



# Eindrapportage Pilot DCC Tijd voor toekomst: RDM voor onderzoek “Gelijke kansen voor een gezonde en kansrijke generatie” (202012P7)

<b>Auteur:</b>	Henk van der Zijden
28 juni 2022	evaluatie met Sanne Visser, Fokke Witteveen, Judith Pijnacker, Marleen Poot
29 augustus 2022	versie 0.9
31 augustus	commentaar verwerkt in versie 1.0

## Inhoudsopgave

<b>1</b>	<b>Inleiding.....</b>	<b>3</b>
<b>2</b>	<b>Beschrijving van het onderzoeksproject .....</b>	<b>4</b>
<b>3</b>	<b>Beoogd resultaat van de pilot DCC.....</b>	<b>5</b>
<b>4</b>	<b>Uitvoering Pilot.....</b>	<b>6</b>
4.1	Stand van zaken onderzoek per 10 juni 2022 .....	6
4.2	Bevindingen pilot RDM .....	6
4.2.1	Codeboek.....	6
4.2.2	Algemene Verordening gegevensverwerking (AVG) .....	6
4.2.3	Datamanagementplan .....	6
4.2.4	Datastewardship .....	7
4.2.5	Durf te vragen? .....	7
<b>5</b>	<b>Evaluatie pilot.....</b>	<b>8</b>
5.1	Conclusies .....	8
5.2	Aanbevelingen: .....	8

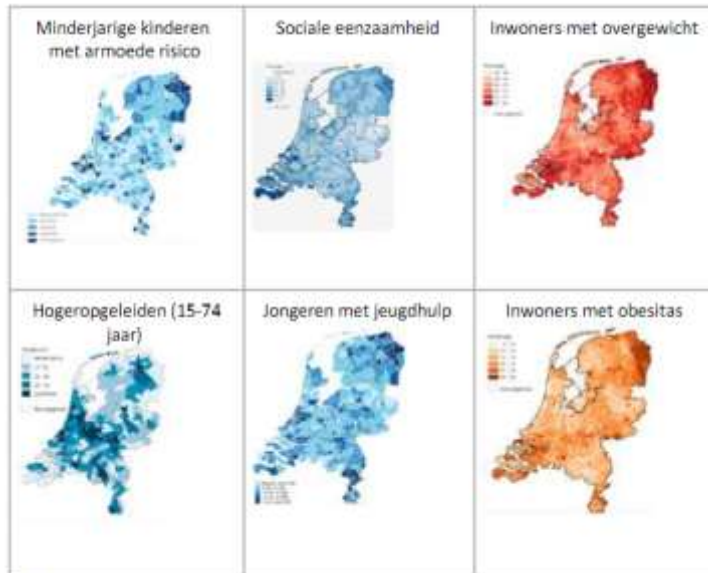
## 1 Inleiding

Van oktober 2021 tot en met juni 2022 is met behulp van SURF ondersteuning gegeven aan het onderzoeksdatamanagement van een meerjarig onderzoeksproject Tijd voor toekomst: “Gelijke kansen voor een gezonde en kansrijke generatie” (202012P7) van de Hanzehogeschool Groningen. Vanuit het Digital Competence Center voor praktijkgericht onderzoek (in oprichting) is in de vorm van subsidie en advies bijgedragen aan de opzet en uitvoering van het onderzoeksdatamanagement. In deze rapportage wordt verslag gedaan van de uitvoering van de pilot en worden conclusies getrokken en aanbevelingen gedaan.

In hoofdstuk 2 wordt het onderzoeksproject kort beschreven. In hoofdstuk 3 wordt de pilot DCC beschreven en in hoofdstuk 4 wordt gerapporteerd over de uitvoering van de pilot. Tenslotte worden in hoofdstuk 5 de pilot geëvalueerd, conclusies getrokken en aanbevelingen gedaan.

## 2 Beschrijving van het onderzoeksproject

Het betreft de ondersteuning van het datamanagement bij een groot meejarenonderzoek