>

7 Dit geldt niet voor de Enquête Beroepsbevolking (EBB), daarin zijn beroepen wél te onderscheiden (ISCO-codes). De EBB betreft echter een roterend panelonderzoek en is dus niet integraal.

8 Tegen deze beperking loopt men vooralsnog bijvoorbeeld aan in de Loopbaanmonitor Onderwijs, waarin aan de hand van CBS-microdata inzicht gegeven wordt in de arbeidsmarktpositie van afgestudeerden aan lerarenopleidingen.

van het schrijven van dit rapport is het nog niet duidelijk of het alleen om administratieve gegevens gaat (zoals school, functie, aanstelling, et cetera) of ook om meer lesinhoudelijke gegevens (bijv. gegeven vakken, uren, klassen vanuit IPTO). In ieder geval zal het met de toekomstige DUO-gegevens wél mogelijk zijn om leraren als dusdanig te kunnen identificeren, een ontwikkeling die van grote waarde is voor een database.

### 3.4 Databenutting: ontsluiting en koppelmogelijkheden

#### *Ontsluiting*

In principe zijn alle microdatabestanden die bij het CBS beschikbaar zijn, toegankelijk voor onderzoekers. Het moment waarop microdatabestanden beschikbaar worden gesteld voor onderzoek verschilt van bestand tot bestand, maar in algemene zin moet rekening worden gehouden met een vertragingfactor. De bestanden zijn (vanzelfsprekend) niet 'realtime'; zo zijn bijvoorbeeld de meest recente gegevens over onderwijsinschrijvingen bij het afronden van dit rapport (juni-juli 2017) beschikbaar over 2016<sup>9</sup>, terwijl de meest recentste gegevens over banen en lonen op basis van de Polisadministratie tot en met september 2016 beschikbaar zijn.<sup>10</sup>

Deze microdatabestanden zijn op individueel niveau onderling koppelbaar via versleutelde persoonscodes. Daarnaast zijn er mogelijkheden om externe microdatabestanden aan te leveren en deze te koppelen aan de andere bestanden.<sup>11</sup> De microdata kunnen slechts benut worden als aan strikte richtlijnen wordt voldaan.<sup>12</sup> Het gaat dan onder meer om autorisatie van gebruikers, onderbouwde projectaanvragen en zorgvuldigheid rond onthullingsrisico. Meer algemeen gesteld zijn microdata alleen te benutten voor statistisch of wetenschappelijk onderzoek, uitgevoerd door instellingen met een primaire onderzoeksdoelstelling, waarbij de onderzoeksresultaten openbaar gepubliceerd worden. Het CBS beschikt over faciliteiten om met deze gegevens te werken. Daarnaast beschikken ook diverse externe organisaties over zogenaamde 'remote access' faciliteiten, waarmee op afstand onderzoek verricht kan worden op de microdata. CBS-microdata zijn zoals gezegd, voor zover voldaan wordt aan de richtlijnen en eisen, breed toegankelijk voor onderzoekers. Het is dus niet zo dat alleen een beperkte groep gebruikers, bijvoorbeeld alleen analisten vanuit lerarenopleidingen of alleen onderzoekers binnen het database-project, toegang kunnen krijgen tot een database. De gegevens moeten breed benut kunnen worden.

9 Het betreft hier het bestand 'ONDERWIJNSCHRITAB' met 'voorlopige cijfers' over 2016 dat op een later tijdstip vervangen wordt door een versie met 'definitieve cijfers'.

10 Het betreft hier de versie 'Eerste plaatsing tot en met september 2016' van het bestand 'SPOLISBUS'.

11 Zie voor meer informatie: <https://www.cbs.nl/-/media/cbs%20op%20maat/zelf%20onderzoek%20doen/161221%20voorschrift%20bestandsformaat%20en%20uploadprocedure.pdf>

12 Zie <https://www.cbs.nl/-/media/8db64c6ff1ee41d5bce09e6b8628cff9.ashx>

### *Koppelmogelijkheden externe bestanden*

Voor het benutten van externe gegevens (dat wil zeggen gegevens die niet als CBS-microdata worden aangeboden) dienen twee typen gegevens onderscheiden te worden. Het eerste type gegevens betreft data die reeds door derden verzameld/aangeleverd worden in het kader van beleidsstatistiek, maar die niet standaard onderdeel zijn van de CBS-microdata. De benutting en koppeling van deze gegevens is volgens het CBS geen probleem, mits de betreffende organisaties de gegevens beschikbaar willen stellen. Het tweede type gegevens betreft data die niet standaard aan het CBS geleverd worden, bijvoorbeeld externe enquêtebestanden. Voor de benutting en koppeling van deze gegevens geldt 'informed consent'. Dit wil zeggen dat respondenten toestemming dienen te verlenen voor het benutten van de onderzoeksgegevens voor andere doeleinden dan het specifieke enquêtedoel. Hetzelfde geldt voor de benutting van studentgegevens voor zover deze nog niet standaard geleverd worden ten bate van beleid(sstatistiek).

Het CBS geeft aan dat er reeds meerdere relevante onderzoeken over leraren worden uitgevoerd door diverse partijen op verzoek van het Ministerie van OCW, het NRO, enzovoorts. Meestal wordt dan wel over de uitkomsten van het onderzoek gerapporteerd, maar worden de onderliggende data niet beschikbaar gesteld voor onderzoek door anderen.<sup>13</sup> Deze data blijven dan alleen beschikbaar voor de partij die het onderzoek uitgevoerd heeft. Het CBS adviseert opdrachtgevers daarom om in hun opdracht te laten opnemen dat het uitvoerende onderzoeksbureau voor een goede 'informed consent' moet zorgen, zodat alle onderliggende data via het CBS beschikbaar gesteld kunnen worden aan andere bevoegde onderzoekers. Ruwweg zijn er twee manieren om externe bestanden die niet standaard geleverd worden aan CBS te koppelen aan de microdata.

#### *Koppelmogelijkheid 1: extern bestand-koppelvariabele-CBS bestand*

Om externe gegevens, die niet al standaard/verplicht aan worden geleverd, te kunnen koppelen aan bestaande CBS-data is 'informed consent' nodig. Om koppelingen vervolgens technisch mogelijk te maken, is het benutten van het Burgerservicenummer (BSN) de ideale manier. Het BSN is voor de Nederlandse bevolking immers (bijna helemaal) integraal en is door het unieke nummer een ideale koppelvariabele. Het BSN mag echter alleen gebruikt worden door overheidsorganisaties om hun taak uit te voeren en door niet-overheidsorganisaties indien dit wettelijk bepaald is (zo is het gebruik van het BSN door zorgorganisaties bijvoorbeeld bepaald in de *Wet gebruik burgerservicenummer in de zorg*).<sup>14</sup> Indien het BSN niet beschikbaar is, bijvoorbeeld bij externe enquêtebestanden, is een alternatieve manier van koppelen mogelijk door een combinatie van diverse persoonsgegevens te hanteren. Het gaat dan in principe om de combinatie

<sup>13</sup> Bij opdrachten voor OCW is het een voorwaarde dat de onderzoeksbestanden door de opdrachtnemer opgenomen worden in het online archief 'DANS', maar dit wordt de facto niet (altijd) nageleefd, aldus het ministerie.

<sup>14</sup> Zie voor meer informatie: <https://www.rijksoverheid.nl/onderwerpen/persoonsgegevens/burgerservicenummer-bsn>

van de geboortedatum, het geslacht en het volledige adres (en het precieze moment waarop iemand op dat adres woont). Namen van personen zijn vaak minder bruikbaar als koppelvariabele, omdat ze erg gevoelig zijn voor typfouten en schrijfwijzen. E-mailadressen zijn ook niet bruikbaar als koppelvariabele, aangezien het CBS niet beschikt over een e-mailadressenbestand.

### *Koppelmogelijkheid 2: dataverzameling via CBS-aanschrijving*

Voor toekomstige dataverzameling via enquêtes zou het mogelijk zijn om een 'CBS-route' te nemen. Dit houdt in dat de aanschrijving verzorgd wordt door het CBS vanuit hun adressenbestanden. Het CBS heeft hier op voorhand geen bezwaren tegen en geeft aan om hier, indien er meer concrete plannen zijn, verder over te praten en hier een kostenindicatie van te geven. Het grote voordeel van deze 'route' is dat de koppelbaarheid met andere CBS-bestanden eenvoudig te bewerkstelligen is. Wel moet nadrukkelijk rekening worden gehouden met het volgende: als de dataverzameling via het CBS verloopt, neemt het CBS ook een inhoudelijke rol in bij het onderzoek. Die betreft bijvoorbeeld het meedenken over de inhoud van de vragenlijst en de steekproeftrekking. Ook is aan samenwerking met het CBS een aantal voorwaarden verbonden. Zo dient het onderzoek op hoofdlijnen aan te sluiten op de doelen en taken van het CBS; dient het CBS als volwaardige samenwerkingspartner in het proces te worden meegenomen; dient het CBS met de onderzoeksbevindingen naar buiten te mogen treden; en dient de samenwerkingspartner een goede reputatie te hebben. Ook is het belangrijk te vermelden dat de gegevens die via het CBS verzameld worden, in principe (in lijn met de andere micro-databestanden) altijd beschikbaar gesteld worden voor verder onderzoek.

Het verzamelen van 'externe' onderzoeksgegevens via het CBS gebeurt reeds bij een aantal samenwerkingsprojecten, bijvoorbeeld voor SCP-studies, voor de Nationale Enquête Arbeidsomstandigheden (TNO) en voor het vo- en mbo-schoolverlatersonderzoek (voor OCW). Voor een database met gegevens over leraren in opleiding, lerarenopleidingen en startende leraren is het denkbaar dat er een CBS-NRO samenwerking tot stand komt, waarmee specifieke aanvullende onderzoeksvragen beantwoord zouden kunnen worden. Belangrijke noot is echter dat het CBS zich niet inmengt in aanbestedingen van onderzoek. Veelal is dit volgens het CBS overigens ook niet nodig voor opdrachtgevers, aangezien zij vaak bij CBS mogen 'inbesteden'.

## Conclusies

- Het CBS staat bekend als dé gegevensverzamelaar in Nederland en diens taak is ook wettelijk verankerd. Het CBS mag in het kader van beleidsstatistiek persoonsgegevens verzamelen en data koppelen. Een groot deel van de gegevens die het verzamelt, worden beschikbaar gesteld als zogenaamde 'microdata'. Deze individuele, onderling koppelbare gegevens hebben betrekking op een zeer groot aantal thema's.
- Voor een database zijn onder meer relevant: studentgegevens, achtergrondgegevens en gegevens over de arbeidsmarktpositie van personen. Deze gegevens worden periodiek geüpdatet en met enige vertraging beschikbaar gesteld (de vertragingfactor verschilt per type gegevens, van enkele maanden tot een jaar of langer). Naast deze bestaande gegevens zijn er initiatieven in gang gezet om ook onderwijspersoneelsgegevens (afkomstig van DUO) beschikbaar te gaan stellen.
- Verder zijn er mogelijkheden om externe gegevens aan te leveren en te koppelen aan bestaande CBS-bestanden. Bestanden die reeds in het kader van beleidsinformatie verzameld worden, zoals CBS-bestanden die nog geen onderdeel uitmaken van de microdata of bijvoorbeeld gegevens van de Onderwijsinspectie (mits zij deze aan willen leveren), kunnen in principe zonder problemen gekoppeld worden aan de microdata.
- Bij andere typen bestanden, bijvoorbeeld externe enquêtes, is 'informed consent' nodig. Dat wil zeggen: toestemming van de personen waarop de data betrekking hebben, voor benutting van de gegevens buiten het directe doel van de enquête om. Om deze data te kunnen koppelen is het BSN de meest ideale koppelvariabele. BSN mag echter alleen gebruikt worden door overheidsorganisaties en niet-overheidsorganisaties indien dit wettelijk bepaald is. Een combinatie van geslacht, geboortedatum en adres is echter ook mogelijk om data koppelingen tot stand te brengen (maar bewerkelijker en foutgevoeliger).
- Een andere mogelijkheid is om de dataverzameling voor dergelijke enquêtes te laten verlopen via het CBS zelf. Dit vergemakkelijkt de koppeling van gegevens sterk, maar betekent wel dat het CBS ook een nadrukkelijke rol in het onderzoeksproces inneemt (o.a. ten aanzien van vragenlijstinhoud, steekproeftrekking et cetera). Bovendien betekent dataverzameling via het CBS dat deze data in principe ook voor vervolgonderzoek beschikbaar gesteld worden als microdata.
- Het CBS is positief over de ontwikkeling van een database voor (startende) leraren en lerarenopleidingen. Wel vindt het CBS het doel heel ambitieus, en adviseert het CBS dat er stapsgewijs te werk wordt gegaan bij de opbouw van een dergelijke database. Leveranciers van data zou bovendien iets bruikbaars 'teruggeven' moeten worden om het welslagen van de database te bevorderen.

Relevantie: groot

Haalbaarheid: groot



# 04

## DUO: REGISTERDATA ONDERWIJS

### 4.1 Inleiding

De Dienst Uitvoering Onderwijs (DUO) is een agentschap van het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW) dat studenten en onderwijsinstellingen financiert en informeert. Hoofdtaken van DUO zijn onder meer:<sup>15</sup>

- Bekostigen van onderwijsinstellingen;
- Verstrekken van studiefinanciering en tegemoetkoming schoolkosten;
- Verzorgen van proces van aanmelding, selectie en plaatsing hoger onderwijs;
- Verzamelen en beheren van onderwijsgegevens in diverse registraties;
- Verrijken van onderwijsgegevens tot informatieproducten.

Gezien deze taken is DUO de centrale organisatie als het gaat om onderwijsgegevens (zowel op het gebied van personeelsgegevens als schoolgegevens als studentgegevens) en derhalve een belangrijke gesprekspartner voor de database.

### 4.2 Visie op database

DUO gaat als uitvoeringsinstantie van OCW zelf niet over beleid en heeft zich te houden aan bestaande wetgeving, waaronder de privacyregelgeving. Gezien deze positie is DUO terughoudend in het formuleren van een visie over een database. Wel is DUO inhoudelijk en technisch zeer goed ingevoerd op het gebied van onderwijsbestanden en om die reden een belangrijke speler om te betrekken bij een database over (startende) leraren en lerarenopleidingen.

<sup>15</sup> Zie: <https://www.rijksoverheid.nl/ministeries/ministerie-van-onderwijs-cultuur-en-wetenschap/inhoud/organisatie/organogram/dienst-uitvoering-onderwijs-duo>

## 4.3 Aanwezige data

### *Personeels- en schoolgegevens*

De afdeling Informatieproducten van DUO beheert databestanden over scholen, studenten en personeel. De personeelsgegevens worden na bewerking (op geaggregeerd niveau) als open databestand via de website van DUO beschikbaar gesteld. Ook verschijnen personeelsgegevens van en over het primair en voortgezet onderwijs op de website scholenopdekaart.nl.

In het kader van de database zijn met name de personeelsgegevens over de leraren interessant. De andere gegevens (bijvoorbeeld schoolgegevens zoals regio, denominatie, schoolgrootte, jaarrekeningcijfers, et cetera) zijn uiteraard ook interessant, maar bieden enkel informatie over de schoolcontext. Scholen hebben een wettelijke verplichting om personeelsgegevens bij DUO aan te leveren. Dit is vastgelegd in zogenaamde 'Programma's van Eisen' (PvE's). In deze PvE's staat aangegeven dat scholen hun personeelsgegevens dienen aan te leveren in het kader van beleidsinformatie. De personeelsgegevens worden grotendeels geleverd door bureaus voor salarisadministratie van het onderwijs, maar ook scholen zelf leveren gegevens aan. Tevens ontvangt DUO lesgegevens uit het voortgezet onderwijs in het kader van de Integrale Personeelstelling Onderwijs (IPTO).

Op het niveau van de individuele leraar beschikt DUO over een viertal typen personeelsgegevens:<sup>16</sup>

- Persoonsgegevens: Burgerservicenummer (BSN)<sup>17</sup>, geslacht, leeftijd (maar géén NAW-gegevens);
- Baangegevens: hoofdfunctie, omvang aanstelling, inschaling. Voor leraren in het voortgezet onderwijs (vanuit IPTO) zijn de gegevens uitgebreider: voor hen is bekend wie welk vak doceert, hoeveel uur, et cetera. Voor het primair onderwijs en het middelbaar beroepsonderwijs zijn de baangegevens algemener van aard, bijvoorbeeld of men al dan niet aangesteld is als leraar (zij kunnen daarnaast andere taken/deelfuncties hebben);<sup>18</sup>
- Loongegevens: circa 40 componenten, zoals brutoloon, loonkosten, uitkeringen (zoals vakantie-uitkering, eindejaarsuitkering etc.), BAPO, enz.;

<sup>16</sup> Naast deze vier hoofdbestanden heeft DUO ook informatie over het gebruik van de Lerarenbeurs. Op basis van het verkennende gesprek kunnen nog geen uitspraken gedaan worden over de exacte inhoud en koppelbaarheid van deze gegevens. Deze informatie zou verkregen kunnen worden bij specialisten binnen DUO of eventueel bij het CPB dat al onderzoek aan de hand van de genoemde data heeft verricht.

<sup>17</sup> Het BSN wordt sinds een aantal jaren (circa 2013/2014) aan DUO geleverd. Sommige leveranciers hebben het BSN met terugwerkende kracht aangeleverd, waardoor ook voor eerdere jaren de persoonsidentificatie redelijk goed mogelijk is.

<sup>18</sup> DUO beschikt niet over soortgelijke gegevens van personeel dat werkzaam is in het hbo en wo. Wel ontvangt DUO een afslag van het Registratie Arbeidsrelaties Hoger Onderwijs (RAHO) bestand van de Vereniging Hogescholen. Dit is een (beperkt) bestand op individueel niveau zonder loon- en verlofgegevens. Voor het wo beschikt DUO alleen over geaggregeerde tabellen uit het Wetenschappelijk Onderwijs Personeels Informatie-systeem (WOPI) zoals beschikbaar gesteld op de website van de VSNU.



- Verlofgegevens: jaarlijkse gegevens over ziekteverzuim, zwangerschapsverlof, studieverlof enz.

Deze verschillende personeelsbestanden zijn onderling koppelbaar en voor 99% van de po- en vo-scholen worden de gegevens aangeleverd. Wel zijn er beperkingen. Zo kennen de vo-lesgegevens vanuit IPTO bijvoorbeeld een hoger dekkingspercentage (100%) dan de administratiegegevens die DUO ontvangt. Hierdoor zijn er soms voor een leraar geen formatiegegevens beschikbaar, maar wel lesgegevens uit IPTO. Een andere beperking is dat de DUO-formatiegegevens alleen personeel in loondienst omvatten. Leraren die niet in dienst zijn bij een school, maar wel lesgeven, komen niet in de formatiegegevens voor, terwijl ze wel in IPTO voor zouden kunnen komen. Het personeel dat niet in loondienst werkt vormt, op basis van de financiële jaargegevens zoals aangeleverd bij DUO, echter slechts een klein deel van het gehele onderwijspersoneel.<sup>19</sup>

### *Studentgegevens*

Naast personeels- en schoolgegevens beschikt DUO over diverse gegevens van studenten aan de verschillende lerarenopleidingen. Op studentniveau zijn (historische) inschrijvings- en diplomeringsgegevens beschikbaar. Deze zijn koppelbaar aan de DUO-formatiegegevens. Zo wordt voor IPTO de bevoegdheid van de docent (in eerste instantie) bepaald aan de hand van de behaalde diploma's. Naast administratieve gegevens, is in de studentbestanden het gemiddeld eindcijfer van de vooropleiding in het voortgezet onderwijs opgenomen. Voor studenten aan de lerarenopleidingen met een mbo- of ho-vooropleiding zijn overigens geen eindcijfers uit het mbo of ho beschikbaar, enkel het gemiddeld vo-eindcijfer is bekend.

De diplomagegevens van afgestudeerde leraren hebben betrekking op 'reguliere' opleidingen. Dat houdt in dat educatieve minoren of pedagogisch-didactische aantekeningen niet apart geregistreerd worden in de diplomabestanden. Ook leraren met buitenlandse diploma's komen niet voor in de diplomabestanden. In IPTO wordt sinds 2014 overigens wel gevraagd naar de benoemingsgrondslag als er geen gegevens uit het diplomaregister of eerdere IPTO-metingen bekend zijn. De mate van vulling is dan afhankelijk van de gegevensaanlevering door scholen.

## **4.4 Databenutting: ontsluiting en koppelmogelijkheden**

### *Ontsluiting*

Als het gaat om studentgegevens dan zijn er reeds veel data voorhanden die beschikbaar worden gesteld via het CBS (microdata). Zo zijn integrale gegevens over onderwijsinschrijvingen en diplomagegevens van het hoger onderwijs beschikbaar vanaf

<sup>19</sup> Zie ook: Leemans et al. (2017), *Personeel niet in loondienst (PNIL) in het po, vo en mbo*, Amsterdam: Regioplan/ROA.

2000. Omdat deze gegevens reeds ontsloten worden, verwijzen we voor meer informatie over microdata naar Hoofdstuk 3. De personeelsgegevens worden nog niet ontsloten via het CBS, maar er zijn momenteel wel initiatieven in gang gezet om deze gegevens via de CBS-microdata-omgeving ter beschikking te stellen.

#### *Koppelmogelijkheden externe bestanden*

DUO beschikt in de formatiebestanden over de volgende personalia: BSN, geslacht en leeftijd. DUO beschikt in principe niet over naam, adres en woonplaats (NAW)-gegevens van onderwijspersoneel, omdat het BSN niet zonder meer gebruikt mag worden om NAW-gegevens uit de Basisregistratie Personen (BRP) te halen. Voor onderwijspersoneel is er derhalve in principe geen koppeling aan de hand van NAW mogelijk. Om NAW aan de personeels-/leraargegevens te matchen zou een wijziging van de wet- en/of regelgeving nodig zijn. Vanuit beleids- en onderzoeksperspectief is dat mogelijk nog wel te verantwoorden, maar vanuit het oogpunt van privacybescherming en administratieve lasten is dat volgens DUO moeilijk te verantwoorden. Technisch gezien zou het matchen van NAW aan formatiegegevens (via BSN uit BRP) wel kunnen. Echter, gezien de initiatieven om personeelsgegevens (gedeeltelijk) beschikbaar te gaan stellen via het CBS, is het koppelen aan de hand van BSN de meest voor de hand liggende optie.

Bij studenten (ingeschrevenen/gediplomeerden) bestaat er overigens wel de mogelijkheid om aan de hand van NAW-gegevens te koppelen. Voor deze groep worden de NAW-gegevens namelijk door DUO zelf verzameld. DUO mag de eigen NAW-gegevens 'vergelijken' met de BRP-gegevens via het BSN. Een groot deel van de deelnemersgegevens wordt echter al in de CBS-omgeving beschikbaar gesteld als microdata.

## Conclusies

- DUO beschikt over een grote hoeveelheid onderwijsgegevens die voor een database van (startende) leraren en lerarenopleidingen zeer relevant zijn. Het gaat dan onder meer om loongegevens, verlofgegevens en baangegevens van leraren (type functie, aanstelling, inschaling). Met name voor het voortgezet onderwijs zijn de leraargegevens zeer rijk omdat door IPTO uitgebreide lesgegevens beschikbaar zijn, waardoor bekend is welk vak(ken) een leraar geeft, hoeveel uur, met welke bevoegdheid et cetera. Ook beschikt DUO over veel studentgegevens, zoals inschrijvings- en diplomeringgegevens. Tevens beschikt DUO over gegevens op schoolniveau, zoals jaarrekeningcijfers, gegevens over de leerlingpopulatie, denominatie, et cetera.
- De mogelijkheden om DUO-gegevens extern te koppelen lijken goed. Zo zijn er voor leraren BSN-gegevens bekend (maar in principe geen NAW-gegevens) en voor studenten BSN- én NAW-gegevens. Voor de ontsluiting van de gegevens, zeker als het gaat om de mogelijkheden om de gegevens vervolgens ook aan andere data te koppelen, lijkt een route via het CBS het meest voor de hand liggend (gezien de koppeling op BSN). Op dit moment worden diverse studentbestanden reeds beschikbaar gesteld als CBS-microdata, waardoor deze daarmee ook aan andere CBS-data gekoppeld kunnen worden. De onderwijspersoneelsgegevens worden op dit moment nog niet ontsloten via het CBS, maar er zijn hier momenteel wel initiatieven voor. Het ontsluiten van de DUO-gegevens is weliswaar een uitdaging, maar er worden dus al stappen gezet in die richting.

Relevantie: groot

Haalbaarheid: groot



# 05

## INSPECTIE VAN HET ONDERWIJS: KWALITEITSINDICATOREN

### 5.1 Inleiding

De Inspectie van het Onderwijs (ofwel Onderwijsinspectie) is een controleorgaan van het Ministerie van Onderwijs, Cultuur en Wetenschap dat toezicht houdt op het Nederlandse onderwijs. Zij bekijkt of er aanwijzingen zijn dat scholen onvoldoende kwaliteit leveren. Op basis van een risicoanalyse en eventueel extra onderzoek wordt gekeken of een school extra toezicht nodig heeft. Daarnaast houdt de Inspectie de (financiële) gezondheid van besturen in de gaten en houdt ze toezicht op het stelsel van hoger onderwijs. De Inspectie voert onder meer schoolinspecties, documentanalyses, peilingsonderzoeken en thema-onderzoeken uit. Gesignaleerde trends en ontwikkelingen in het onderwijs worden jaarlijks gepubliceerd in het rapport *'De Staat van het Onderwijs'*.

### 5.2 Visie op database

Volgens de Inspectie zijn de diverse, op dit moment voorradige (onderzoeks)gegevens over leraren erg gefragmenteerd en wordt het zeer nuttig geacht om deze meer in relatie tot elkaar te gebruiken. Vanuit onderzoeksperspectief vormt een structuur waarin verschillende data aan elkaar gekoppeld kunnen worden een 'ideaalbeeld'. Voor de Onderwijsinspectie zou het benutten van gekoppelde data onder meer nuttig zijn voor de jaarlijkse rapportage *'De Staat van het Onderwijs'*. Hiervoor put de Onderwijsinspectie reeds uit vele bronnen die nu op verschillende plekken opgevraagd moeten worden. Een meer gecentraliseerde plek voor lerarendata zou het proces vergemakkelijken en zou het bovendien eenvoudiger maken om data te kunnen 'hergebruiken'. De Onderwijsinspectie geeft verder aan momenteel een zogenaamde 'Stelselmonitor' te ontwikkelen, een soort dashboard waarin beleidsrelevante, (vooral) structurele onderwijsdata geïntegreerd worden en waarmee een 'vinger aan de pols' van het onderwijs gehouden kan worden. Ook voor deze monitor zou het nuttig zijn als gebruik gemaakt kan worden van gekoppelde, gestructureerde leraargegevens. Daarnaast zou een database bruikbaar kunnen zijn voor het toezicht op individuele scholen en besturen.

Waarborgen ten aanzien van betrouwbaarheid en de eisen aan de kwaliteit en volledigheid van de gegevens zouden dan echter veel strenger zijn.

De Onderwijsinspectie geeft aan dat er grote winst behaald kan worden door registerdata beter te benutten. De Onderwijsinspectie heeft al ervaring met het afzonderlijk analyseren van DUO- en CBS-data, maar juist de combinatie van deze databronnen kan veel nieuwe inzichten opleveren, bijvoorbeeld als het gaat om (onderwijs-)arbeidsmarktstromen. Ook kan het benutten van registerdata bijdragen aan een verminderde bevragingslast van onderwijsinstellingen door onderzoeksbureaus, bijvoorbeeld doordat er minder achtergrondgegevens opgevraagd hoeven te worden, maar ook omdat aan de hand van gegevens uit de database een gerichtere aanschrijving (bijvoorbeeld specifieke groepen) mogelijk zou zijn.

### *Praktische overwegingen*

Hoewel een database op diverse vlakken zinvol wordt geacht, plaatst de Onderwijsinspectie met nadruk enkele kanttekeningen. Zo dient de privacybescherming van zowel scholen als leraren absoluut gewaarborgd te worden. Dit betekent niet alleen dat gegevens geanonimiseerd dienen te worden, maar ook dat er beperkingen gesteld moeten worden aan de onthullingsmogelijkheden van de data. Door gegevens te koppelen kunnen er immers al snel kleine groepen ontstaan waardoor er een risico op persoonsonthulling ontstaat, iets wat te allen tijde voorkomen moet worden. Wanneer het beheer van de database bij het CBS zou komen te liggen, dan lijkt de privacybescherming volgens de Onderwijsinspectie in principe wel gewaarborgd te zijn gezien hun strenge voorwaarden rond het gebruik van microdata.

## **5.3 Aanwezige data**

### *Lesobservaties*

De Onderwijsinspectie voert schoolinspecties uit waarbij lesobservaties gehouden kunnen worden. Dit is echter niet per definitie het geval. Uitsluitend bij bezoeken in het kader van 'De Staat van het Onderwijs' houdt de Onderwijsinspectie lesobservaties die verzameld worden om op schoolniveau analyses te doen. De lesobservaties worden gehouden bij drie leerjaren (groep 3, 5 en 7 in het basisonderwijs). De lesobservatiedata zijn derhalve verre van integraal, ook omdat scholen (in principe) steekproefsgewijs worden bezocht en niet alle leraren op een school worden geobserveerd. De bevindingen uit de lesobservaties worden wel door de Onderwijsinspectie bewaard, maar zijn niet tot individuele personen herleidbaar. Er worden bovendien slechts enkele basale achtergrondkenmerken van de leraren verzameld. Wel is er altijd een identifica-

tienummer van de BRIN-vestiging<sup>20</sup>, waardoor de data wel op schoolniveau te benutten zijn. De Onderwijsinspectie beschouwt deze data als privacygevoelig, omdat er bij kleine scholen alsnog redelijk eenvoudig achterhaald kan worden welke leraren het betreft.

De Onderwijsinspectie benadrukt dat zij gegevens verzamelt om de kwaliteit van scholen en het onderwijs(stelsel) in bredere zin in kaart te kunnen brengen, niet om de kwaliteit van individuele leraren te beoordelen. De onderzoeksinstrumenten van de Onderwijsinspectie zijn daar ook niet voor bedoeld. Het is de verantwoordelijkheid van de individuele schoolbesturen om de kwaliteit van individuele leraren te waarborgen.

### *Vragenlijstgegevens*

Naast lesobservatiegegevens beschikt de Onderwijsinspectie ook over leraargegevens vanuit vragenlijsten. Dergelijke vragenlijsten spelen meestal in op een specifieke onderzoeksvraag en zijn daarom vaak incidenteel van aard. Het betreft bijvoorbeeld thema-onderzoeken naar professionalisering, de begeleiding van beginnende leraren en percepties op toezicht. Een concreet voorbeeld hiervan is het vierdelige onderzoek 'Beginnende leraren kijken terug' uit 2015.<sup>21</sup> In deze onderzoeksreeks is nagegaan in hoeverre leraren die recent zijn afgestudeerd aan een lerarenopleiding, zich door hun opleiding voldoende voorbereid voelen op het leraarschap. De focus van dit onderzoek is met name gericht op het kennisniveau en de pedagogisch-didactische vaardigheden van beginnende leraren. Naast opvattingen van leraren over hun vaardigheden en kennisniveau, gaat de onderzoeksreeks ook in op de opvattingen van schoolleiders hierover. Onderzoeksgegevens zoals die in 'Beginnende leraren kijken terug' zijn meestal grootschalig van aard (dat wil zeggen breed uitgezet in het veld). Echter: net als de lesobservatiedata, zijn ook deze gegevens niet tot individuen herleidbaar, enkel tot scholen.

## **5.4 Benutting: ontsluiting en koppelmogelijkheden**

### *Ontsluiting en koppelmogelijkheden*

De hiervoor behandelde gegevens (lesobservaties en vragenlijstdata) zijn zoals gezegd niet tot individuen te herleiden. Wel kunnen ze op schoolniveau benut worden om zicht te krijgen op de schoolcontext waarbinnen leraren werken. De koppelmogelijkheden liggen op het niveau van de BRIN-vestiging. De hiervoor benoemde gegevens worden niet actief openbaar gemaakt door de Onderwijsinspectie. Wel staat zij er voor open om dergelijke gegevens te ontsluiten voor onderzoek, mits aan bepaalde voorwaarden voldaan worden. Het gaat dan om afspraken over anonimisering en onthulling (van scholen) en over aan wie toegang tot de data wordt verleend.

<sup>20</sup> De Basisregistratie Instellingen (BRIN) is het landelijke register van onderwijsinstellingen. Elke instelling wordt hierin geïdentificeerd aan de hand van een unieke code van vier alfanumerieke karakters. Deze code wordt aangevuld met twee cijfers die de specifieke vestiging aanduiden.

<sup>21</sup> De vier rapportages gaan respectievelijk over afgestudeerden van: de pabo, de tweedegraads lerarenopleidingen, de universitaire lerarenopleidingen, en de eerstegraads lerarenopleidingen op het hbo.

## Conclusies

- Een database van (startende) leraren en lerarenopleidingen wordt als potentieel nuttig beschouwd door de Inspectie van het Onderwijs. De Onderwijsinspectie geeft aan dat de bestaande lerarendata momenteel erg gefragmenteerd zijn. Een database zou het makkelijker maken om gegevens te analyseren en te hergebruiken voor onder meer de jaarlijkse publicatie 'De Staat van het Onderwijs'. Ook zou deze op den duur benut kunnen worden rondom de inspectie van individuele scholen, hoewel de eisen aan de data (zoals volledigheid en kwaliteit) dan zeer hoog moeten zijn.
- Een belangrijke kanttekening bij een database is de wijze waarop omgegaan wordt met persoonsgegevens. Een database zou niet alleen anoniem moeten zijn, maar ook voorwaarden moeten stellen aan het onthullingsrisico wanneer gegevens gekoppeld worden. Een CBS-model waarbij strenge voorwaarden aan de onderzoeksoutput worden gesteld zou hiervoor een oplossing zijn.
- De Onderwijsinspectie verzamelt veel gegevens om toezicht te houden op de onderwijskwaliteit in Nederland. Op het niveau van individuele leraren beschikt de Inspectie over gegevens afkomstig uit lesobservaties en vragenlijstonderzoeken, maar deze gegevens zijn niet tot individuen herleidbaar. Dit is ook niet het doel van de Onderwijsinspectie: zij controleert niet de kwaliteit van individuele leraren, maar houdt toezicht op de kwaliteit van scholen en het onderwijsstelsel in het algemeen. De gegevens zijn derhalve enkel te benutten op het schoolgeaggregeerd niveau (BRIN-vestiging) en dus vooral te benutten als contextuele kenmerken. De Inspectie is bereid om, mits goede afspraken gemaakt worden over anonimisering, onthulling en datatoegang, gegevens beschikbaar te stellen voor onderzoek.

Relevantie: Redelijk  
Haalbaarheid: Groot



# 06

## LERARENOPLEIDINGEN: STUDENTGEGEVENS

### 6.1 Inleiding

In Nederland kunnen in principe drie typen onderwijsbevoegdheden onderscheiden worden: een bevoegdheid voor het primair onderwijs, een bevoegdheid voor het tweedegraads domein (onderbouw havo/vwo en het gehele vmbo; mbo en praktijk-onderwijs) en een bevoegdheid voor het eerstegraads domein (het gehele tweedegraads domein plus de bovenbouw havo/vwo). De bevoegdheid is gerelateerd aan een bepaald vak binnen het type onderwijs waarin les gegeven wordt. De bevoegdheid voor het primair onderwijs kan gehaald worden via de pabo (hbo-bachelor 'leraar basisonderwijs') of de academische pabo (combinatie van de universitaire bachelor 'onderwijskunde' en de hbo-bachelor 'leraar basisonderwijs' [pabo]). De tweedegraads bevoegdheid kan gehaald worden via een tweedegraads lerarenopleiding (hbo-bachelor) aan een hogeschool.<sup>22</sup> De eerstegraads bevoegdheid kan gehaald worden via een eerstegraads lerarenopleiding (master) op een hogeschool of universiteit.

Aangezien er drie typen bevoegdheden onderscheiden kunnen worden, kunnen er op basaal niveau dus ook drie typen lerarenopleidingen onderscheiden worden. Om een breed beeld te krijgen van de diverse opvattingen die binnen lerarenopleidingen kunnen bestaan over een 'database' en de typen studentgegevens die binnen opleidingen verzameld worden, is binnen dit verkennend onderzoek daarom met alle drie typen opleidingen gesproken. Bij het lezen van dit hoofdstuk dient de lezer zich ervan te vergewissen dat er gesprekken zijn geweest met een aantal representanten van specifieke lerarenopleidingen c.q. van de overlegorganen en/of koepelorganisaties van de opleidingen (LOBO, ADEF en ICL).<sup>23</sup> De bevindingen zijn derhalve wel *indicatief* voor de (variatie aan) opvattingen van de lerarenopleidingen over een database, maar mogen niet zonder meer als *representatief* worden beschouwd.

<sup>22</sup> Daarnaast zijn er enkele hbo-opleidingen die leiden tot een zogenaamde 'ongegradeerde' bevoegdheid. Het betreft de (hbo-bachelor) lerarenopleidingen voor de 'kunstvakken' en de lerarenopleiding 'lichamelijke opvoeding'. Afgestudeerden aan deze opleidingen kunnen voor het betreffende vak in het gehele voortgezet onderwijs lesgeven, evenals in het basisonderwijs en het beroepsonderwijs.

<sup>23</sup> LOBO: Landelijk Overleg Lerarenopleiding Basisonderwijs (pabo-opleidingen), ADEF: Algemeen Directeurenoverleg Educatieve Faculteiten (hbo-opleidingen voor leraren vo), ICL: Interuniversitaire Commissie Lerarenopleidingen (universitaire lerarenopleidingen).

## 6.2 Visie op database

### *Scepsis over toegevoegde waarde*

Over het nut van een database bestaat de nodige scepsis. Dit lijkt met name het geval onder de geïnterviewden van de hbo-lerarenopleidingen (pabo en tweedegraads lerarenopleidingen). Zo is het diverse geïnterviewden onduidelijk waarom een database nodig is, zeker wanneer de database nadrukkelijk in verband wordt gebracht met 'kwaliteit'. De kwaliteit van opleidingen, en daarmee in zekere zin van studenten die de opleiding volgen, is volgens diverse geïnterviewden immers al geborgd in het accreditatiestelsel. Opleidingen die onvoldoende kwaliteit bieden, worden immers niet geaccrediteerd. Dit stelsel wordt als erg belastend<sup>24</sup> ervaren en zou voldoende moeten zijn om (opleidings)kwaliteit in kaart te brengen.

Diverse geïnterviewden geven aan dat onderzoek naar de kwaliteit van leraren als risico heeft dat lerarenopleidingen er op 'afgerekend' worden. Lerarenopleidingen kunnen volgens hen het risico lopen dat zij aangewezen worden als 'schuldige' als zaken niet goed lopen. Lerarenopleidingen liggen sowieso al sterk onder het vergrootglas, zo wordt gesteld. Er worden grote vraagtekens geplaatst bij het bestuderen van eventuele relaties tussen de gevolgde lerarenopleiding en de kwaliteit van leraren (zeker als het gaat om leraren in latere fasen van hun loopbaan). Zo wordt betoogd dat de lerarenopleiding natuurlijk een zeer belangrijk onderdeel uitmaakt van de ontwikkeling van de student en de (startende) leraar, maar dat er talloze aspecten samenhangen met de kwaliteit van leraren. Er valt dan te denken aan hun persoonlijke achtergrond, de school waar zij werken, enzovoorts. Bovendien ontwikkelen studenten en leraren zich op allerlei manieren en is een goede student nog niet per definitie een goede leraar en andersom. Dit maakt het zeer complex om aan te geven wat kwaliteit bepaalt, nog afgezien van de vraag wat kwaliteit precies is. Zo wordt er gesteld: "Het is vrijwel onmogelijk om met alles rekening te kunnen houden en dus zal er snel gewezen worden naar lerarenopleidingen" en "De 'kwaliteit' van leraren komt tot uiting in vele aspecten en is niet met een paar variabelen te vatten". Een geïnterviewde noemt de kwetsbaarheid van opleidingen door onderzoek (zeker als hierover niet anoniem gerapporteerd zou worden) als grote uitdaging voor de database.

Voor enkele geïnterviewden is de meerwaarde van een database vanuit het oogpunt van 'kennis' niet evident, aangezien er al voorzien wordt in hun kennisbehoefte. Zo wordt de voortgang van studenten binnen de opleidingen al op allerlei manieren gemonitord. Ook tussen de instellingen onderling wordt hier aandacht aan besteed. De (directies van de) pabo's komen maandelijks bijeen in het LOBO om het een en ander te bespreken. In hun overleg komen onder andere actuele beleidsontwikkelingen aan bod en worden de scores op de landelijke kennistoetsen (en verschillen daarin tussen instellingen) globaal

<sup>24</sup> Iets wat overigens onderschreven wordt door het Ministerie van OCW. Minister Bussemaker kondigde in haar kamerbrief van 18 februari 2016 (betreft: 'Accreditatie op maat III') daarom aan om te gaan starten met de pilot instellingsaccreditatie met lichte opleidingsaccreditatie.

besproken. Ook als het gaat om informatie over de arbeidsmarktpositie van afgestudeerden en/of over de aansluiting tussen opleiding en arbeidsmarkt, wordt hier volgens veel geïnterviewden al voor een belangrijk deel in voorzien vanuit contacten met alumni en vanuit bestaand onderzoek (ofwel eigen onderzoek, ofwel landelijk onderzoek zoals de HBO-Monitor). Verder kunnen lerarenopleidingen er altijd voor kiezen om zelf onderzoek te (laten) verrichten als er een specifieke kennisvraag is. In algemene zin wordt gesteld dat de lerarenopleidingen verschillende indicatoren gebaseerd op student- en alumnigegevens al veelvuldig met elkaar uitwisselen en bespreken. Kennis wordt dus al volop gedeeld en benut door de opleidingen.

### *Potentiële meerwaarde*

Hoewel er de nodige scepsis bestaat over de meerwaarde van een database, worden er door enkele geïnterviewden ook voordelen genoemd. In algemene zin kan een dergelijke database bijdragen aan het verkrijgen van een breder beeld over (de kwaliteit van) leraren en de relatie met lerarenopleidingen. Zo kan beter in beeld worden gebracht in hoeverre verschillen tussen (de kwaliteit of het curriculum van) lerarenopleidingen op latere momenten in de loopbaan gevolgen hebben. Een geïnterviewde stelt: *“Lerarenopleidingen hebben een globaal beeld van wat ‘goede didactiek’ is en hoe je studenten daarin bekwaamt. Welke opleidingsaspecten, bijvoorbeeld meer of minder aandacht voor vakdidactiek, in de praktijk en op langere termijn beter of minder goed werken is echter nog niet goed in beeld. Een database zou daar bij kunnen helpen.”*

Meer specifiek wordt een verbeterd zicht op de competentieontwikkeling na afronding van de studie genoemd als potentiële opbrengst van een database. Er is immers wel veel bekend over de arbeidsmarktontwikkelingen van afgestudeerden vanuit alumnistudies, maar nauwelijks over de inhoudelijke, kwalitatieve ontwikkeling van afgestudeerden, zo wordt aangegeven. Ook kan een database bijdragen aan kennis over uitval van studenten en leraren. Onderzoeksvragen die hierbij centraal staan zijn bijvoorbeeld: wat typeert de uitvallers, wat zijn risicofactoren voor uitval en hoe kan de opleiding daarop inspelen?

### *Kennisbehoefte in relatie tot de database*

Hoewel de opleidingen aangeven al veel informatie te hebben over hun studenten en afgestudeerden, zijn er dus wel degelijk onderwerpen waarop kennisvermeerdering gewenst is. Zo zijn de opleidingen geïnteresseerd in de mate waarin afgestudeerden goed voorbereid worden op het uitoefenen van het beroep. Het gaat daarbij niet alleen over de vraag of afgestudeerden zich ‘staande’ kunnen houden in het onderwijs, maar ook over de mate waarin zij ‘geoutilleerd’ zijn om een daadwerkelijke bijdrage aan onderwijsverbetering te kunnen leveren. Door de pabo’s wordt verder uitgesproken dat er behoefte bestaat aan meer kennis over het thema ‘mannen in het onderwijs’. Zo zou men graag succesfactoren onderscheiden om meer mannen te kunnen werven en te behouden voor de lerarenopleidingen en het onderwijs in het algemeen.

Vooral de representanten van de universitaire lerarenopleidingen zien een meerwaarde van een database om vraagstukken rondom de kwaliteit(sontwikkeling) en arbeidsmarktpositie van leraren met betere data in kaart te brengen. Andere geïnterviewden geven echter aan dat zij wel kennisbehoeften hebben, maar dat zij niet zien hoe een database daarin zou kunnen voorzien. Zo wordt er vanuit de pabo's aangegeven dat er geen behoefte is aan een brede database met generieke basisgegevens over leraren, maar dat juist de contextspecifieke, 'gefragmenteerde' onderzoeken naar specifieke thema's nuttige inzichten op kunnen leveren. Juist deze specifieke studies in concrete, bekende situaties (zoals in de eigen regio/omgeving) worden door de representanten van de pabo hoog gewaardeerd. Zo worden lokale interventiestudies als voorbeeld genoemd van praktisch toepasbaar, nuttig onderzoek.

### *Praktische overwegingen bij database*

Naast de nodige twijfels over het nut en de mogelijkheid om gedegen onderzoek uit te voeren naar kwaliteit, vragen diverse geïnterviewden zich nadrukkelijk af wie de eigenaar van een database zou zijn en wie daar toegang tot heeft. De geïnterviewden geven aan dat het in ieder geval voor de hand ligt dat de lerarenopleidingen daar zelf toegang toe hebben. Sommige geïnterviewden gaan hier nog wat verder in, en vinden het belangrijk – of zelfs essentieel – dat zij zelf de eigenaar zijn. Een geïnterviewde: *“Als de lerarenopleidingen de eigenaar van de database zijn, dan kan het helpen om specifieke kennisvragen te beantwoorden die op een gegeven moment rijzen. Als lerarenopleidingen niet de eigenaar zijn, dan ontstaat het gevaar dat er met de gegevens aan de haal wordt gegaan en de vinger al snel naar de lerarenopleiding wijst.”*

Een breed gedeelde zorg heeft betrekking op de privacygevoeligheid van een database, zeker als het gaat om het benutten van gegevens vanuit lerarenopleidingen. De grote vraag voor de opleidingen is: (op wat voor manier) zijn mensen herkenbaar? Maar ook de herkenbaarheid van de opleidingen zelf is (zoals eerder genoemd) een issue. Het borgen van de privacy van de studenten is essentieel en een absolute voorwaarde voor eventuele benutting van gegevens van lerarenopleidingen.

### **6.3 Aanwezige data**

De lerarenopleidingen beschikken over de nodige gegevens over de intrede, voortgang en uitstroom van studenten. Wat betreft intredegegevens gaat het bijvoorbeeld om jaarlijkse toelatingsexamens en instaptoetsen voor Nederlandse taal en rekenen/wiskunde. Daarnaast beschikken de opleidingen natuurlijk over voortgangsgegevens tijdens de studie. Het gaat dan onder meer om kennistoetsen en voortgangsgegevens over de ontwikkeling van competenties zoals algemene didactische vaardigheden, pedagogische vaardigheden en coaching. Uitstroomgegevens waarover opleidingen beschikken, zijn bijvoorbeeld het wel of niet behalen van een diploma binnen de nominale studieduur. Daarnaast beschikken de opleidingen over alumniegegevens die worden verza-

meld via de algemene alumni-onderzoeken, maar ook via persoonlijke contacten met afgestudeerden.

Bij het benutten van studentgegevens voor onderzoek worden diverse kanttekeningen gezet. Meermaals wordt aangegeven dat de gegevens tussen lerarenopleidingen niet één op één vergelijkbaar zijn, bijvoorbeeld omdat opleidingen en instellingen over verschillende systemen beschikken die niet zomaar te harmoniseren zijn. Maar vooral ook inhoudelijk zijn er verschillen: zo zijn er verschillen tussen opleidingen en instellingen in de aanpak en de curricula, alsmede in de toetsen en toetswijzen (bijvoorbeeld via Rubrics evaluaties<sup>25</sup> of schriftelijke toetsen), eindtermen enzovoorts.

Met name het vergelijken van studentcompetenties wordt moeilijk geacht. Zo wordt er door enkele representanten van de tweedegraads lerarenopleidingen aangegeven dat hier geen vergelijkbare data over zijn. Een vertegenwoordiger van de universitaire lerarenopleidingen geeft overigens aan dat er tussen de wo-lerarenopleidingen waarschijnlijk wél een behoorlijke mate van vergelijkbaarheid bestaat in competentiebeoordelingen (het leraar-competentieprofiel is ulo-breed), hoewel de toetswijzen niet gelijk zijn. Verder geven enkele geïnterviewden aan dat de leerwegen en vakken die studenten volgen zeer gepersonaliseerd kunnen zijn, hetgeen 'gestandaardiseerde' gegevens in een database bemoeilijkt.

## 6.4 Databenutting: ontsluiting en koppelmogelijkheden

### *Ontsluiting*

Het eventuele ontsluiten van studentgegevens vanuit de lerarenopleidingen zal aan zeer strikte voorwaarden moeten voldoen. Naast bedenkingen over de toegevoegde waarde van een database en de ingeschatte methodologische problemen van het onderzoek op een dergelijke database, is er zoals gezegd grote zorg over de privacygevoeligheid van de gegevens. Mogelijke ontsluiting van gegevens via het CBS als een 'vertrouwde derde partij' wordt daarbij overigens niet per definitie als borging van de privacy gezien. De levering van studentgegevens zonder een absolute waarborg van de privacybescherming is dan ook uitgesloten. Veel geïnterviewden geven daarbij aan dat zij zonder expliciete toestemming van de studenten geen gegevens aan zouden willen en mogen leveren.<sup>26</sup> Een geïnterviewde oppert een zogenaamd 'opt in / opt out'-

<sup>25</sup> Rubrics zijn analytische beoordelingsschalen om een product of vaardigheid te beoordelen op kwaliteit en om de manier van werken te beoordelen om tot een product te komen (zie: <http://www.slo.nl/voortgezet/vmbo/themas/Rubrics/>).

<sup>26</sup> Uit ervaringen van het Nationaal Cohortonderzoek Onderwijs (NCO) lijkt het '*niet mogen leveren*' overigens niet aan de orde. In het NCO worden leerlinggegevens in de beveiligde omgeving van het CBS (die daar als 'entrusted third party' optreedt) opgeslagen en bewerkt. Voortgangsgegevens over scholieren, afkomstig uit het leerlingvolgsysteem, kunnen benut worden als schoolbesturen daarvoor toestemming geven. Wel wordt aanbevolen om de (ouders van) scholieren hierover te informeren. Voor 'bijzondere persoonsgegevens' waarover instellingen beschikken (denk bijvoorbeeld aan gezondheidsgegevens van studenten) is overigens wél actieve toestemming nodig.

principe, waarbij studenten expliciet moeten aangeven of ze hun gegevens beschikbaar willen stellen en op ieder moment kunnen aangeven dat zij de gegevenslevering willen beëindigen en/of hun gegevens verwijderd willen zien. Een belemmering bij de benutting van studentgegevens kan de 'houdbaarheid' van deze data zijn. Zo geven geïnterviewden aan dat studentgegevens niet oneindig bewaard worden, er worden bewaartermijnen genoemd van vijf tot zeven jaar. Het langdurig bewaren van dergelijke gegevens zou "een hele goede reden" vereisen. De bewaartermijn (of het hanteren van een bewaartermijn in algemene zin) lijkt te verschillen van instelling tot instelling.

Een andere beperking bij de ontsluiting van studentgegevens heeft betrekking op het extraheren van de gegevens uit de instellings- en opleidingsssystemen. Enkele geïnterviewden geven aan dat het niet zo eenvoudig is om studentgegevens gestructureerd uit te draaien c.q. aan te leveren voor onderzoeksdoeleinden. "Het is geen druk op de knop, in de praktijk kan het uitermate ingewikkeld zijn." Het idee dat het éénmalig een forse investering betreft en in het vervolg min of meer automatisch kan gebeuren, wordt bovendien bestreden. Het opvragen van data van opleidingen zou dus volgens de opleidingen een forse belasting kunnen opleveren.

#### *Koppelmogelijkheden externe bestanden*

De koppelbaarheid van studentgegevens met andere databronnen lijkt in principe niet problematisch. Zo zijn voortgangsgegevens zoals toetsresultaten te herleiden tot studentnummers. Met deze studentnummers zijn verdere individuele gegevens te achterhalen zoals bekend bij de onderwijsinstelling, bijvoorbeeld NAW-gegevens. Bovendien wordt binnen het onderwijs het burgerservicenummer (BSN) gebruikt, daar ook wel het persoonsgebonden nummer (PGN) of 'onderwijsnummer' genoemd.<sup>27</sup> Iedereen die onderwijs volgt dat door de overheid wordt betaald, heeft te maken met het PGN. Instellingen zijn verplicht het PGN te gebruiken in hun administratie. Externe datakoppeling aan de hand van het BSN/PGN is technisch gezien dus eenvoudig te bewerkstelligen.

We sluiten dit hoofdstuk over de lerarenopleidingen af met twee korte 'case studies' waarin we de studentadministratie van twee instellingen bespreken (HAN en De Nieuwste Pabo van Zuyd Hogeschool). Het betreffen hier twee praktijkvoorbeelden van hoe instellingen studentgegevens administreren (dat wil zeggen het type gegevens, de systematiek) en in hoeverre deze gegevens ontsluitbaar zijn. Deze case studies zijn illustratief en geven geen representatief beeld van hoe studentgegevens in het hoger onderwijs opgeslagen worden.

<sup>27</sup> Zie ook: <https://www.rijksoverheid.nl/onderwerpen/persoonsgegevens/burgerservicenummer-bsn/bsn-in-het-onderwijs>.

## 6.5 Case study 1: Hogeschool van Arnhem en Nijmegen<sup>28</sup>

### *Context*

De Faculteit Educatie van de Hogeschool van Arnhem en Nijmegen (HAN) biedt praktijkgerichte bachelor- en masteropleidingen aan op het gebied van onderwijs en opleiden. De bacheloropleidingen worden verzorgd binnen de uitvoerende instituten Instituut voor Leraar en School (ILS) en het Instituut HAN Pabo. Binnen het ILS worden de tweedegraads lerarenopleidingen voor het voortgezet onderwijs aangeboden. Het gaat om de lerarenopleidingen Aardrijkskunde, Economie, Geschiedenis, Nederlands, Duits, Frans, Engels, Biologie, Natuurkunde, Scheikunde, Wiskunde en Pedagogiek. Tevens wordt binnen de ILS de bachelor Opleidingskunde aangeboden. Het Instituut HAN Pabo verzorgt de lerarenopleiding basisonderwijs (pabo) en biedt de Academische Lerarenopleiding Primair Onderwijs (ALPO)<sup>29</sup> aan. Verder worden binnen het interfacultair instituut 'HAN Masterprogramma's' (HMP) de volgende onderwijsmasters aangeboden: Begeleidingskunde, Leraar Economie, Leraar Engels, Leraar Nederlands, Leraar Wiskunde en Pedagogiek. Jaarlijks kennen de diverse opleidingen binnen het pedagogisch domein circa 1.000 instromers. Het gaat daarbij in iets minder dan de helft van de gevallen om instromers in de verschillende tweedegraads lerarenopleidingen. Ruwweg drie tiende volgt de opleiding tot leraar basisonderwijs en ongeveer één op de zeven volgt de opleiding tot leraar lichamenlijk opleiding. De overige instromers (circa een tiende) volgen de bachelor Opleidingskunde of één van de masters.<sup>30</sup>

### *Studentadministratie en type gegevens*

De HAN maakt gebruik van het studentinformatiesysteem ALLURIS. Hoewel de geïnterviewde vermoedt dat dit systeem niet breed gebruikt wordt binnen het hbo, lijkt een eerste online inventarisatie dit wat te nuanceren.<sup>31</sup> Het is de geïnterviewde niet bekend in hoeverre de studentgegevens tussen de verschillende hogescholen op een (datatechnisch) overeenkomstige manier worden geadmistreerd. Wel wordt ALLURIS in ieder geval binnen de HAN instellingsbreed benut. Alle onderliggende faculteiten en opleidingen maken er gebruik van (in ieder geval voor het bekostigd onderwijs binnen de HAN).

<sup>28</sup> Informatie afkomstig van gesprek met deskundige aangevuld met gegevens van [www.han.nl](http://www.han.nl) en [www.vereniginghogescholen.nl](http://www.vereniginghogescholen.nl).

<sup>29</sup> Studenten aan de ALPO volgen de helft van de vakken aan de Radboud Universiteit bij Pedagogische Wetenschappen en Onderwijskunde en de andere helft bij de HAN Pabo Nijmegen. Na afronding van de ALPO beschikken zij over twee diploma's.

<sup>30</sup> Aantallen afkomstig van de kennisbank van de Vereniging Hogescholen. Zie: <http://www.vereniginghogescholen.nl/kennisbank/feiten-en-cijfers/artikelen/feiten-en-cijfers-qlikview>

<sup>31</sup> Zo treffen we via Google ALLURIS-portals aan voor de Aeres Hogeschool, Amsterdamse Hogeschool voor de Kunsten, HAS Hogeschool, Hogeschool van Arnhem en Nijmegen, IPABO, Thomas More Hogeschool.

In ALLURIS wordt de voortgang van studenten bijgehouden. HAN-studenten halen (net als bij andere instellingen) studiepunten door het succesvol afronden van onderwijseenheden. Deze onderwijseenheden kunnen echter bestaan uit meerdere onderliggende deeltentamens. Zo kan het bijvoorbeeld zijn dat studenten meerdere toetsen moeten behalen om studiepunten te halen. Dit betekent dat het voorkomt dat studenten soms géén (nul) studiepunten behalen, omdat zij bijvoorbeeld slechts 4 van de 5 onderliggende toetsen hebben behaald. Alle voortgangsgegevens worden echter opgeslagen in het systeem. Zowel de toegekende studiepunten als de onderliggende beoordelingen van de verschillende onderdelen zijn in ALLURIS herkenbaar. De onderwijsdelen (en onderliggende onderdelen) zijn herkenbaar via een specifieke benaming zoals de toetscode. Aan de hand van die benaming is overigens niet direct te zien op wat voor manier beoordeeld is. Om te achterhalen of het bijvoorbeeld gaat om een kennistoets, een vaardigheidentoets of een andere beoordeling zou informatie vergaard moeten worden bij de specifieke opleiding.

Naast gegevens uit het studievolsysteem beschikt de HAN over een studentinschrijvingsbestand met achterliggende persoonsgegevens. De gegevens uit ALLURIS en het inschrijvingsbestand zijn onderling koppelbaar. Het is dus in theorie mogelijk om voortgangsgegevens via de persoonsgegevens uit het inschrijvingsbestand aan externe datasets te koppelen. Naast inschrijvings- en studentgegevens beschikt de HAN over gegevens uit de Nationale Studenten Enquête<sup>32</sup> (NSE).

### *Exporteren van gegevens*

Het studentvolgsysteem is (vanzelfsprekend) niet ontwikkeld om onderzoeksbestanden mee te construeren. Het uitdraaien van gegevens voor onderzoeksdoeleinden is dan ook bepaald 'geen druk op de knop'. Mocht dit in de toekomst benodigd zijn dan vergt dit vermoedelijk flinke inspanningen. Het systeem maakt het wel mogelijk om snel gegevens in te zien van studenten, hetgeen bijvoorbeeld handig is voor studiebegeleiders, maar de gegevens zijn niet eenvoudig te exporteren. Er worden momenteel wel acties ondernomen binnen de HAN om gegevens uit het systeem te gaan ontsluiten voor de eigen informatievoorziening. Hoewel hier stappen in zijn gezet, zal dit nog de nodige tijd en moeite gaan kosten.

### *Ontsluiten van data*

De HAN hanteert een strikt beleid wat betreft privacy(bescherming). De hogeschool is dan ook terughoudend in het ontsluiten van gegevens op 'lage' niveaus, zeker als het gaat om gegevens van individuele studenten. Deze voorzichtigheid rondom de ontsluiting van gegevens geldt niet alleen voor ontsluiting aan derden, ook intern worden

<sup>32</sup> Instellingen kunnen er voor kiezen om in de uitnodigingsmail voor de NSE aan te geven dat ze de resultaten verder willen analyseren met gebruikmaking van de achtergrondgegevens en student-ID van de respondent. In dat geval worden de NSE-gegevens inclusief sleutelbestand teruggestuurd aan de instelling. Dit gebeurt onder strikte voorwaarden. Zie: <https://www.studiekeuze123.nl/faq/student/nationale-studenten-enquete/wat-gebeurt-er-met-de-resultaten>



hier strenge voorwaarden aan gesteld. Als men binnen de HAN bijvoorbeeld specifieke gegevens uit de NSE wil benutten voor de eigen informatievoorziening, moet verantwoord worden waarom dat kenmerk relevant is voor onderzoek. Hierbij valt bijvoorbeeld te denken aan een studentkenmerk als 'geboortedatum': veelal zal dat gegeven niet noodzakelijk zijn voor de benodigde informatie. Indien niet goed verantwoord kan worden waarom dit wel nodig is, worden die gegevens niet ontsloten.

Het strenge beleid ten aanzien van gegevensontsluiting maakt het naar verwachting niet eenvoudig om in de toekomst studentgegevens te benutten voor onderzoek. Als een externe partij dergelijke gegevens wil benutten voor onderzoek, zal een specifieke informatievraag neergelegd moeten worden bij (onder meer) de juristen binnen de HAN. Een dergelijke vraag zal zeer sterk onderbouwd moeten worden, wil men het überhaupt overwegen. De hogeschool dient er bovendien zelf het nut van in te zien: data-ontsluiting in het kader van een database zou meerwaarde moeten bieden voor de instelling zelf.

Om studentgegevens in de toekomst te gebruiken voor een database, zou de meest kansrijke route vermoedelijk lopen via de Vereniging Hogescholen. Dat wil zeggen: indien de vereniging (en dus de hogescholen) daar achter staan. De kans op ontsluiting van gegevens is vermoedelijk het grootst als er een breed draagvlak bestaat vanuit de sector en hier gezamenlijk afspraken over gemaakt kunnen worden. Een alternatieve route om studentgegevens te kunnen benutten, zou zijn om studenten individueel om toestemming te vragen voor het beschikbaar stellen van hun gegevens. Deze route wordt echter als zeer arbeidsintensief beschouwd en vormt een grote logistieke opgave.

## 6.6 Case study 2: De Nieuwste Pabo / Zuyd Hogeschool

### *Context*

De bachelor 'leraar basisonderwijs' (pabo) wordt gezamenlijk door Zuyd Hogeschool en Fontys Hogescholen aangeboden binnen de Nieuwste Pabo te Sittard. Administratief betekent dit dat een deel van de studenten ingeschreven staan binnen Zuyd en een deel binnen Fontys (ongeveer 50/50), ondanks dat deze studenten hetzelfde onderwijs volgen. De inschrijving is relevant bij het toekennen van stageplekken: grosso modo krijgen de Fontysstudenten een plek in Sittard of ten noorden daarvan en Zuydstudenten in het zuidelijk deel van de regio. In deze paragraaf gaan we in op de studentgegevens van De Nieuwste Pabo, vanuit het perspectief van Zuyd Hogeschool.

### *Managementinformatiesysteem*

Zuyd Hogeschool heeft een managementinformatiesysteem ontwikkeld waarin drie hoofdcomponenten voorkomen. Het gaat om gegevens op het gebied van onderwijs en studenten, op het gebied van finance, en op het gebied van HR. Het systeem wordt

automatisch geüpdatet vanuit onderliggende databronnen (zoals personeelsgegevens vanuit salarisadministraties). Wat betreft de studentgegevens wordt het systeem 'gevuld' vanuit OSIRIS.<sup>33</sup> Het managementsysteem maakt het mogelijk om geaggregeerde gegevens over opleidingen, faculteiten of de instelling Zuyd als geheel te presenteren, maar kan ook gegevens over individuele studenten tonen. Studentgegevens in het systeem kunnen ruwweg in drie categorieën worden ingedeeld: ten eerste inschrijvingsgegevens vanuit Studielink, ten tweede voortgangsgegevens vanuit de opleidingen en ten derde 'belevingsgegevens' vanuit enquêtes (zoals eigen enquêtes en landelijke enquêtes zoals de Nationale Studenten Enquête). De studentgegevens blijven in principe ook na uitschrijving of diplomering in het systeem bewaard. Er lijkt geen sprake te zijn van een maximum-bewaartermijn. De geïnterviewden geven aan dat het exporteren van data vanuit het systeem als vrij eenvoudig en niet-tijdrovend kan worden beschouwd.

### *Type studentgegevens 1: personalia uit Studielink*

Studenten melden zich aan voor opleidingen via Studielink. Van iedere student kan bekeken worden of deze op een bepaald moment ingeschreven staat binnen (een) bepaalde opleiding(en) en binnen Zuyd Hogeschool. Hierdoor is inzichtelijk hoeveel uitvallers en overstappers er zijn. Aangezien Studielink verloopt via DigID zitten hier allerlei persoonsgegevens achter. Van de ingeschrevenen is daardoor onder andere bekend: geslacht, geboortedatum, vooropleiding, et cetera. In principe is het ook mogelijk om het BSN te benutten voor het eventueel koppelen van gegevens aan externe gegevens. Wel wordt nadrukkelijk de kanttekening geplaatst dat zoiets sowieso alleen te overwegen is mits er geen problemen zijn met de Wet Bescherming Persoonsgegevens.

### *Type studentgegevens 2: voortgangsgegevens*

Een studiejaar binnen Zuyd Hogeschool bestaat uit 4 blokken (van 10 weken) van elk 15 EC (studiepunten). Van alle studenten wordt bijgehouden hoeveel punten zij in de desbetreffende blokken hebben gehaald. Voor iedere student kan daardoor per blok bekeken worden hoe het gerealiseerde aantal studiepunten zich verhoudt tot het maximaal aantal punten dat behaald had kunnen worden. Ook kan voor iedere student van alle individuele opleidingsonderdelen die studiepunten opleveren, bekeken worden wat de beoordeling (cijfer) is geweest. Zodoende zijn alle resultaten op alle tentamens en stages voor studenten op te vragen. De Nieuwste Pabo kent geen hard onderscheid tussen verschillende 'vaktoetsen' in de opleiding, zoals elders het geval is. De studenten van De Nieuwste Pabo leggen kennistoetsen af aan het eind van iedere periode: geïntegreerde 'thematoetsen'. Wel wordt net als op de andere pabo's de landelijke kennistoets rekenen (WISCAT) en taal afgenomen.

Tussen de Nederlandse pabo's zijn er flinke verschillen in de curricula, zo geeft de opleidingsdirecteur van de Nieuwste Pabo aan. De Nieuwste Pabo kenmerkt zich bijvoor-

<sup>33</sup> OSIRIS is een studentinformatiesysteem dat door tientallen hogeronderwijsinstellingen wordt gebruikt om o.a. studieresultaten bij te houden.

beeld door een sterke focus op de praktijk en het scholenveld: zo vindt circa 40% van het programma plaats in de beroepspraktijk. Het gaat dan om het ontwikkelen van lessen en lespakketten, het voorbereiden op lessen, en praktijkstages. Deze elementen worden beoordeeld door de pabo (SLB'er) en op school door de betrokken opleidingsdocent en stagebeleider. Verder volgen studenten onderwijswerkgroepen: interactieve sessies waarin studenten bijvoorbeeld de gemaakte keuzes onderling bespreken en de casussen verdiepen. Ook deze werkgroepen worden beoordeeld. Ieder studiejaar kent een andere focus. In de eerste twee jaar volgt men een gezamenlijk basistraject met een kennismaking van de onder-, midden- en bovenbouw in het basisonderwijs. In het derde jaar kiezen de studenten voor de profilering 'Jonge Kind', 'Oudere Kind' of 'Scholier' en maken ze kennis met het speciaal onderwijs en de onderbouw van het vo. In het laatste, vierde, jaar is er daarnaast nog een inhoudelijke profilering. Studenten kunnen zich dan richten op 'Wetenschap & Techniek', 'Maatschappijvakken' (geschiedenis, aardrijkskunde, maatschappijleer), 'Beeldende Vorming', of 'Muziek & Drama'.

Zoals gezegd geeft het systeem de mogelijkheid om de scores op te vragen van alle onderwijsonderdelen die studiepunten opleveren. Dit betekent dat scores op toelatingstoetsen<sup>34</sup> (de zogeheten 'Toetsen voor de Poort') voorafgaand aan de pabo niet in het managementsysteem te vinden zijn. Deze gegevens worden wel (elders) geregistreerd, maar vergen vermoedelijk aanzienlijk meer inspanning om ze te extraheren. Ook kwalitatieve (tekstuele) gegevens rondom onderwijsonderdelen komen niet in het managementinformatiesysteem voor. Hiervoor zal OSIRIS c.q. Blackboard geraadpleegd moeten worden. Dit is waarschijnlijk een arbeids- en tijdsintensieve klus.

### *Type studentgegevens 3: belevingscijfers en overige gegevens*

Het managementinformatiesysteem van Zuyd biedt inzicht in de resultaten uit de Nationale Studenten Enquête (NSE). Ook eventuele andere tevredenheidsgegevens (bijvoorbeeld verzameld door Zuyd zelf) kunnen geraadpleegd worden. Verder biedt het systeem inzage in gegevens over vooraanmeldingen, instroom, doorstroom, uitstroom, rendementen, beoordelingen en studentkenmerken.

### *Potentiële dataontsluiting en vergelijkbaarheid*

De geïnterviewden van Zuyd Hogeschool staan in principe positief tegenover ontsluiting van de data. Het exporteren van data vanuit het systeem wordt zoals gezegd als vrij eenvoudig en niet-tijdrovend betiteld. Eventuele toekomstige ontsluiting van de data zal echter alleen geschieden indien de Colleges van Bestuur van Zuyd en Fontys daarmee akkoord zouden gaan.

<sup>34</sup> Afhankelijk van de vooropleiding dienen aanmelders wel of geen (specifieke) toelatingstoetsen te behalen alvorens men toegelaten kan worden tot de opleiding, zie: <http://www.denieuwstepabo.nl/studeren/toelaten/toelatingstoetsen/>

Het vermoeden bij de geïnterviewden bestaat dat het managementsysteem van Zuyd een relatieve 'voorloper' is binnen het hbo. Er kan dus niet zonder meer gesteld worden dat andere hogescholen ook (dezelfde typen) gegevens kunnen exporteren zonder al te veel tijd en moeite. Een andere beperking betreft de verschillen in curricula tussen instellingen. Ook zijn de toetsgegevens, de landelijke reken- en taaltoets uitgezonderd, volgens de opleidingsdirecteur van De Nieuwste Pabo niet zomaar vergelijkbaar. Verder wordt door de geïnterviewden aangegeven dat als studentgegevens voor onderzoek ontsloten worden, er goed nagedacht dient te worden over de wijze van aanlevering. Het zou dan noodzakelijk zijn om een standaardformat in te richten waarbinnen de gegevens gepresenteerd worden, opdat de vergelijkbaarheid van gegevens tussen instellingen zo goed mogelijk geborgd wordt. Het CBS zou hier wellicht een mediërende rol in kunnen spelen.

### Conclusies

- Er bestaat onder de lerarenopleidingen de nodige scepsis over het nut van een eventuele database, vooral onder de lerarenopleidingen in het hbo.. Zo wordt aangegeven dat de kwaliteit van het onderwijs reeds geborgd wordt door het accreditatiestelsel. Verder wordt aangegeven dat de lerarenopleidingen al voorzien worden in hun kennisbehoefte, dat de voortgang van studenten doorlopend gemonitord wordt en er al veel kennis is uit alumnistudies.
- De aanwezige kennis wordt daarbij al volop gedeeld door opleidingen onderling, zo wordt gesteld. Daar waar opleidingen wél aanvullende kennisvragen hebben, lijkt er vooral interesse te bestaan voor context-specifiek (bijvoorbeeld lokaal) en themaspecifiek onderzoek: een generieke database wordt als weinig nuttig beschouwd.
- Onderzoek naar leraarkwaliteit via een database wordt verder als uitermate complex of zelfs onmogelijk betiteld, waarbij de angst bestaat dat er al snel naar de lerarenopleidingen 'wordt gewezen' als schuldige voor eventuele mindere kwaliteit. Bij de eventuele ontwikkeling van een database is het dan ook een grote uitdaging om deze zorg weg te nemen, en met praktische opbrengsten aan te kunnen tonen wat de meerwaarde voor de opleidingen zelf kan zijn.
- De universitaire lerarenopleidingen lijken echter meer open te staan voor een database en geven aan meerwaarde te zien. Bijvoorbeeld meer inzicht verwerven in de risicofactoren van uitval van studenten en leraren, en in manieren om daarmee om te gaan.
- Naast de twijfels over de meerwaarde van een database bij veel van de opleidingen, is er bovendien grote zorg over het eigenaarschap van de data en de privacygevoeligheid van studentgegevens. Lerarenopleidingen beschikken over veel potentieel zeer interessante gegevens, zoals scores van intrede-toetsen of voortgangstoetsen, maar zullen deze zeker niet 'zomaar' prijs geven. Meerdere malen wordt 'informed consent' door studenten als minimumvereiste genoemd om data eventueel te ontsluiten.
- De bruikbaarheid van de data wordt overigens ook door diverse geïnterviewden betwijfeld: zo kunnen opleidingen en instellingen verschillen in curricula en toetswijzen, hetgeen de vergelijkbaarheid bemoei-

lijkt. Andere geïnterviewden geven echter aan dat er vermoedelijk wel een bepaalde mate van vergelijkbaarheid zal zijn, zeker binnen dezelfde 'typen' opleidingen (zoals universitaire lerarenopleidingen). De externe koppelbaarheid van studentgegevens lijkt goed, aangezien deze te relateren zijn aan individuele studentnummers en daarmee aan persoonsgegevens.

#### 'Case studies': HAN en De Nieuwste Pabo/Zuyd Hogeschool

- Een eerste blik in de studentadministratiesystemen van twee instellingen (HAN en Zuyd Hogeschool) leert dat de voortgang van studenten in principe goed digitaal gedocumenteerd wordt. Aangezien ook andere hogescholen digitale studentinformatiesystemen zullen gebruiken, kan verwacht worden dat de voortgang van studenten ook bij andere hogescholen op soortgelijke wijze inzichtelijk kan worden gemaakt.
- Zowel binnen de HAN als Zuyd is het mogelijk om individuele resultaten op onderwijsonderdelen (toetsen, stagebeoordelingen et cetera) op te vragen. Het feit dat dergelijke resultaten digitaal bewaard worden, betekent echter niet dat deze ook zonder meer geëxporteerd kunnen worden als onderzoeksbestand. Zo wordt door de HAN aangegeven dat het uitdraaien van studentgegevens voor onderzoeksdoeleinden een tijdrovende, arbeidsintensieve klus is en bepaald geen 'druk op de knop'. Wel worden er binnen de HAN op dit gebied stappen gezet om studentgegevens beter te gaan benutten voor de interne informatievoorziening. Binnen Zuyd Hogeschool is de afgelopen jaren een managementinformatiesysteem ontwikkeld dat behalve met financiële en HR-gegevens voor een belangrijk deel automatisch wordt 'gevuld' door data uit het studentvolgsysteem OSIRIS. Met behulp van het managementinformatiesysteem is het mogelijk om op verschillende niveaus (instelling, faculteit, opleiding, maar ook de individuele student) relatief eenvoudig datasets te exporteren.
- Naast verschillen in de technische mogelijkheden om studentgegevens te ontsluiten, lijkt ook de bereidheid tussen instellingen te verschillen om dergelijke gegevens te ontsluiten voor onderzoeksdoeleinden. Zo geeft de HAN aan doorgaans terughoudend te zijn in het ontsluiten van gegevens op lage niveaus (zoals individuen) vanwege privacybescherming. Bij Zuyd Hogeschool wordt vanzelfsprekend ook groot belang gehecht aan privacybescherming, maar lijkt men meer mogelijkheden te zien. Over dergelijke verzoeken zal echter altijd (onder meer) het CvB zich moeten buigen.

Relevantie: groot

Haalbaarheid: verschillend



# 07

## LERARENREGISTER: PROFESSIONALISERINGSGEGEVENS

### 7.1 Inleiding

Het Lerarenregister is een beroepsregister voor leraren waarmee zij kunnen aantonen dat ze bevoegd en bekwaam zijn en doorlopend investeren in hun professionele ontwikkeling. Leraren houden in het register hun professionaliseringsactiviteiten bij, waarbij zowel formele als informele activiteiten kunnen worden toegevoegd. Er is sprake van een herregistratiecyclus van vier jaar. In die vier jaar tijd dienen leraren minimaal 160 uur te hebben geïnvesteerd in hun professionele ontwikkeling, waarvan minimaal 100 uur aan vakinhoudelijke, didactische en pedagogische bekwaamheid. Het bijhouden van het type activiteit, de omvang en de evaluatie van de activiteit zou tevens de discussie moeten aanzwengelen over welke activiteiten nuttig zijn, welke minder, en indien ze nuttig zijn voor wie, wanneer en waarom. Aan de hand van dergelijke reflectie hoopt de Onderwijscoöperatie tot een overzicht te komen van activiteiten die zinvol worden geacht door de beroepsgroep, waar leraren uit kunnen kiezen en waarmee opleidingstrajecten ontwikkeld kunnen worden.

Het Lerarenregister is in 2012 door de Onderwijscoöperatie<sup>35</sup> gestart als een vrijwillig register, maar wordt omgezet naar een verplicht beroepsregister. Op 21 februari 2017 stemde de Eerste Kamer in met de wet die dit regelt: de wet *Beroep Leraar en Lerarenregister*. Leraren kunnen zich nog tot 1 augustus 2017 aanmelden in het vrijwillige register. Daarna wordt de wet van kracht en wordt alle focus gelegd op de inrichting en implementatie van het nieuwe lerarenregister. Vanaf 1 augustus 2018 kunnen bevoegde leraren zich inschrijven in het nieuwe, verplichte register. Per 1 augustus 2019 dienen alle bevoegde leraren geregistreerd te zijn in het nieuwe register en gaan ze hun (professionaliserings)activiteiten bijhouden.

<sup>35</sup> De Onderwijscoöperatie is in 2011 opgericht door de vijf grootste onderwijsvakverenigingen AOb, CNV Onderwijs, FvOv, BON en Platform VVVO. Doel is het stimuleren van de ontwikkeling van de beroepsgroep, waarbij leraren zelf verantwoordelijkheid nemen en zeggenschap hebben over hun beroep.

## 7.2 Visie op database

### *Nuttig maar complex*

In algemene zin wordt een database als potentieel nuttig gezien, waarmee sterker onderbouwd onderzoek naar de beroepsgroep kan worden verricht. Er kan bijvoorbeeld gedacht worden aan onderzoek naar de effecten van (bepaalde typen) professionaliseringsactiviteiten op het onderwijs. De Onderwijscoöperatie geeft aan blijvend belang te hechten aan kennis over het nut en de effectiviteit van professionalisering en wil daar ook adviezen over uitbrengen. Inzichten in de doelmatigheid van scholing en andere vormen van professionele ontwikkeling zijn dan ook zeker gewenst.

Wel worden er nadrukkelijke kanttekeningen bij dergelijk onderzoek geplaatst. Zo is het uiterst complex om 'onderwijskwaliteit' te meten: *“Waar heb je het dan over? Gaat het om leeropbrengsten, cognitieve opbrengsten, sociale ontwikkeling of weer iets anders?”* Daarbij spelen er zeer veel factoren op allerlei niveaus (zoals op het niveau van het stelsel, van de school en van het individu) die kwaliteit kunnen beïnvloeden. Hier dient men zich volgens de Onderwijscoöperatie zeer bewust van te zijn. Het is daarbij belangrijk om vooraf goede, heldere onderzoeksvragen te formuleren die door analyse van de data beantwoord kunnen worden. Het opzetten van een database zonder achterliggende vragen, maar simpelweg 'omdat het kan', wordt als zeer onverstandig betiteld: *“Aan de hand van de vragen die je hebt, moet je gericht op zoek naar de juiste data, en niet andersom.”*

### *Praktische overwegingen*

Om draagvlak voor een database te ontwikkelen, is het belangrijk om goede, sprekende voorbeelden te geven van de mogelijke opbrengsten van gekoppelde data. Dat draagvlak is er namelijk zeker niet zonder meer: *“In algemene zin heerst er een zekere 'schrik' voor databases, een soort 'Big Brother' bestanden. Er is angst voor eventuele negatieve maatschappelijke implicaties.”* Om het veld mee te krijgen, is het daarom van groot belang om bijvoorbeeld aan te kunnen geven wat de concrete verwachte opbrengsten zijn.

Bij de haalbaarheid van een database worden de nodige vraagtekens gezet. Nadrukkelijke vragen die bij de Onderwijscoöperatie rijzen zijn: Hoe zit het met de herleidbaarheid van de gegevens tot personen? Mogen gegevens voor andere doeleinden gebruikt worden dan waarvoor deze oorspronkelijk verzameld zijn? Heeft de oorspronkelijke gegevensverstrekker invloed op het verdere gebruik van de gegevens? En wie is uiteindelijk 'eigenaar' van de gegevens?

## 7.3 Aanwezige data

Het Lerarenregister is ontwikkeld om de positie van de leraren te versterken: *“Met het register maakt de leraar maatschappelijk zichtbaar dat hij de professionele standaard van*



*zijn beroep onderschrijft en daaraan voldoet.*" Het register is dan ook expliciet niet ontwikkeld als potentiële onderzoeksdatabase. Dit heeft grote consequenties voor de eventuele bruikbaarheid van de gegevens voor onderzoeksdoeleinden. Het Lerarenregister staat bovendien nog maar aan het begin en is volop in ontwikkeling.<sup>36</sup> Op moment van dit schrijven hebben ruim 63 duizend leraren zich vrijwillig ingeschreven,<sup>37</sup> maar het register wordt – indien het wetsvoorstel wordt aangenomen – op termijn verplicht en daarmee in principe integraal. Ook technisch is het register nog sterk in ontwikkeling. De datastructuur is op dit moment nog lang niet voldoende uitgekristalliseerd om het register te benutten voor onderzoek, zo wordt aangegeven. De Onderwijscoöperatie geeft aan dat het nog wel tien jaar kan duren voordat er een robuuste datastructuur ontwikkeld is die daar wel 'geschikt' voor is.

Er wordt doorlopend nagedacht over het opnemen en verder onderverdelen van activiteiten in het systeem. Op dit moment kent het register een standaardclustering van zes typen professionalisering, namelijk: cursussen; conferenties; activiteiten binnen school; activiteiten buiten school; individuele activiteiten. Dit betreft een zeer algemene indeling en is bedoeld als een eerste, handzame sjabloon. Bovendien kunnen leraren gebruik maken van deze zes categorieën, maar dat hoeft niet. Het is ook mogelijk om zelf activiteiten te benoemen en in te vullen. Naast het '6-categorieënmodel' bestaat het systeem momenteel uit een 'urenteller' (dat wil zeggen het aantal uur besteed aan bepaalde activiteit) en een evaluatiesysteem (dat wil zeggen oordelen over de activiteit, zoals: heeft de leerling er iets aan, zou u het aanbevelen, et cetera).

### *Ontsluiting*

Mochten gegevens uit het Lerarenregister in de toekomst benut kunnen worden voor onderzoek, dan zou in ieder geval geëist worden dat hier op een zorgvuldige, van tevoren gereguleerde manier mee omgegaan moet worden. Een reglementering waarmee de organisaties van de Onderwijscoöperatie en de achterban akkoord gaan. De leraar dient daarbij de eigenaar te zijn en te blijven van zijn eigen gegevens, waarbij actieve toestemming nodig is van leraren én overheid om dergelijke data voor andere doelen te gebruiken dan waarvoor ze oorspronkelijk zijn geleverd.

### *Koppelmogelijkheden externe bestanden*

Mochten gegevens uit het Lerarenregister op termijn ontsloten worden, dan lijkt het koppelen van gegevens aan andere data niet problematisch. Op dit moment is het Lerarenregister nog vrijwillig, wat betekent dat leraren nu een account aanmaken en een aantal persoonsgegevens invullen: NAW, geboortedatum, geslacht, school (BRIN) en hun bevoegdheid. Als het register verplicht wordt, zal de registratieprocedure verder verbeterd worden. Bij het creëren van een account kunnen de achtergrondgegevens

<sup>36</sup> Ook wordt naast het registratiesysteem momenteel een herregistratiesysteem ontwikkeld.

<sup>37</sup> Cijfers afkomstig van <http://www.wijzijnvoorhetlerarenregister.nl/>

dan al worden 'klaargezet' door DUO via DigiD, en dient de leraar enkel nog de gegevens te controleren en goed te keuren.

#### Conclusies

- Het Lerarenregister is een beroepsregister voor leraren waarin zij hun professionele ontwikkeling (zowel formeel als informeel) bijhouden. Het Lerarenregister is in 2012 gestart als een vrijwillig register, maar wordt conform de wet Beroep Leraar en Lerarenregister omgevormd tot een verplicht register (per augustus 2018 dienen alle bevoegde leraren geregistreerd te zijn en moeten ze hun activiteiten bijhouden).
- In algemene zin wordt een database als potentieel nuttig gezien, bijvoorbeeld om onderzoek te verrichten naar de effecten van professionaliseringsactiviteiten op het onderwijs. Bij het opzetten van een database is het noodzakelijk om concrete vragen te hebben. Het opzetten enkel 'omdat het kan', wordt onwenselijk geacht. Om draagvlak voor een database te creëren en argwaan weg te nemen, is het belangrijk om sprekende voorbeelden te geven van de opbrengsten van gekoppelde data. Verder dient goed nagedacht te worden over hoe omgegaan wordt met persoonsgegevens, wie de eigenaar van de database is et cetera.
- Het Lerarenregister is niet ontwikkeld als potentiële onderzoeksdatabase en is nog volop in ontwikkeling. De datastructuur is momenteel nog lang niet voldoende uitgekristalliseerd om te benutten voor onderzoek. De Onderwijscoöperatie geeft aan dat het nog wel tien jaar kan duren voordat er een robuuste datastructuur ontwikkeld is die daar wel 'geschikt' voor is. Op dit moment kent het systeem een '6-categorieënmodel' van professionaliseringsactiviteiten, een 'urenteller' en een evaluatiesysteem. Er wordt continu nagedacht over verbetering van het systeem (bijvoorbeeld over het onderbrengen van activiteiten). Het register bevat veel interessante data over de professionalisering van leraren, maar de benutting is voorlopig dus nog niet aan de orde.
- Ook wanneer de gegevens in de toekomst 'bruikbaar' worden voor onderzoek, dan zullen deze zeker niet zonder meer ontsloten worden. Er zal in ieder geval sprake moeten zijn van een reglementering waarmee de organisaties van de Onderwijscoöperatie en de achterban akkoord gaan. De leraar dient daarbij de eigenaar te zijn en te blijven van zijn eigen gegevens, waarbij actieve toestemming nodig is van leraren én overheid om dergelijke data voor andere doelen te gebruiken dan waarvoor ze oorspronkelijk zijn geleverd.
- De externe koppelbaarheid van de gegevens lijkt goed mogelijk (in vrijwillige register momenteel o.a. NAW, geboortedatum, BRIN etc.), zeker nadat het register verplicht wordt (registratie loopt dan via DigiD).

Relevantie: groot

Haalbaarheid: (komende jaren) gering

# 08

## VERDERE ONDERZOEKSBESTANDEN

### 8.1 Inleiding

In de voorgaande hoofdstukken hebben we aan de hand van gesprekken met inhoudelijke experts diverse typen bestanden in detail besproken. Het ging daar onder meer om CBS-bestanden, DUO-bestanden, studentgegevens van lerarenopleidingen, et cetera. In die hoofdstukken hebben we bovendien een inschatting gemaakt van de inhoudelijke relevantie voor een database van (startende) leraren en lerarenopleidingen en de (technische) haalbaarheid van het benutten van de bestanden. Naast de voorgaande bestanden zijn er echter nog vele andere datasets die van waarde kunnen zijn voor een database. In Hoofdstuk 2 gaven we hiervan al een breed overzicht. Een aantal van deze bestanden willen we gezien hun potentiële relevantie voor een database in dit hoofdstuk nader bespreken. Het gaat hier om databestanden in de projecten 'Begeleiding Startende Leraren', de 'Loopbaanmonitor' en de 'HBO-Monitor'. Deze drie bestanden hebben gemeen dat ze (op verschillende manieren) inzicht geven in de eerste stappen en ervaringen van afgestudeerde leraren op de arbeidsmarkt. Het kan in eventuele latere fasen van een database daarom de moeite waard zijn om de mogelijkheden omtrent deze onderzoeksbestanden nader te verkennen (ten aanzien van ontsluiting, externe koppelbaarheid et cetera). Van belang is daarbij tevens dat de laatste twee onderzoeken veel overlap kennen wat betreft de bevraging van dezelfde groep afgestudeerden van tweedegraads lerarenopleidingen in ongeveer dezelfde periode. Twee eerdere pogingen gericht aan OCW om de Loopbaanmonitor in de HBO-monitor te integreren zijn echter mislukt.

### 8.2 Begeleiding Startende Leraren<sup>38</sup>

Binnen het project 'Begeleiding Startende Leraren' (BSL), uitgevoerd door de Rijksuniversiteit Groningen (op verzoek van OCW), wordt onderzoek verricht naar de effecten van inductiemanagement op de ontwikkeling van startende leraren in het voortgezet onderwijs. Het doel van BSL is om de uitval van beginnende leraren in het voortgezet onderwijs te verminderen en om hun kwaliteit te versterken. Het landelijke onderzoek richt zich op de vragen als 'Zijn de (regionale) inductiearrangementen effec-

<sup>38</sup> Informatie afkomstig van [www.begeleidingstartendeleraren.nl](http://www.begeleidingstartendeleraren.nl).

tief?’ en ‘Kunnen de inzichten die voortkomen uit de samenwerking tussen de scholen en de universitaire lerarenopleiding (ulo) en het hoger beroeps onderwijs (hbo) leiden tot verbeterde inwerktrajecten voor starters?’ Naast het landelijk onderzoek wordt in de negen deelnemende regio’s flankerend onderzoek verricht onder leiding van de regionale projectleider. Om na te gaan of de begeleiding van starters oplevert wat het beoogt, wordt onderzocht:

- de daadwerkelijke implementatie van het inductiearrangement;
- de kwaliteit van pedagogisch-didactisch handelen van een starter vanuit het perspectief van iemand met ruime ervaring in het onderwijs;
- de kwaliteit van pedagogisch-didactisch handelen van een starter door de ogen van leerlingen;
- de longitudinale ontwikkeling van het pedagogisch-didactisch handelen van starters;
- eventuele uitval uit het beroep en vertrek naar andere scholen.

Scholen in het voortgezet onderwijs kunnen startende leraren (dat wil zeggen lesbevoegde leraren met maximaal twee jaar ervaring) aanmelden voor het onderzoek en committeren zich voor drie jaar aan het landelijke onderzoek. De scholen kunnen aanspraak maken op een financiële vergoeding van 2.000 euro per starter. Vanaf schooljaar 2014-2015 doen 102 instellingen mee, vanaf schooljaar 2015-2016 gaat het om 165 schoolinstellingen. In het eerste projectjaar 2014-2015 begonnen 856 starters aan de regionale projecten. In het tweede projectjaar 2015-2016 begonnen 1.353 starters aan de regionale projecten, onder wie 751 vrouwen.

Om de ontwikkeling van leraren te monitoren, zijn er (per deelnemend cohort) vier dataverzamelingsrondes. In iedere ronde gaat het in ieder geval om een lesobservatie via ICALT (observatie-instrument waarmee vastgesteld kan worden welk gedrag de leraar in de klas vertoont op het terrein van de pedagogische en didactische competenties) en om een leerlingvragenlijst (leerlingpercepties over pedagogisch-didactisch handelen van de leraar). Daarnaast vult de leraar driemaal een zogenaamde ‘begeleidingsmonitor’ in (perceptie van de leraar over de aangeboden ondersteuning). Aan het begin van het traject vult de leraar een toestemmingsverklaring in: zonder deze verklaring kunnen de gegevens niet benut worden voor het onderzoek.

### 8.3 Loopbaanmonitor<sup>39</sup>

De Loopbaanmonitor wordt sinds 2004 uitgevoerd op verzoek van OCW (huidige uitvoering door CentERdata en MOOZ) en heeft als doel het verkrijgen van inzicht in de arbeidsmarktsituatie van pas afgestudeerde leraren en het beroepsrendement van de lerarenopleidingen een half, één, twee, drie, vier en vijf jaar na diplomering, dat wil

<sup>39</sup> Informatie afkomstig van [www.centerdata.nl](http://www.centerdata.nl) en rapportage ‘Loopbaanmonitor onderwijs’ (Fontein et al., 2016).

zeggen het aandeel afgestudeerden dat ook daadwerkelijk in het onderwijs aan de slag gaat. De Loopbaanmonitor richt zich zowel op leraren in het po als het vo. De focus van de Loopbaanmonitor ligt op de eerste fase na het afstuderen, omdat leraren vooral in de eerste jaren na hun afstuderen het onderwijs verlaten. De gegevens worden tevens als input gebruikt in andere onderzoeken. Zo wordt de monitor door CentERdata geraadpleegd ten behoeve van de modellering van de onderwijsarbeidsmarkt (via Mirror). De Loopbaanmonitor bestaat uit twee onderdelen:

- een halfjaarlijkse webenquête (april/mei en oktober/november) onder leraren die in het jaar daarvoor hun eerste diploma voor een lerarenopleiding hebben behaald;
- analyses op de microdata van het CBS: gegevens op persoonsniveau die vanuit verschillende bestanden bij elkaar zijn gebracht.

Het voordeel van de microdata is dat het om administratieve gegevens van de gehele populatie gaat. Met andere woorden, de informatie over loopbaanstappen is niet 'gekleurd' door een responsbias, die veroorzaakt kan worden doordat pas afgestudeerden die werkzaam zijn in het onderwijs eerder geneigd zijn om deel te nemen aan het invullen van een vragenlijst voor de Loopbaanmonitor dan degenen die buiten het onderwijs of niet werkzaam zijn. Nadeel van het gebruik van de administratieve data is dat deze informatie steeds een jaar 'achterloopt' op de enquêtegegevens, aangezien de bestanden niet 'realtime' beschikbaar zijn. De enquêteresultaten zijn daarmee actueler en geven bovendien verdiepend inzicht in onder meer het type begeleidingsactiviteiten voor beginnende leraren, de tevredenheid van leraren over de begeleiding en de baan als geheel, maar bijvoorbeeld ook in het geloof in het eigen kunnen. Daarnaast geeft de enquête zicht op baankenmerken die niet vanuit de administratieve gegevens van het CBS te extraheren zijn, zoals het type aanstelling en dienstverband.

Voor de enquêtegegevens uit de Loopbaanmonitor zouden voor een database voor (startende) leraren en lerarenopleidingen interessant kunnen zijn. De CBS-microdata zijn dat vanzelfsprekend ook, maar op deze gegevens zijn we in Hoofdstuk 3 al uitgebreid ingegaan. De enquêtegegevens worden specifiek in het kader van de Loopbaanmonitor verzameld en kunnen derhalve van aanvullende waarde zijn op een database. Voor de webenquête worden, aan de hand van bestanden van DUO en via briefaanschrijving dóór DUO, alle afgestudeerden benaderd die een diploma hebben gehaald aan de pabo of de lerarenopleidingen voor het vo. De respons op de briefaanschrijving in april/mei 2015 betrof 28 procent, oftewel ruim 2.400 respondenten. Vervolgens is aan de deelnemers aan de webenquête gevraagd of zij bereid waren om een half jaar later deel te nemen aan een vervolgenquête (waarvan 45 procent ook daadwerkelijk repondeerde, oftewel ongeveer 1.100 respondenten).

## 8.4 HBO-Monitor<sup>40</sup>

Onder auspiciën van de Vereniging Hogescholen voert het ROA (Universiteit Maastricht) sinds de jaren negentig jaarlijks de HBO-Monitor uit. De HBO-Monitor is een instrument waarmee hogescholen ieder jaar de arbeidsmarktpositie van hun afgestudeerden kunnen vaststellen en waarmee zij kunnen zien in hoeverre opleidingen aansluiten op de eisen van de beroepspraktijk. Ongeveer negen op de tien hogescholen nemen deel aan de HBO-Monitor en alle hbo-opleidingssectoren zijn er in vertegenwoordigd. In de HBO-Monitor worden drie componenten onderscheiden: een landelijke kern, een domeinspecifiek deel en een instellingsspecifiek deel. De landelijke kern omvat gegevens over:

- de arbeidsmarktpositie van afgestudeerden (zoals type baan, type aanstelling, lonen et cetera);
- generieke competenties (zoals kennis van vakgebied, logisch redeneren, zelfstandigheid et cetera);
- de relatie tussen de opleiding en de arbeidsmarkt (aansluiting opleiding-werk qua niveau en richting; tevredenheid over voorbereiding op loopbaan et cetera);
- de mening van de afgestudeerden over de opleiding (studietevredenheid, oordeel over docenten, ingeschatte moeilijkheidsgraad et cetera).

De gegevens zijn hbo-breed vergelijkbaar en voor een aantal kernindicatoren zelfs onderwijsbreed, door vergelijkbare enquêtes in vo, mbo en wo. Het domeinspecifieke deel omvat de domeinspecifieke competenties en eventueel toegespitste vragen over de domeinspecifieke arbeidsmarkt. Het instellingsspecifieke deel betreft de vragen in de enquête die door de instellingen zelf kunnen worden bepaald.

De HBO-Monitor bestaat uit een enquête via internet; een schriftelijke vragenlijst wordt toegestuurd aan de afgestudeerden die niet responderen via internet. De enquête wordt ieder najaar gehouden onder afgestudeerden van een jaar daarvoor. Dit betekent dat de enquête ongeveer anderhalf jaar na het afstuderen plaatsvindt. Aangezien de HBO-Monitor vrijwel het gehele hbo afdekt, levert het onderzoek een zeer grote responsgroep op. In de meting van de arbeidsmarktpositie in 2016 namen ongeveer 24.000 afgestudeerden deel aan de enquête, een respons van 41 procent. Hoewel de monitor niet specifiek gericht is op afgestudeerden aan de lerarenopleidingen, zijn zij hier wel met veel personen in vertegenwoordigd. Aan de meting van 2016 namen ruim 3.200 afgestudeerden uit de verschillende lerarenopleidingen in het hbo deel (pabo, tweedegraads lerarenopleiding, master educational needs, enzovoorts).

<sup>40</sup> Informatie onder andere afkomstig van [www.hbomonitor.nl](http://www.hbomonitor.nl) en [www.roa.nl](http://www.roa.nl).

## Conclusies

In de voorliggende rapportage wordt over een aantal kernbestanden inzicht gegeven in de meerwaarde en de technische haalbaarheid van de data wat betreft de benutting voor een database. Er zijn echter nog vele andere bestanden die een nadere verkenning behoeven, waarvan we in dit hoofdstuk een drietal bespreken. Het betreffen onderzoeksbestanden die zich (vooral) richten op de eerste stappen van afgestudeerde leraren op de arbeidsmarkt.

- **Begeleiding Startende Leraren (BSL):** sinds 2014 uitgevoerd door de Rijksuniversiteit Groningen op verzoek van OCW. Landelijk onderzoek naar de effecten van inductiemanagement op de ontwikkeling van startende leraren in het voortgezet onderwijs. Het onderzoek richt zich vooral op de vragen of inductiearrangementen effectief zijn en hoe deze verbeterd kunnen worden. Vanaf schooljaar 2015-2016 nemen ca. 1.350 starters deel. Zij worden gemonitord via lesobservaties (ICALT) en vragenlijsten (zowel onder leerlingen als starters).
- **Loopbaanmonitor:** jaarlijks sinds 2004 op verzoek van OCW, huidige uitvoering door CentERdata en MOOZ. De Loopbaanmonitor heeft als doel het verkrijgen van inzicht in de arbeidsmarktsituatie van pas afgestudeerde leraren (po en vo) en bestaat uit twee onderdelen. Enerzijds een halfjaarlijkse webenquête onder pas afgestudeerde leraren over o.a. de begeleiding die zij hebben ontvangen en hun inschatting van het eigen kunnen (respons eerste meting ca. 2.400). Anderzijds analyses op microdata van het CBS, onder andere om het beroepsrendement (wie komt in het onderwijs terecht, wie daarbuiten) in kaart te brengen. Voor een database zijn vooral de enquêtegegevens potentieel interessant, de CBS-microdata hebben we immers al afzonderlijk in voorliggend rapport behandeld.
- **HBO-Monitor:** jaarlijks sinds 1990, uitvoering door ROA (Universiteit Maastricht) onder auspiciën van de Vereniging Hogescholen. De HBO-Monitor geeft inzicht in de arbeidsmarktpositie van afgestudeerde hbo'ers anderhalf jaar na diplomering en toont in hoeverre opleidingen aansluiten op de eisen van de beroepspraktijk (zoals beschikking over door werkgever gevraagde competenties, match tussen opleiding en werk qua niveau en richting et cetera). Tevens geeft het inzicht in het oordeel van afgestudeerden over de opleiding (bijvoorbeeld over de diepgang en de kwaliteit van de docenten). De HBO-Monitor bestaat uit een webenquête (of eventueel papieren enquête). In de meting van 2016 namen ongeveer 24.000 afgestudeerden deel (circa 41 procent respons) waarvan ongeveer 3.200 afkomstig uit de verschillende lerarenopleidingen in het hbo.

De twee laatste onderzoeken kennen een grote overlap wat betreft de bevraging van dezelfde groep afgestudeerden van tweedegraads lerarenopleidingen. Bij de totstandkoming van een Leraren database zou een dergelijke dubbele bevraging uiteraard voorkomen moeten worden.

**Relevantie:** Redelijk/groot

**Haalbaarheid:** Nader te onderzoeken





# 09

## OCW: KIJK OP DATABASE

### 9.1 Inleiding

Het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW) draagt de zorg voor het scheppen van een wettelijk kader voor het onderwijs, voor het uitvoeren van onderwijswetten, en voor het verstrekken van financiële middelen daarvoor. Daarnaast heeft het ministerie de zorg voor de bevordering van wetenschappelijk onderwijs en wetenschapsbeleid en voor het cultuur- en mediabeleid. Onder het ministerie vallen een aantal diensten en instellingen, zoals DUO, de Inspectie van het Onderwijs en de Onderwijsraad. Binnen OCW is de Directie Kennis verantwoordelijk voor beleids- en scenario-ontwikkeling op het terrein van onderwijs, cultuur en wetenschap.

### 9.2 Visie op database

Het idee van een database is niet nieuw voor OCW. Zo zijn er begin 2015 contacten geweest met het NRO over een verkennend onderzoek naar een dergelijke database. Het ministerie staat in principe positief tegenover het beter benutten van gegevens, en ziet duidelijke meerwaarde in het koppelen van bestanden. Zo wordt een database waardevol geacht voor de monitoring en evaluatie van allerhande beleidsmaatregelen op het vlak van leraren. Er kan dan gedacht worden aan maatregelen als de lerarenbeurs, subsidie zij-instroom, opleiden in de school, aanscherping toelatingseisen pabo's, subsidieregeling begeleiding startende leraren, tegemoetkoming studiekosten lerarenopleidingen tekortvakken, Eerst de Klas, Educatieve Minor, Versterking functiemix, enzovoorts.

Het vaak gebruikte alternatief is enquête-onderzoek, waarbij altijd een behoorlijk risico bestaat op een selectieve en lage respons, en waarmee meestal veel hogere kosten zijn gemoeid. Ook wordt in dit soort onderzoeken veelal slechts op één moment gemeten, terwijl meerdere metingen in de tijd voor beleid van belang zijn om middellange en langetermijneffecten in beeld te brengen. Dat kan juist uitstekend met administratieve data. In het VK en de VS zijn hier al ruime ervaringen mee opgedaan en dit levert waardevolle data voor beleid en wetenschap op. Deze meerwaarde kan (op termijn) nog sterk

worden vergroot als koppelingen (via NCO) kunnen worden gemaakt met leerlingresultaten, lesobservaties en Inspectiedata.

Ook is het verminderen van de onderzoekslast een groot voordeel. Niet alleen omdat er minder vragen gesteld hoeven worden (zo is het bijvoorbeeld niet meer nodig om naar achtergrondgegevens te vragen in enquêtes), maar ook omdat er zoals gezegd zonder enige bevraginglast nuttig onderzoek kan worden verricht. Een goed voorbeeld daarvan is het CPB-onderzoek 'Paden naar Leraarschap' (zie ook hoofdstuk 10) waarin onder meer personele transities tussen regio's en tussen scholen aan de hand van administratieve data in kaart worden gebracht.

### 9.3 Uitdagingen

Bij het eventueel opzetten van een database dient goed nagedacht te worden over de wettelijke mogelijkheden en het databeheer. Zo zou moeten worden nagegaan wat dit juridisch vergt aan de voorkant van het proces in termen van informed consent van betrokken studenten en leraren. Een voor de hand liggende insteek zou zijn om het databeheer bij het CBS te beleggen, maar dan moeten de leraargegevens wel aansluiten bij de doelstelling van het CBS (gegevensverzameling in het kader van beleid, maatschappij, wetenschap). Een andere insteek zou zijn om de database te laten beheren door DUO. Een database buiten het CBS of DUO om lijkt niet haalbaar vanuit praktisch oogpunt en ook niet zo wenselijk. Zo is de meest betrouwbare manier om bestanden aan elkaar te koppelen aan de hand van BSN. Echter mogen alleen overheidsinstanties (en enkele andere organisaties waarvoor dit wettelijk bepaald is) het BSN gebruiken (om hun taak uit te voeren), waardoor koppelingen buiten het CBS of DUO niet voor de hand liggen.

Of het databeheer belegd wordt bij het CBS of DUO is voor het ministerie afhankelijk van het type gegevens dat in de database wordt opgenomen. Als het bijvoorbeeld alleen gaat om gegevens die DUO verzamelt, dan zou een CBS-route omslachtig zijn. DUO is immers dé expert op het gebied van integrale onderwijsdata op persoonsniveau. Als de database naast DUO-gegevens ook uit andere gegevens zou bestaan, of als er vergelijkingen met andere (bedrijfs)sectoren gemaakt moeten worden, dan is beheer door het CBS logischer.

Het ligt voor OCW voor de hand om de reeds bestaande, rijke (personeels)data bij DUO breder te gaan benutten. Er zijn echter ook een aantal beperkingen waar rekening mee gehouden dient te worden. Zo zijn de gegevens weliswaar uitgebreid, maar niet integraal. Er wordt bijvoorbeeld aangegeven dat er in het mbo soms door meer dan 10 procent van de instellingen geen gegevenslevering tot stand komt. Ook komt het soms voor dat gegevens afkomstig van salarisverwerkers niet herleidbaar zijn tot de formatiegegevens, omdat er geen uniek identificatienummer in wordt opgenomen. In de analyses kan het dan zijn dat iemand onterecht wordt aangemerkt als uitstromer (als

deze in de voorgaande 'meting' nog wél gekoppeld kon worden) of als instromer (als deze in de vorige 'meting' niet gekoppeld kon worden).

#### Conclusies

- OCW staat in principe positief tegenover het benutten van gekoppelde leraargegevens en geeft aan contact te hebben gehad met NRO om dit te verkennen. Een groot voordeel van het benutten van bestaande data is de vermindering van de onderzoekslast in het veld. Enerzijds kan het de actieve bevraging verminderen, anderzijds kan er zelfs zónder actieve bevraging veel verdiepend en verbredend onderzoek verricht worden.
- Aan de hand van gekoppelde data zou bijvoorbeeld meer inzicht verkregen kunnen worden in de effectiviteit van beleidsmaatregelen, in arbeidsmarktstromen, en in factoren die samenhangen met leskwaliteit (bijvoorbeeld via lesobservatiegegevens). Ook wordt het koppelen van leraargegevens aan leerlinggegevens (NCO) genoemd als interessante insteek voor de toekomst. Hiermee zou meer kennis kunnen worden verkregen in de mogelijke voorspellers van leerlingprestaties.
- Als een database ontwikkeld zou worden, lijkt het voor de hand te liggen om het databeheer te beleggen bij CBS of DUO. Afhankelijk van welk type data opgenomen wordt ligt ofwel het CBS ofwel DUO meer voor de hand. Als het gaat om het koppelen van onderwijsdata waar DUO reeds over beschikt, dan is het logisch om de database te laten beheren door DUO. Als de database breder ingestoken wordt en/of men vergelijkingen wil maken met andere sectoren dan is de CBS-route logischer.



# 10

## CPB: ERVARING MET GEKOPPELDE LERAARGEGEVENS

### 10.1 Inleiding

Het Centraal Planbureau (CPB) is een onderzoeksinstituut dat economische beleidsanalyses maakt en organisatorisch onder het Ministerie van Economische Zaken (EZ) valt. Het betreft onderzoek op eigen initiatief of op verzoek van de regering, het parlement, Kamerleden, vakbonden en werkgeversorganisaties. In de recente CPB-studie 'Paden naar leraarschap'<sup>41</sup> is via de geanonimiseerde, beveiligde microdataomgeving van het CBS onderzoek verricht op data van het CBS (zoals achtergrondgegevens en baangegevens), DUO (zoals inschrijvings-, diplomering- en onderwijssubsidiedata) en het Platform Bètatechniek (gegevens over onderwijstrainees in dit gebied). In zekere zin kan hier dus gesproken worden van een (beperkte) database van (startende) leraren. De ervaringen van het CPB zijn derhalve leerzaam voor deze verkenning.

### 10.2 'Paden naar leraarschap': de mogelijkheden

De CPB-studie 'Paden naar leraarschap' biedt een interessant voorbeeld van de onderzoeksmogelijkheden die ontstaan door leraargegevens te koppelen en in relatie tot elkaar te analyseren. De gegevens hebben betrekking op afgestudeerden van de lerarenopleidingen tussen 2000 en 2015. Het gaat in het onderzoek om administratieve, kwantitatieve opleidings- en arbeidsmarktindicatoren (dus geen kwaliteitsindicatoren).

In de studie staan zes onderzoekselementen centraal die met de combinatie van data van het CBS, DUO en Platform Bètatechniek inzichtelijk gemaakt worden. Het betreft:

- Het aantal inschrijvingen bij lerarenopleidingen;
- Verschillen in diplomarendement tussen lerarenopleidingen;
- Verschillen in instromerskenmerken;
- Het baanrendement van de verschillende lerarenopleidingen (oftewel het aandeel dat werkzaam is in het onderwijs);

<sup>41</sup> CPB (2017), *Paden naar leraarschap*, Den Haag: Centraal Planbureau.

- De loonverdeling tussen afgestudeerde leraren die wel/niet in het onderwijs werken;
- De samenhang tussen het eindexamencijfer op de middelbare school en de kans op het afronden van de lerarenopleiding c.q. de kans op het werken in het onderwijs;
- De sectoren waar afgestudeerde leraren werken als zij geen baan in het onderwijs hebben.

Door de koppeling van deze gegevens is het bijvoorbeeld mogelijk om te herkennen wie er als leraar werkzaam zijn (functieherkenning die in de reguliere administratieve CBS-data niet mogelijk is, enkel onderscheid naar sectoren is daarin mogelijk) en (gedeeltelijk) wie de zij-instromers zijn.

Het integrale karakter van de administratieve data maakt het verder mogelijk om op lage niveaus uitsplitsingen te doen. Zo onderscheidt het CPB in 'Paden naar leraarschap' afgestudeerden van de pabo, de tweedegraads en de eerstegraads lerarenopleiding op het hbo, de universitaire eerstegraads lerarenopleiding en de ongegradeerde lerarenopleidingen (kunstvakken en lichamelijke opvoeding). Ook is het vaak mogelijk gebleken om onderscheid tussen voltijders en deeltijders te maken en/of onderscheid naar schoolvakken te maken.

Het koppelen van verschillende databestanden wordt door het CPB als veelbelovend beschouwd en de onderzoeksmogelijkheden zouden sterk uitgebreid kunnen worden door aanvullende gegevens te koppelen en te benutten. Als voorbeelden op het niveau van individuele leraren worden genoemd: ICALT-gegevens (lesobservaties) en IPTO-gegevens, waarmee door aanstellings- en roostergegevens van scholen onder andere bevoegdheden inzichtelijk gemaakt kunnen worden. Ook kan gedacht worden aan gegevens over leerlingen, zoals school- en examencijfers.

### 10.3 Kanttekeningen bij datakoppelingen

Het koppelen van leraargegevens wordt zoals gezegd als veelbelovend getypeerd vanwege de brede, vernieuwende analysemogelijkheden. Wel dienen vanuit onderzoeksperspectief de nodige kanttekeningen te worden geplaatst. We bespreken hier enkele door het CPB aangedragen aandachtspunten en beperkingen bij onderzoek op basis van datakoppelingen.

#### *Definities en selecties*

Allereerst dient bij onderzoek rekening gehouden te worden met uiteenlopende of juist beperkende definities van begrippen. Als voorbeeld wordt het begrip 'diplomarendement' genoemd. Een gangbare definitie van 'diplomarendement' in het hoger onderwijs is het afronden van een hbo- of wo-opleiding, ongeacht welke opleiding dit betreft.<sup>42</sup> Het CPB hanteert in 'Paden naar leraarschap' echter een scherpere definitie, namelijk het afronden van de opleiding waar men mee is gestart. Door dergelijke verschillen in

<sup>42</sup> Deze definitie wordt bijvoorbeeld gehanteerd door de Vereniging Hogescholen.

definities kunnen resultaten fors uiteenlopen. Zelfs wanneer definities wel consistent gehanteerd worden, kunnen resultaten verschillen doordat er altijd methodische keuzes en selecties gemaakt moeten worden in het onderzoeksproces, waarbij de details bij het programmeren in de syntax soms veel verschil kunnen uitmaken. Het zou dus niet alleen goed zijn om de definities helder te hebben, maar ook om keuzemomenten en selecties goed te documenteren en deze ook breed toegankelijk te maken, zodat er zoveel mogelijk consistentie ontstaat tussen onderzoeken, of zodat eventuele verschillen verklaard kunnen worden. Tegelijkertijd is het wel goed om maatwerk te kunnen leveren door, indien de onderzoeksdoelstelling dit vraagt, flexibel te kunnen zijn in definities en keuzes.

### *Administratieve data*

Registraties zijn bijzonder waardevol vanwege de (vaak) hoge kwaliteit en de hoge mate van integraliteit in termen van de populatieomvang. De administratieve gegevens in registraties zijn echter niet inhoudelijk integraal, of anders gezegd: niet alle informatie is daarin opgenomen.<sup>43</sup> Het formele karakter van de administratieve gegevens kan soms beperkingen met zich meebrengen, omdat er in de onderwijspraktijk talloze specifieke situaties en uitzonderingen mogelijk zijn. In de CPB-studie is bijvoorbeeld gebruik gemaakt van CBS-bestanden over (DUO-) onderwijsinschrijvingen, vooropleidingen en behaalde diploma's. Uit de praktijk is bekend dat er met name bij de tweedegraads lerarenopleidingen vaak mogelijkheden bestaan om een verkort opleidingstraject te volgen als studenten reeds beschikken over een hbo- of wo-diploma. Deze studenten worden ingeschreven bij de reguliere opleiding en krijgen op individuele basis een groot aantal vrijstellingen. Over deze specifieke vrijstellingen is in de registratiedata echter geen informatie beschikbaar. Hierdoor kan er in de analyses geen onderscheid gemaakt worden tussen studenten die de volledige opleiding volgen en studenten die een verkorte variant volgen.

Een ander voorbeeld waar het CPB tegenaan liep, heeft betrekking op zij-instromers. Zij hebben (in eerste instantie) vaak geen lerarenopleiding gevolgd (of niet voor het benodigde vak). Daardoor komen ze niet voor in het gekoppelde bestand van inschrijvingen en diploma's dat door het CPB gebruikt wordt om studenten aan de diverse lerarenopleidingen te identificeren. Zij-instromers zijn moeilijk als dusdanig te identificeren: alleen zij-instromers waarvoor zij-instroomsubsidie is aangevraagd zijn op die manier bij DUO bekend. In 'Paden naar leraarschap' heeft het CPB derhalve alleen analyses op de specifieke groep zij-instromers kunnen verrichten die een zij-instroomsubsidie hebben aangevraagd.

---

<sup>43</sup> Zo zijn bijvoorbeeld educatieve minoren in de door het CBS-beschikbaar gestelde onderwijsbestanden niet opgenomen.

Een andere belemmering bij onderzoek op basis van administratieve gegevens betreft de vertragingfactor. Dergelijke gegevens kunnen vaak pas na een jaar of langer beschikbaar gesteld worden, waardoor analyses altijd betrekking hebben op niet-actuele data. De updatefrequentie verschilt van bestand tot bestand, en de vertragingfactor van de database als geheel wordt in principe bepaald door het te koppelen administratieve bestand dat het minst actueel is.

### *Onderzoek in de CBS-omgeving*

Toegang tot de CBS-microdata kan alleen plaatsvinden door geautoriseerde gebruikers (met een geautoriseerd onderzoeksvoorstel) op het CBS zelf, of door geautoriseerde organisaties die beschikken over 'remote access' faciliteiten. De CBS-microdata zijn derhalve niet laagdrempelig en openbaar toegankelijk. Sommige data worden overigens wel op geaggregeerde niveaus beschikbaar gesteld via DANS.<sup>44</sup> Aan de analyses op CBS-microdata worden strenge voorwaarden gesteld. Zo is er strikt toezicht op het onthullingsrisico van privacygevoelige informatie, zowel op persoons- als op organisatieniveau. Dit betekent dat er bijvoorbeeld geen bevindingen over individuele scholen gepubliceerd mogen worden, tenzij deze scholen daar actieve toestemming voor geven. Het gebruik van CBS-data kent volgens het CPB dan ook zijn beperkingen, niet alleen vanwege de strenge voorwaarden en regels, maar ook vanwege de relatief hoge gebruikskosten.<sup>45</sup>

44 Data Archiving and Networked Services (DANS) is een instituut van het KNAW en NWO dat fungeert als online archiveringssysteem voor onderzoeksdata.

45 Zo zijn er onder meer kosten verbonden aan de installatie van remote access-faciliteiten, het autoriseren van een nieuwe gebruiker, het opstarten van een project binnen de CBS-omgeving en het controleren van output. Voor een overzicht van de tarieven, zie: [https://www.cbs.nl/-/media/cbs%20op%20maat/zelf%20onderzoek%20doen/2016fd37%20folder%20microdata\\_web.pdf](https://www.cbs.nl/-/media/cbs%20op%20maat/zelf%20onderzoek%20doen/2016fd37%20folder%20microdata_web.pdf).



## Conclusies

- Het CPB heeft in de studie 'Paden naar leraarschap' analyses verricht op administratieve bestanden over leraren c.q. over afgestudeerden aan lerarenopleidingen, afkomstig van verschillende bronnen (CBS, DUO, Platform Bètatechniek). Deze bestanden zijn binnen de CBS-microdataomgeving aan elkaar gekoppeld en zijn derhalve een interessant voorbeeld van het combineren van verschillende datasets over leraren. De studie is gericht op arbeidsmarktstromen van leraren en geeft onder meer inzicht in ontwikkelingen in studierendementen van leraren in opleiding, in baankenmerken van leraren en van afgestudeerde leraren die niet in het onderwijs werken, en in ontwikkelingen aangaande de zij-instromers.
- Het grote voordeel van analyses op basis van administratieve data is de hoge mate van bijzondering naar verschillende detailniveaus. Zo kunnen uitspraken gedaan worden op het niveau van specifieke lerarenopleidingen en specifieke vakken. De combinatie van CBS- en DUO-bestanden maakt het onder meer mogelijk om de functie 'leraar' te herkennen in de baanbestanden van het CBS. In de administratieve bestanden die het CBS beschikbaar stelt kunnen namelijk alleen sectoren onderscheiden worden, geen functies. De CPB-studie laat daarmee zien dat de analysemogelijkheden door datakoppelingen sterk uitgebreid worden waardoor nieuwe onderzoeksvragen beantwoord kunnen worden die anders vermoedelijk onbeantwoord zouden blijven. Voor de toekomst vindt het CPB het dan ook interessant om bijvoorbeeld lesobservatiegegevens over leraren en IPTO-gegevens te gaan benutten.
- Het werken met microdata in een CBS-omgeving kent naast grote voordelen (bijvoorbeeld de mogelijkheid om persoonsgegevens te koppelen en externe data aan te leveren) ook zijn kanttekeningen. Zo kunnen definities van onderzoek tot onderzoek uiteenlopen waardoor analyses op dezelfde bestanden verschillende resultaten kunnen opleveren. Heldere, eenduidige definities zouden daar bij kunnen helpen, hoewel de flexibiliteit van definities (naar gelang de onderzoeksvraag) ook mogelijk moeten zijn.
- Een beperking van onderzoek aan de hand van administratieve data is onder meer de algemeenheid en het administratieve karakter van de data (waardoor uitzonderingen op specifieke situaties vaak niet herkenbaar zijn) en het feit dat niet alles geregistreerd wordt. Verder hebben dergelijke bestanden te maken met een vertragingfactor: ze zijn niet erg actueel, vooral niet als het CBS ze aan meerdere administratieve bestanden moet koppelen. Tot slot zijn de voorwaarden streng (bijvoorbeeld wat betreft onthullingsrisico van individuen, of wat betreft de toegang tot remote access faciliteiten) en zijn de gebruikerskosten volgens het CPB fors.



# 11

## ONDERWIJSONDERZOEKERS: MOGELIJKHEDEN VOOR ONDERZOEK

### 11.1 Inleiding

Om meer zicht te krijgen op de (nieuwe) onderzoeksmogelijkheden die een database mogelijk kan bieden, is hierover gesproken met Dinand Webbink, hoogleraar 'Policy evaluation, with specific focus on Health, Education and Strategic Philanthropy' aan de Erasmus School of Economics van Erasmus Universiteit Rotterdam en Lex Borghans, hoogleraar 'Arbeidseconomie en Sociaal Beleid' aan de School of Business en Economics van Universiteit Maastricht.<sup>46</sup> Beide hoogleraren verrichten onderzoek naar de rol van leraren in de schoolse ontwikkeling van kinderen en jong volwassenen.

### 11.2 Visie op database

Beide onderzoekers vinden het idee van een database op zichzelf erg interessant vanuit onderzoeksperspectief. Volgens zowel Webbink als Borghans is er met name één centraal wetenschappelijk belang bij het tot stand brengen van een database over leraren: meer inzicht in de invloed van de leraar op de leerlingprestaties. Van belang is daarbij onder andere te onderzoeken welke persoons- en curriculumkenmerken van aankomende leraren invloed hebben op de leerlingen die ze later lesgeven. Er bestaan tot nog toe weinig harde wetenschappelijke onderzoeksresultaten over welke lerarenkenmerken er werkelijk toe doen voor de onderwijsprestaties van Nederlandse leerlingen. In de VS is hier al veel meer onderzoek naar verricht. Borghans noemt met name een studie van Rivkin, Hanushek en Kain<sup>47</sup> als voorbeeld van goed onderzoek naar de invloed van de leraar op leerlingprestaties.

<sup>46</sup> Daarnaast is binnen dit verkennend onderzoek onder meer gesproken met onderwijsonderzoekers Prof. Paulien Meijer, Prof. Klaas van Veen en Prof. Rolf van der Velden. In het gesprek met Prof. Meijer en Prof. Van Veen is vooral ingegaan op hun rol binnen de universitaire lerarenopleidingen (zie hoofdstuk 6). Met Prof. Van der Velden is vooral gesproken over zijn ervaringen binnen het Nationaal Cohortonderzoek Onderwijs (NCO), zie hoofdstuk 11.

<sup>47</sup> Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005), Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.

Dergelijk onderzoek zou voor de Nederlandse situatie ook moeten worden verricht, en een database kan daarbij een belangrijk rol spelen. Wel merkt Borghans op dat er in de VS veel meer centrale onderzoeken plaatsvinden naar de schoolprestaties van leerlingen dan in Nederland. In de VS leggen leerlingen jaarlijks een centraal examen af, in Nederland zijn er enkel centrale toetsgegevens voor groep 8 (CITO Eindtoets) en het eindexamenjaar van het vo (centraal eindexamen). Hierdoor worden de onderzoeksmogelijkheden naar de centrale invloed van leraren op leerlingprestaties sterk beperkt. Een alternatief voor een centrale database, dat volgens Borghans niet zou moeten worden uitgesloten, is om een database in te richten die slechts betrekking heeft op een (relatief kleine) representatieve groep van leraren. Op deze manier zouden veel rijkere data over leraren kunnen worden verzameld, met bijvoorbeeld evaluaties van leraren door schoolbesturen, evaluaties van leraren door leerlingen, evaluaties van leraren door collega's, evaluaties van leraren door ouders, inspectiedata, et cetera. Een tweede factor van wetenschappelijk belang voor een database over leraren die zowel door Borghans als Webbink werd genoemd is een groter inzicht in de factoren die een rol spelen bij lerarenmobiliteit, waaronder de selectiviteit in wie er met welke persoonskenmerken in het lerarenberoep in- en uitstromen.

### 11.3 Kritische noten bij de organisatie van de database

Bij het eventuele oprichten van een database moet er volgens Borghans met een aantal belangrijke punten rekening worden gehouden. Zo voorziet hij ten eerste mogelijke complicaties bij het centraal organiseren van een database. Des te meer partijen hierbij betrokken worden, des te meer compromissen er gesloten zullen moeten worden, des te 'saaier' de data volgens hem zullen worden. Hierdoor kunnen de mogelijkheden voor interessante onderzoeksvragen drastisch verkleind worden. De vraag is dan ook of de baten van een database wel voldoende opwegen tegen de ongetwijfeld hoge kosten die de inrichting van een database met zich mee zal brengen. Ten tweede voorziet hij mogelijke problemen bij een centrale rol van het CBS bij de totstandkoming van een database. Zijn ervaring met het CBS is dat het doorgaans erg lang duurt voordat je opgevraagde data van het CBS ontvangt en dat het bovendien altijd maar de vraag is of je wel echt de data krijgt die je nodig hebt. Bovendien zijn de kosten die het CBS in rekening brengt voor data doorgaans erg hoog. Het zou volgens hem daarom het beste zijn als de database door het NRO zelf zou worden beheerd.

## Conclusies

- Het idee van een database wordt door de onderwijsonderzoekers erg interessant bevonden. De belangrijkste meerwaarde die een database zou kunnen bieden, is het leveren van meer inzicht in de invloed van leraren op leerlingprestaties. Zo is het van belang meer te weten te komen over de mogelijke invloed van persoonskenmerken van leraren en over de opleiding die zij gevolgd hebben (curriculumkenmerken).
- Het ontbreekt volgens de onderzoekers tot nog toe aan 'harde', Nederlandse onderzoeksresultaten over welke leraaraspecten er toe doen voor de onderwijsprestaties. Naast onderzoek naar de invloed van leraren op de onderwijsprestaties van leraren, kan een database meer zicht bieden op mobiliteit van leraren op de arbeidsmarkt.
- Een alternatief voor een integrale database met leraren zou een kleine database kunnen zijn met een representatieve groep leraren. Dit maakt het door de kleinschaligheid eenvoudiger om bijvoorbeeld evaluaties van leraren door scholen, leerlingen, collega's, ouders, enzovoorts te benutten.
- Een risico bij het ontwikkelen van een database kan zijn dat er te veel partijen bij betrokken worden, waardoor er te veel compromissen gesloten moeten worden en de data onvoldoende 'interessante' gegevens zal bevatten.
- Ook kan het eventueel beleggen van een database bij het CBS mogelijke problemen opleveren, zo wordt aangegeven. Zo worden er kritische kanttekeningen geplaatst over de trage werking van het CBS en de hoge kosten die worden gerekend. Het oprichten van een database bij het NRO wordt daarvoor als eventueel te verkennen alternatief genoemd.



# 12

## NCO: DATAKOPPELINGEN EN HET VELD

### 12.1 Inleiding

In het afgelopen decennium heeft het grootschalig longitudinaal onderzoek COOL5-18 ('Cohortonderzoek onderwijsloopbanen') plaatsgevonden, waarbij ongeveer 20.000 leerlingen die instroomden in het po of vo gevolgd zijn in hun ontwikkeling. Onder meer vanwege de teruglopende respons door scholen is het onderzoek in 2015 geëvalueerd.<sup>48</sup> Het advies luidde om voor een andere opzet te kiezen, waarbij de basis gevormd wordt door registergegevens over onderwijsloopbanen van leerlingen, aangevuld met onder andere gegevens over cognitieve prestaties en sociaal-economische achtergrondgegevens. Naam van deze nieuwe opzet: het Nationaal Cohortonderzoek Onderwijs, ofwel NCO.

In september 2015 zijn Prof. Rolf van der Velden en Dr. Mark Levels aangesteld als kwartiermakers om de technische, juridische en organisatorische haalbaarheid van de voorgestelde opzet te toetsen. In deze kwartiermakersfase hebben de onderzoekers gesproken met een groot aantal partijen, zoals OCW, CBS, Inspectie van het Onderwijs, DUO, CITO, NRO, PO-Raad, VO-Raad, diverse schoolbesturen in po en vo, en juristen. Van der Velden en Levels concluderen dat de gekozen opzet voor het NCO technisch, juridisch en organisatorisch haalbaar is. In deze opzet speelt het CBS op twee manieren een centrale rol. Op de eerste plaats betreft dat de koppeling van BRON-data<sup>49</sup> aan andere registerdata van het CBS en op de tweede plaats gaat het over de rol van het CBS als 'entrusted third party' bij de koppeling van gegevens uit de leerlingvolgsystemen<sup>50</sup> aan de onderwijsloopbanen zoals verkregen uit BRON. Deze centrale rol van het CBS heeft directe consequenties voor de organisatorische vormgeving van het NCO en betekent dat het CBS een belangrijke speler is in de verdere ontwikkeling van het NCO.

<sup>48</sup> Van der Velden & Van der Maas (2015), *Naar een toekomstbestendig cohortonderzoek*, Maastricht: ROA.

<sup>49</sup> BRON staat voor Basisregister Onderwijs en wordt beheerd door DUO. Alle leerlingen en studenten staan in BRON, van po tot wo. BRON bevat gegevens over de in- en uitschrijving van leerlingen, en over examens en diploma's. Zie ook: <https://www.rijksoverheid.nl/contact/contactgids/bron-primair-onderwijs>

<sup>50</sup> Hiermee houden scholen vorderingen en resultaten van leerlingen en groepen bij. Dit gebeurt in ieder geval voor Nederlandse taal en rekenen-wiskunde. Zie ook: <https://www.rijksoverheid.nl/onderwerpen/basionderwijs/vraag-en-antwoord/hoe-legt-de-basisschool-de-prestaties-van-mijn-kind-vast>

In mei 2016 is het NCO officieel van start gegaan met Rolf van der Velden en Mark Levels als Nationaal Coördinatoren. Er is sindsdien gestart met het daadwerkelijk koppelen van gegevens uit BRON aan achtergrondgegevens van het CBS. Inmiddels zijn de eerste schoolrapportages opgesteld. Vanwege de inhoudelijke parallellen tussen het NCO en een mogelijke database over (startende) leraren en lerarenopleidingen, en de ervaring die vanuit het NCO is opgedaan in de stadia voorafgaand aan de officiële start, bespreken we hier enkele leerpunten.

## 12.2 Basisvoorwaarden: draagvlak, focus en realisme

In de voorliggende rapportage signaleren we een zekere terughoudendheid in het onderwijsveld, met name bij de (hbo-)lerarenopleidingen, wat betreft het ontsluiten en koppelen van gegevens. Ook in het NCO wordt deze terughoudendheid bij scholen in het po en vo ervaren. Daardoor moe(s)ten er de nodige barrières beslecht worden voor het welslagen van onderzoek gebaseerd op koppelingen van bestanden die door scholen dienen te worden aangeleverd. Vanuit de ervaringen van het NCO kan een aantal aspecten worden onderscheiden die van belang zijn om draagvlak voor dergelijk onderzoek te creëren.

Het is allereerst essentieel dat het veld overtuigd wordt van het *belang en het nut*. Het is hierbij noodzakelijk om aan te kunnen geven welke concrete (onderzoeks)vragen met de data beantwoord kunnen worden, waarom het zinvol is om die vragen te beantwoorden en op welke manier de belasting van het veld verminderd kan worden. Hierbij is een goede *communicatie(strategie)* zeer belangrijk. Het kan lonen om hiervoor professionele ondersteuning te vragen. Zo is voor het NCO een communicatiebureau ingeschakeld dat helpt om een duidelijk, sterk verhaal naar buiten te brengen.

*Breng focus aan.* In de loop van het NCO-project is steeds beter afgebakend wat wel en wat niet meegenomen zal worden in het onderzoek. De doelstelling moet haalbaar/realistisch zijn. Het is bijvoorbeeld onverstandig om in te zetten op 'gevoelige' gegevens (denk bijvoorbeeld aan gezondheidsdata), zeker in de beginfasen. In het verlengde hiervan is het belangrijk om een *stapsgewijze opbouw* te hanteren. Te grote plannen in één keer zijn intimiderend en het koppelen van bestanden zonder achterliggende onderzoeksvragen zijn weinig zinvol. Een betere aanpak is stapsgewijs: van 'makkelijk' naar steeds moeilijker. Als de eerste fases lukken en zinvol blijken, dan kan verder gekeken worden naar wat in de nieuwe ontstane situatie nog meer interessant en haalbaar is.

Om goodwill te creëren en iets terug te mogen verwachten van (bijvoorbeeld) scholen of opleidingen is het belangrijk om *eerst iets geven en dan pas iets te nemen*. Door partijen in het veld iets 'tastbaars' te kunnen geven als voorbeeld van de zinvolle benutting van onderzoeksgegevens, groeit de relevantie en het draagvlak. Binnen het NCO krijgen scholen bijvoorbeeld (benchmark)rapporten over de doorstroom van hun leerlingen



die specifieker zijn dan de gegevens waarover ze al beschikken. Pas nadat ze iets (bruikbaar) ontvangen, kun/mag je ze vragen om ook iets te leveren. In het geval van NCO gaat het dan bijvoorbeeld om gegevens uit het leerlingvolgsysteem.

*Plan realistisch en evalueer doorlopend.* De opbouw van een database kost veel tijd en inspanning, niet alleen data-technisch (zoals het daadwerkelijke koppelen, analyseren, et cetera), maar zeker ook om het veld mee te krijgen. Stel daarom een realistische planning op en kijk in de eerste jaren naar de (soms kleine) concrete 'wins' die behaald zijn, en evalueer de opbrengsten om vervolgens nieuwe doelen te stellen. Heel veel kleine stapjes maken samen één grote stap.

### Conclusies

- Het 'meekrijgen' van het veld is misschien wel de grootste uitdaging bij onderzoek gebaseerd op gekoppelde data. Er bestaat vaak een grote terughoudendheid om gegevens te ontsluiten voor onderzoek, bijvoorbeeld omdat de meerwaarde van het onderzoek niet duidelijk is, of omdat er twijfels zijn over wat er met de data gebeurt. Dezelfde terughoudendheid van scholen die in de voorliggende rapportage wordt gesignaleerd, met name bij de (hbo-)lerarenopleidingen, wordt herkend door de betrokkenen bij het Nationaal Cohortonderzoek Onderwijs (NCO).
- Het NCO is inmiddels in uitvoering, maar heeft een lange verkenningsfase doorlopen waarin het nut, de mogelijkheden en de beperkingen van leerlingcohortonderzoek (gebaseerd op bestaande data) in kaart werd gebracht. Op basis van de ervaringen vanuit het NCO is een aantal elementen te benoemen die kunnen helpen om draagvlak te creëren in het veld:
  1. Overtuig het veld van het *belang en het nut* (concrete meerwaarde en opbrengsten) en zorg hierbij voor een *goede communicatie(strategie)*
  2. *Breng focus aan*: baken af wat wel en niet onderdeel zal zijn van het onderzoek (c.q. de database) en stel realistische doelstellingen.
  3. Hanteer een *stapsgewijze opbouw*: begin met 'laaghangend fruit' (data die makkelijk te verkrijgen zijn) en breid stap voor stap uit met 'moeilijkere' typen gegevens.
  4. *Eerst iets geven, dan iets te nemen*: lever 'tastbare' producten op waaruit het praktisch nut van de inspanningen blijkt. Pas dan kun men iets terugvragen van het veld.
  5. *Plan realistisch en evalueer doorlopend*: stel een realistische planning op en evalueer de opbrengsten om vervolgens nieuwe doelen te stellen.



# 13

## CONCLUSIES EN DISCUSSIE

### 13.1 Inleiding

In dit slothoofdstuk blikken we terug op de belangrijkste bevindingen van het onderzoek. Het centrale doel van het onderzoek was om de mogelijkheden te verkennen om gegevens over de kenmerken van leraren en de door hen doorlopen opleidingen op een meer systematische wijze te verzamelen dan nu het geval is. Om hier meer inzicht in te verkrijgen, zijn we nagegaan:

- Welke kennisbehoefte er bestaat bij de mogelijke afnemers uit de wetenschap, het beleid en het onderwijsveld en in hoeverre een database hieraan kan bijdragen;
- Welke onderzoeksbestanden over leraren (c.q. lerarenopleidingen) relevant zijn om op te nemen in een database;
- In hoeverre bestaande gegevens ontsloten en gekoppeld kunnen worden in een database en welke hindernissen daarbij overwonnen moeten worden.

In dit hoofdstuk gaan we allereerst in op de kennisbehoefte die er bestaat bij de mogelijke afnemers van de database. Vervolgens bespreken we welke databestanden mogelijk deel uit zouden kunnen maken van de database. Ten slotte komen enkele beperkingen van het huidige onderzoek aan bod en geven we suggesties voor toekomstig onderzoek.

### 13.2 Wenselijkheid en meerwaarde

#### *Koepels lerarenopleidingen kritisch over nut database*

De reacties op een database vanuit het veld zijn zeer divers. Het spectrum loopt van 'onnodig, niet zinvol en niet aansluitend op de kennisbehoefte' tot 'zeer interessant, bruikbaar en zinvol'. De lerarenopleidingen, in het bijzonder de lerarenopleidingen in het hbo (de pabo en de tweedegraads lerarenopleidingen), zitten vooral aan de sceptische zijde van het spectrum. De door ons geïnterviewde representanten van de koepelorganisaties LOBO en ADEF geven aan geen grote meerwaarde te zien in een database. De kwaliteit van de opleidingen wordt volgens hen geborgd door het accreditatiestelsel, en ook in de eigen kennisbehoeften wordt voor een belangrijk deel al voorzien. Zo wordt

de voortgang van studenten reeds doorlopend gemonitord en ook is er al veel kennis uit alumnistudies en contacten met het veld. De aanwezige kennis wordt daarbij volop gedeeld door opleidingen onderling. Daar waar opleidingen wél aanvullende kennisvragen hebben, lijkt er vooral interesse te bestaan voor contextspecifiek (bijvoorbeeld lokaal/regionaal) en themaspecifiek onderzoek: een generieke database wordt derhalve als weinig nuttig beschouwd. Onderzoek naar leraarkwaliteit via een database wordt bovendien als uitermate complex of zelfs vrijwel onmogelijk betiteld, waarbij de zorg bestaat dat er al snel naar de lerarenopleidingen 'wordt gewezen' als schuldige voor eventuele mindere 'kwaliteit' van bepaalde leraren.

Representanten van de universitaire lerarenopleidingen (ICL) zijn positiever over het idee van een database voor (startende) leraren en lerarenopleidingen. Een dergelijk onderzoeksbestand zou volgens hen bijvoorbeeld meer inzicht kunnen bieden in de mogelijke effecten van verschillen in (het curriculum van) lerarenopleidingen op de kwaliteit van leraren tijdens hun loopbaan. Ook wordt aangegeven dat een database meer zicht kan bieden in studenten- en arbeidsmarktstromen, bijvoorbeeld door onderzoek te verrichten naar de risicofactoren voor de uitval van (startende) leraren. De geïnterviewden binnen de ICL staan dan ook niet onwelwillend tegenover een database en eventuele toekomstige ontsluiting van studentgegevens. Wel zullen in dat geval goede afspraken moeten worden gemaakt over de wijze waarop dat zou gebeuren en over de privacybescherming.

#### *Onderwijscoöperatie: database mogelijk zinvol, mits...*

De Onderwijscoöperatie ziet een database in principe als potentieel nuttig instrument, bijvoorbeeld om onderzoek te verrichten naar de effecten van professionaliseringsactiviteiten op het onderwijs. De doelmatigheid van het scholingsaanbod is dan ook een belangrijk thema voor de Onderwijscoöperatie, en een database zou mogelijk bij kunnen dragen aan het aanbieden van een selectie van 'bewezen effectieve' professionaliseringsactiviteiten. Over onderzoek naar de kwaliteit van leraren is men wel sceptisch. Kwaliteit is lastig te definiëren en het is complex om aan te tonen waar kwaliteit door beïnvloed wordt. De Onderwijscoöperatie staat echter niet per definitie onwelwillend tegenover een database. Het is voor hen essentieel dat er goede, heldere onderzoeksvragen ten grondslag liggen aan de eventuele ontwikkeling van een database. Een database enkel 'omdat het kan' wordt als onwenselijk en weinig zinvol geacht. Mogelijke toekomstige benutting van gegevens uit het Lerarenregister – iets wat de komende jaren in ieder geval zeker niet haalbaar wordt geacht – zal alléén overwogen worden indien dit in goed overleg gaat met de achterban en onder strenge voorwaarden. Uitgangspunt is daarbij dat de leraar eigenaar is en blijft van zijn eigen gegevens en er actieve toestemming nodig is van leraren om data voor andere doeleinden te gebruiken dan waarvoor ze oorspronkelijk bedoeld zijn.

### *Positieve geluiden vanuit overheid en onderzoek*

De eerste reacties vanuit de kant van landelijk onderwijsbeleid, zoals OCW en de Inspectie van het Onderwijs, over een database voor (startende) leraren en lerarenopleidingen zijn positief. Ook vanuit onderzoekskant, zoals het CPB en onderwijsonderzoekers aan universiteiten, zijn er vooral positieve geluiden. Er worden weliswaar de nodige kanttekeningen geplaatst, bijvoorbeeld over het gebruik van registerbestanden vanwege de vertragingfactor (dergelijke bestanden zijn vaak niet erg actueel) en over soms moeizame ervaringen met het CBS, maar in algemene zin wordt een database als nuttig gezien. Meerwaarde van een database ligt vooral in de vermindering van de bevraginglast in het onderwijsveld, en verdiepende onderzoeksmogelijkheden. Zo kan vragenlijstonderzoek onder leraren bondiger worden doordat veel achtergrondvragen overbodig worden. Tevens kan, indien aangeschreven zou worden vanuit een database, veel meer focus aangebracht worden. Zo kunnen er specifieke doelgroepen geselecteerd worden waardoor de 'bijvangst' (dat wil zeggen respondenten die niet tot de eigenlijke onderzoeksdoelgroep behoren) geminimaliseerd kan worden. De CPB-studie 'Paden naar leraarschap' is hier een voorbeeld van.

Op meer onderzoeksinhoudelijk gebied kan een database nieuwe mogelijkheden bieden voor verdiepend onderzoek waarbij voor vele factoren 'gecorrigeerd' kan worden (denk aan achtergrondkenmerken, schoolkenmerken, enzovoorts). Ook biedt het mogelijkheden om de effectiviteit van beleidsmaatregelen beter in kaart te brengen, bijvoorbeeld onderzoek naar de lerarenbeurs, subsidie zij-instroom, opleiden in de school, aanscherping toelatingseisen pabo's, subsidieregeling begeleiding startende leraren, tegemoetkoming studiekosten lerarenopleidingen tekortvakken. Tevens kan een database zorgen voor minder 'versnippering' in de dataverzameling: op dit moment worden gegevens over leraren en opleidingen door veel verschillende instanties en onderzoeksinstellingen verzameld en worden deze data vaak niet centraal gedocumenteerd. Hierdoor is het bijvoorbeeld niet altijd duidelijk of en waar dergelijke data gevonden kunnen worden. Een centrale verzamelplek kan de toegankelijkheid van leraaronderzoek sterk vergroten. Een ander voordeel van een database kan zijn dat er gestandaardiseerde afbakeningen en definities voor onderzoek naar leraren en lerarenopleidingen opgesteld kunnen worden. In huidig onderzoek, onder meer op registerdata, worden selecties en definities vaak 'ad hoc' bepaald binnen het betreffende onderzoek. De vergelijkbaarheid van leraaronderzoek kan vergroot worden als een database wordt opgebouwd rondom bepaalde standaarddefinities. Welke partijen betrokken dienen te worden bij de afstemming van deze selecties en definities zal in de toekomst nader verkend moeten worden. Overigens wordt door diverse geïnterviewden benadrukt dat er altijd wel ruimte moet blijven om in specifieke gevallen beredeneerd af te kunnen wijken van een standaardafbakening.

## Overeenstemming over privacybescherming

Hoewel de meningen over het nut en de wenselijkheid van een database flink uiteenlopen, bestaat er wel grote overeenstemming dat dit in ieder geval bijzonder zorgvuldig en 'veilig' dient te gebeuren. In een dergelijke database dient de anonimiteit van leraren gewaarborgd te zijn en dienen onthullingsrisico's voorkomen te worden. Onthullingsrisico's kunnen ontstaan door specifieke selecties en koppelingen toe te passen, waardoor personen en/of scholen herkenbaar worden in rapportages. Een 'CBS-model' waarbij alleen geautoriseerde gebruikers toegang hebben tot onderzoeksgegevens, waarin personen en scholen versleuteld en anoniem in bestanden voorkomen en waarin scherpe eisen worden gesteld aan het voorkomen van onthullingsrisico's, lijkt derhalve essentieel voor een database van (startende) leraren en lerarenopleidingen. Omdat het CBS een unieke positie bekleedt in het mogen koppelen van persoonsgegevens met BSN, en strenge voorwaarden hanteert voor onderzoek op microdata, is de ontwikkeling van een database binnen het CBS een voor de hand liggende optie.

### 13.3 Onderliggende databestanden: haalbaarheid en ontwikkeling

#### *Database in CBS-omgeving lijkt haalbaar*

Naast de vraag in hoeverre een database wenselijk is en toegevoegde waarde biedt, staat de vraag in hoeverre een dergelijke database praktisch haalbaar is centraal in deze verkenning. Op basis van de gesprekken concluderen we dat er in beginsel geen onoverkomelijke juridische en technische beperkingen zijn voor het oprichten van een database van (startende) leraren en lerarenopleidingen. Het inrichten van een database binnen het CBS lijkt op dit moment het meest voor de hand liggend. Het CBS verzamelt, bewerkt en publiceert statistische gegevens over de Nederlandse samenleving (personen, bedrijven en instellingen) en mag deze ook koppelen aan andere persoonsgegevens in het kader van statistiek. Dit is wettelijk verankerd in de *Wet op het Centraal bureau voor de statistiek*.<sup>51</sup> Het CBS beschikt reeds over een grote hoeveelheid demografische, sociaal-culturele en sociaal-economische databestanden op individueel niveau ('microdata'), afkomstig vanuit administraties en surveys, en stelt deze voor een belangrijk deel beschikbaar voor onderzoek. Deze bestanden zijn onderling koppelbaar waardoor zeer rijke databases ontwikkeld kunnen worden. Wel dient er rekening gehouden te worden met een vertragsingsfactor die varieert naar type bestand, bijvoorbeeld van een aantal maanden tot een jaar of langer, en die kan toenemen bij de koppeling aan administratieve bestanden die met de meeste vertraging worden opgeleverd. De data worden periodiek geüpdatet.

<sup>51</sup> Het ten behoeve van statistische doeleinden gebruiken van het Burgerservicenummer (BSN) en verwerken van persoonsgegevens als bedoeld in artikel 16 van de *Wet bescherming persoonsgegevens* (WbP) is geregeld in de artikelen 34 en 35 van de CBS-wet.

In de administratieve CBS-bestanden is het mogelijk om werkenden in de onderwijssector te herkennen, maar is het beroep van leraar niet als dusdanig herkenbaar. Om een gerichte database te ontwikkelen zijn derhalve aanvullende, specifiekere gegevens benodigd. Het CBS biedt de mogelijkheid om externe bestanden te matchen aan de eigen microdata. Deze koppeling is het eenvoudigst en minst foutgevoelig wanneer het Burgerservicenummer (BSN) benut wordt, aangezien het BSN integraal is en uniek voor individuen. Het BSN mag echter alleen gebruikt worden door overheidsorganisaties om hun taak uit te voeren en door niet-overheidsorganisaties indien dit wettelijk bepaald is. In veel externe bestanden, zoals extern vragenlijstonderzoek, zal het BSN dus niet bekend zijn. Een alternatieve manier om externe bestanden te koppelen aan CBS-bestanden is op basis van persoonsgegevens. Het gaat dan in principe om de combinatie van geboortedatum, geslacht en het volledige woonadres. Overigens is het ook mogelijk om nieuwe 'externe' data te laten verzamelen via en in samenwerking met het CBS. Zo is het mogelijk om vragenlijstonderzoek te laten verrichten via het CBS, zoals momenteel bijvoorbeeld gebeurt voor de BVE-Monitor (afgestudeerdenonderzoek onder mbo'ers). Doordat er vanuit het CBS aangeschreven wordt, kunnen de resultaten eenvoudig gekoppeld worden aan andere CBS-bestanden. Wel dient het onderzoek aan te sluiten bij de doelen en taken van het CBS en zijn er vrij strenge voorwaarden aan verbonden. Zo heeft het CBS als volwaardige partner in het proces onder meer invloed op de inhoud van de vragenlijst en de procedure van steekproeftrekking, dient het CBS met de onderzoeksbevindingen naar buiten te mogen treden; en dient de samenwerkingspartner een goede reputatie te hebben.

#### *Potentiële kern database: CBS- plus DUO-gegevens*

In eerste instantie is DUO de meest voor de hand liggende 'leverancier' van aanvullende leraargegevens. DUO beschikt immers over persoonsgegevens, baangegevens, loongegevens en verlofgegevens van individuele leraren. Aangezien het in principe om administratieve gegevens gaat, kan de kwaliteit van de data als hoog worden beschouwd en is het bereik van de data zeer groot. Ondanks dat niet alle leraren in DUO-bestanden aanwezig zijn, bijvoorbeeld door onvolledigheid vanuit salarisadministraties of leraren die niet in dienst zijn bij scholen, leveren de DUO-gegevens een zeer grote hoeveelheid kennis op over werkzame leraren. Naast gegevens over leraren beschikt DUO over administratieve studentgegevens en daarmee dus over gegevens over toekomstige leraren en over de opleidingsachtergrond van huidige leraren. Verder beschikt DUO over gegevens op schoolniveau, zoals jaarrekeningcijfers, gegevens over de leerlingpopulatie, denominatie, et cetera.

Op dit moment worden diverse DUO-bestanden reeds beschikbaar gesteld als CBS-microdata, waardoor deze ook aan andere CBS-data gekoppeld kunnen worden. De onderwijspersoneelsgegevens worden op dit moment nog niet ontsloten via het CBS, maar de koppelmogelijkheden lijken in principe goed. Overigens zijn er op dit moment al initiatieven om onderwijspersoneelsgegevens van DUO (deels) beschikbaar te gaan stellen via het CBS. Gezien de relevantie van de DUO-leraargegevens, de relatief

hoge mate van integraliteit en het structurele karakter van de data, dient een database volgens ons in beginsel te bestaan uit een combinatie van registerdata van CBS en DUO. De CPB-studie 'Paden naar leraarschap' heeft al laten zien dat het benutten van deze verschillende bronnen praktisch haalbaar is en veel meerwaarde biedt. Wel dient ook voor de DUO-gegevens rekening gehouden te worden met een zekere vertraging-factor: de bestanden worden periodiek geüpdatet.

#### *Aanvullende bestanden: wisselend kansrijk*

Gezien de beschikbaarheid en grote mate van integraliteit van de bestaande registerbestanden is het voor de hand liggend om hiermee de basis van een database te leggen. Naast CBS- en DUO-bestanden zijn er echter vele typen gegevens die potentieel relevant kunnen zijn voor een dergelijke database. Een belangrijke bron zouden studentgegevens vanuit lerarenopleidingen kunnen zijn. Het gaat hier immers om de leraren van de toekomst. Voortgangsgegevens zoals toetsscores of stagebeoordelingen kunnen wellicht vroege indicatoren zijn van de latere kwaliteit van leraren, alsook van soorten arbeidsmarktstromen. Studentgegevens lijken derhalve een grote inhoudelijke relevantie te hebben, maar zijn zeker niet zonder meer te benutten.

Allereerst is het de vraag in hoeverre opleidingen en instellingen bereid zullen zijn om deze gegevens te ontsluiten en of dit juridisch haalbaar is in het kader van bescherming van persoonsgegevens. Vanuit de ervaringen uit het Nationaal Cohortonderzoek Onderwijs (NCO) blijken er op juridisch vlak geen grote belemmeringen te bestaan. In het NCO worden leerlinggegevens in de beveiligde omgeving van het CBS (dat daar als 'entrusted third party' optreedt) opgeslagen en bewerkt. Voortgangsgegevens over scholieren, afkomstig uit het leerlingvolgsysteem, kunnen benut worden als schoolbesturen daarvoor toestemming geven.<sup>52</sup> Wel wordt aanbevolen om de (ouders van) scholieren hierover te informeren. Een tweede potentiële moeilijkheid heeft betrekking op de technische kant van de ontsluiting: de mogelijkheden om studentgegevens vanuit een informatie/studentvolgsysteem naar een onderzoeksbestand te exporteren zal verschillen van instelling tot instelling. Dit is afhankelijk van het systeem dat gebruikt wordt en lijkt op basis van de interviews te variëren van 'een druk op een knop' tot 'een arbeidsintensieve, complexe klus'. Een derde belangrijk punt is tot slot dat curricula tussen instellingen kunnen verschillen. De vergelijkbaarheid van de data is daarmee mogelijk in het geding. Hierover bestaat onder de geïnterviewden echter discussie: sommige gesprekspartners veronderstellen een behoorlijke mate van vergelijkbaarheid, anderen juist een zeer beperkte. Om hier verdere uitspraken over te doen is vervolgonderzoek noodzakelijk. Het benutten van gegevens van lerarenopleiding voor een database lijkt op zijn minst een flinke uitdaging.

<sup>52</sup> Zie ook: <https://nationaalcohortonderzoek.nl/privacy/>



Ook de haalbaarheid van diverse andere typen bestanden zal per bron verschillen en noodzaakt verder onderzoek. Zo is het Lerarenregister inhoudelijk erg interessant voor een database, maar kunnen deze data in ieder geval in de komende jaren niet benut worden voor onderzoek. De datastructuur van het register leent zich hier nog niet voor en is hier ook niet voor bedoeld. Het benutten van data afkomstig van de Inspectie van het Onderwijs lijkt veel beter haalbaar; de Inspectie staat niet negatief tegenover ontsluiting. Aangezien de Inspectie-data in principe alleen op schoolniveau benut kan worden (en dus niet op het niveau van individuele leraren), is de inhoudelijke relevantie van deze gegevens voor een database van (startende) leraren en lerarenopleidingen in eerste instantie weer wat kleiner. Andere typen bestanden zoals vragenlijstdata uit bijvoorbeeld de Loopbaanmonitor, HBO-Monitor of Nationale Studenten Enquête kunnen zeker relevant zijn, maar zullen ook niet zonder meer te benutten zijn voor een database. Hiervoor zal in principe 'informed consent' noodzakelijk zijn: deelnemers dienen geïnformeerd te worden dat de gegevens benut kunnen worden voor ander onderzoek en men dient hiermee akkoord te gaan. Op voorhand is de benutting van dergelijke bestanden zeker een optie, maar de specifieke mogelijkheden en meerwaarde van dergelijke bestanden dient nader uitgezocht te worden. Dit geldt zeer zeker ook voor onderzoeksdata die verzameld worden in het kader van het Begeleiding Startende Leraren-project naar inductieprogramma's van starters in het onderwijs. Voor de bevraging van afgestudeerden in de Loopbaanmonitor, de HBO-monitor, de Nationale Alumni Enquête en Studie & Werk geldt dat een groot deel van de pas afgestudeerden van lerarenopleidingen in dezelfde periode twee of drie keer worden benaderd met een soortgelijke vragenlijst. Een betere afstemming, minder enquêtedruk, een hogere respons en kostenbesparingen zouden mogelijk moeten zijn.

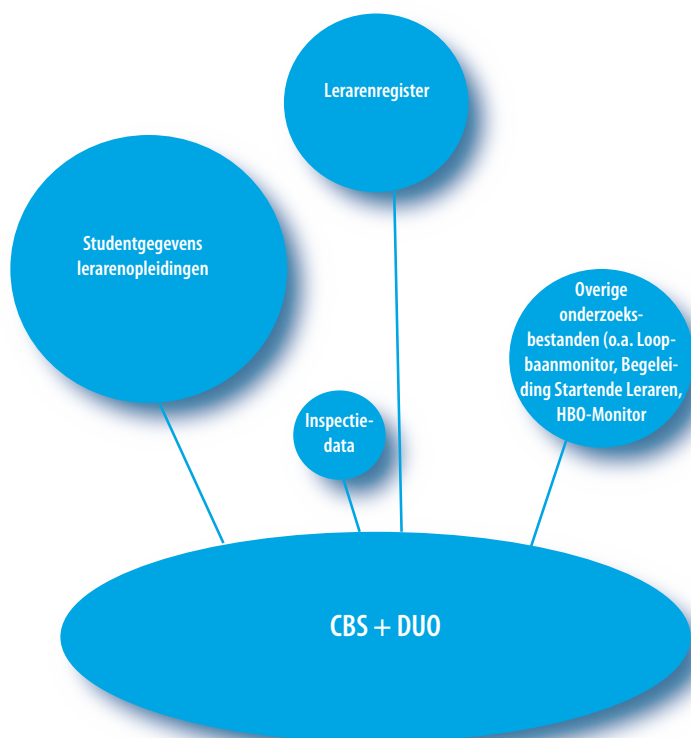
#### 13.4 Potentie en haalbaarheid verschillend per bestand

Als we de inschatting van de relevantie en haalbaarheid van een aantal in dit rapport verkende bestandstypen grafisch uitdrukken, volgt hieruit Figuur 13.1. In de figuur is de ingeschatte relevantie van de typen bestanden voor een database uitgedrukt in de omvang van de cirkel: een grote cirkel correspondeert met een hoog ingeschatte relevantie. Aangezien we uitgaan van een database gebaseerd op de kernbronnen CBS en DUO, staan deze aan de basis van de figuur. We veronderstellen daarnaast een grote inhoudelijke relevantie van met name de studentgegevens van lerarenopleidingen. De relevantie van Inspectie-data beschouwen we op dit moment als minder groot, aangezien het hier in principe niet om gegevens op individueel niveau gaat, maar enkel om gegevens op schoolniveau. De cirkel met Inspectie-data is derhalve relatief klein.

De lengte van de lijnen vanaf de 'basis' (CBS- en DUO-data) richting de andere bronnen indiceert de ingeschatte (voorlopige) haalbaarheid van de betreffende bron. Een lange lijn correspondeert met een naar verwachting lage praktische haalbaarheid en een grote benodigde inspanning om de bron te kunnen benutten. Het benutten van CBS- en DUO-bestanden lijkt naast inhoudelijk relevant ook relatief eenvoudig te zijn.

Deze bestanden vormen daarom, zoals eerder gezegd, de ‘basis’. Inspectiedata lijken eveneens praktisch haalbaar om te benutten voor een database, mits hierover goede afspraken worden gemaakt. Het benutten van studentgegevens van lerarenopleidingen of gegevens vanuit andere onderzoeksbestanden (zoals Loopbaanmonitor, Begeleiding Startende Leraren) zal vermoedelijk meer inspanning vergen. De lijn naar het Lerarenregister is het langst: volgens de Onderwijscoöperatie zullen deze gegevens gezien de structuur van het register in de komende jaren niet bruikbaar zijn om onderzoek op te verrichten.

**FIGUUR 13.1** Ingeschatte relevantie en haalbaarheid van bestandstypen voor database



### 13.5 Beperkingen in het huidige onderzoek

Er is een aantal kanttekeningen bij het huidige onderzoek te plaatsen. Ten eerste hebben we in dit onderzoek vooral aandacht gevestigd op grootschalige databestanden (vaak registerdata) en bekende databestanden uit langlopende onderzoeken van goede reputatie. Kleinschalige dataverzameling die incidenteel en ad hoc is verricht, valt buiten het

gezichtsveld van dit onderzoek, wat vooral een beperking is als dit innovatieve dataverzameling met allerhande nieuwe mogelijkheden zou betreffen.<sup>53</sup>

Een andere beperking die uit het voorgaande volgt is dat we vrijwel uitsluitend hebben gekeken naar individuele data over leraren en studenten, en veel minder naar de kenmerken van lerarenopleidingen en van ontvangende scholen. Over systematische gegevensverzameling met betrekking tot deze 'contextuele' factoren is niet veel bekend en deze dient nader onderzocht te worden.

Een volgende kanttekening betreft de lijst van geselecteerde betrokken en belanghebbende partijen. De interviews met deze partijen zijn gehouden met vertegenwoordigers van de leveranciers en gebruikers van de databestanden die wij als meest relevant hebben beoordeeld. Bovendien waren deze gesprekken vrij informeel. Hoewel de gespreksverslagen achteraf zijn goedgekeurd door de geïnterviewden, is het maar de vraag in hoeverre zij een formeel standpunt vertegenwoordigen van bijvoorbeeld het CBS of de Onderwijscoöperatie. De standpunten die zijn geformuleerd in de verschillende hoofdstukken van ons rapport kunnen dus persoonlijke inkleuringen bevatten, maar ook belangen van de betreffende organisaties die niet per se het algemeen belang vertegenwoordigen.

Een laatste belangrijke kanttekening bij ons onderzoek is dat we in onze gesprekken meer in algemene zin hebben gesproken over het belang van een database bestaande uit onderzoeksgegevens die inzicht bieden in de kwaliteit van (startende) leraren en lerarenopleidingen. Veel gesprekspartners vonden dit een te globaal vertrekpunt en vonden het opbouwen van een database voor (startende) leraren en lerarenopleidingen zonder specifiekere onderzoeksvragen niet gewenst.

### 13.6 Mogelijke vervolgstappen

#### *Duidelijke doelstelling van database*

Voordat verder wordt gegaan met de opbouw van een database over startende leraren en lerarenopleidingen dient duidelijk te zijn welke concrete vragen er beantwoord dienen te gaan worden met de gegevens. Welke onderzoeksvragen en indicatoren zijn voor lerarenopleidingen, scholen en beleidsmakers van belang? Daar is tot nu toe onvoldoende zicht op. De vragen zouden vanuit beleidsperspectief betrekking kunnen hebben op een regelmatige monitoring van indicatoren over bijvoorbeeld het doorlopen van de curricula in de lerarenopleidingen, de arbeidsmarktpositie van afgestudeerden en de intrede bij ontvangende scholen, het functioneren van startende leraren bij deze scholen en de professionaliseringsactiviteiten. Vanuit wetenschappelijk perspectief lijkt het er vooral op dat men wil weten wat een leraar tot een goede leraar

<sup>53</sup> Een voorbeeld hiervan is Leemans et al. (2017), *Personeel niet in loondienst (PNIL) in het po, vo en mbo*, Amsterdam: Regioplan/ROA.

maakt. Welke lerarenkenmerken doen er werkelijk toe voor de onderwijsprestaties van leerlingen? En: welke kenmerken van curricula van leraaropleidingen zijn vooral van invloed op de kwaliteit van (startende) leraren? Het spreekt vanzelf dat deze vragen ook voor het voeren van onderwijsbeleid en de inrichting van de lerarenopleidingen van groot belang zijn. Daarnaast hebben lerarenopleidingen en ontvangende scholen waarschijnlijk specifieke behoeftes als het gaat om de aanlevering van instellingsrapportages op basis van een leraren database. Het is de moeite waard uitgebreid met vertegenwoordigers van verschillende lerarenopleidingen en lerarenorganisaties te spreken over de dataverzameling en de informatiebehoefte.

Een belangrijk aandachtspunt, gerelateerd aan de functie van de database, is wellicht ook de wijze waarop de database van (startende) leraren en lerarenopleidingen wordt ingericht. In plaats van een complete database waarin alle mogelijke databestanden aan elkaar worden gekoppeld zou men ook kunnen denken aan een 'romp' van databestanden van CBS-, DUO- en eventueel IPTO-gegevens. Deze romp zou altijd zoveel mogelijk up-to-date beschikbaar moeten zijn, en snel te koppelen met bestaande databestanden als dat nodig wordt bevonden voor onderzoek op specifieke thema's op initiatief van beleidsmakers, wetenschappers of het onderwijsveld. Daarnaast zou de dataromp van dienst kunnen zijn bij het opzetten van nieuwe databestanden in samenwerking met de lerarenopleidingen, bijvoorbeeld over de studenten en startende leraren.

### *Draagvlak creëren*

In samenhang met het bespreken van de functie van de database, het formuleren van specifieke onderzoeksvragen en het vaststellen van de informatiebehoefte, kan er draagvlak gecreëerd worden onder de relevante partijen. Dit is vooral nodig bij de lerarenopleidingen. Hierbij kunnen de ervaringen met het Nationaal Cohort Onderzoek (NCO) van dienst zijn. Het uitvoeren van een 'charmeoffensief' om het onderwijsveld te informeren en mee te krijgen is waarschijnlijk van groot belang om de lerarenopleidingen bij de database te betrekken. Daarbij gelden voor de individuele lerarenopleidingen principes als 'Eerst iets geven, dan pas iets nemen'. Samen met een aantal voorlopers onder de lerarenopleidingen zou men aan het veld kunnen laten zien welke mogelijkheden er met een nieuwe database allemaal zijn om te voorzien in de verschillende informatiebehoefte. Pas daarna zou men de meer lastige dataverzameling bij de lerarenopleidingen moeten opstarten. Belangrijk zijn tevens het aanbrengen van focus, het formuleren van realistische doelstellingen, goed plannen en het doorlopend evalueren van opbrengsten. Het gaat dus om een stapsgewijze opbouw, waarbij wordt begonnen met 'laaghangend fruit' (data die gemakkelijk te verkrijgen zijn) en stap voor stap wordt uitgebreid met 'moeilijkere' typen gegevens.

Bij aanvang zou met een groep van voorlopers onder de lerarenopleidingen kunnen worden verkend in hoeverre administraties, systemen en opleidingsgegevens vergelijkbaar zijn tussen opleidingen en instellingen. Een eerste traject zou bijvoorbeeld kunnen eindigen in een aantal min of meer vergelijkbare instellingsrapportages met indicatoren die tegemoet komen aan de informatiebehoeftes van onderwijsinstellingen.

### *Kosten en gebruik*

Van belang is verder dat duidelijk wordt welke kosten er verbonden zijn aan de opzet van een database voor (startende) leraren en lerarenopleidingen. Er zou bijvoorbeeld een kostenindicatie moeten worden gegeven van het jaarlijks samenstellen van de eerder genoemde romp van de database door het CBS en DUO. Ook de procedure en de kosten van het koppelen van andere databestanden aan deze romp zouden wellicht kunnen worden gestandaardiseerd. Om het gebruik van een 'CBS-route' voor een database te stimuleren in plaats van een 'stand-alone' dataverzameling zou er goede bekendheid aan moeten worden gegeven, en wellicht enige subsidie. Bij maatwerk en dataverzameling voor ministeries of sectororganisaties zou als voorwaarde kunnen worden gesteld dat er zoveel mogelijk gebruik wordt gemaakt van de database om het afnemen van geheel nieuwe enquêtes af te remmen.

Door samenwerking met het CBS zou ook bespaard kunnen worden op de kosten van meervoudige monitoring in verschillende onderzoeken. Een bekend voorbeeld dat jaarlijks terugkomt, is de overlap tussen de Loopbaanmonitor, de HBO-monitor en ook Studie & Werk. Deze onderzoeken bevragen dezelfde groep afgestudeerden van tweedegraads lerarenopleidingen in ongeveer dezelfde periode over de aansluiting tussen opleiding en werk.

Op termijn zou tevens onderzocht moeten worden wat de mogelijkheden en kosten zijn van het koppelen van de database van (startende) leraren en lerarenopleidingen aan de NCO-database van leerlingen. Hiermee zou men niet alleen de beschikking kunnen krijgen over leraar-leerling-schoolgegevens, maar ook over bijvoorbeeld de toetsgegevens van leraren toen zij zelf nog in het primair en voortgezet onderwijs zaten.



# BIJLAGE I: OVERZICHT VAN GEÏNTERVIEWDEN

Organisatie	Naam	Functie <sup>54</sup>	Project / Onderdeel / Specificatie	Datum gesprek	Type gesprek
Algemeen Directeuren- overleg Educatieve Faculteiten (ADEF)	Diverse representanten	Directeuren lerarenopleidingen hbo	ADEF	6 oktober 2016	Face-to-face
CBS	Ivo Gorissen Marijke Hartgers	Relatiebeheerder Senior statistical researcher	Centrum voor Beleidsstatistiek	24 november 2016	Face-to- face
CentERdata	Peter Fontein	Hoofd afdeling Kwantitatieve Analyse	IPTO	12 oktober 2016	Telefonisch
CPB	Jonneke Bolhaar	Wetenschappelijk medewerker	Arbeid, Onderwijs en Pensioenen	19 januari 2017	Face-to-face
DUO	André Dickman Tris Serail	Onderzoeker Onderzoeker	DUO Informatie- producten	7 november 2016	Video-gesprek
Erasmus Universiteit Rotterdam (EUR)	Dinand Webbink	Hoogleraar Policy Evaluation	Erasmus School of Economics	7 april 2017	Telefonisch
Hogeschool van Arnhem en Nijmegen (HAN)	Koen Willems	Beleidsmedewerker Managementinfor- matie	Beleid	18 april 2017	Telefonisch
Inspectie van het Onderwijs	Harry Heemskerk Francien Berndszen	Afdelingshoofd Kennis Senior onderzoeker	Directie Kennis	25 januari 2017	Face-to-face
Interuniversitaire Commissie Lerarenopleidingen (ICL)	Paulien Meijer	Vice-voorzitter Wetenschappelijk directeur	ICL Radboud Docenten Academie	4 november 2016	Video-gesprek
	Klaas van Veen	Voorzitter Directeur	ICL Lerarenopleiding RUG	14 november 2016	Video-gesprek

<sup>54</sup> Naast de benoemde functies kunnen de geïnterviewden nog andere functies bekleden. We noemen hier de functies die in het kader van het gevoerde gesprekken het meest relevant zijn.

Landelijk Overleg Lerarenopleiding Basisonderwijs (LOBO)	Barbara de Kort	Voorzitter Hogeschoolhoofd-docent	LOBO Marnix College	21 november 2016	Video-gesprek
	Diverse representanten	Directeuren pabo's	LOBO / diverse instellingen	8 februari 2017	Face-to-face
Maastricht University (UM)	Lex Borghans	Hoogleraar Arbeidseconomie en Sociaal Beleid	School of Business and Economics	7 april 2017	Face-to-face
de Nieuwste Pabo (Fontys / Zuyd Hogeschool)	Henk van den Heuvel	Directeur	de Nieuwste Pabo	6 april 2017	Telefonisch
OCW	Thijs Noordzij	Adviseur	Directie Kennis	7 november 2016	Video-gesprek
	Marc van der Steeg	Adviseur	Directie Kennis	7 december 2016	Face-to-face
Onderwijscoöperatie	Frank Jansma	Adviseur	Lerarenregister	25 november 2016	Face-to-face
Zuyd Hogeschool	Paul Reijns Marcel Starmans	Beleidsadviseur Informatiespecialist	Beleid en Management-informatie	23 maart 2017	Face-to-face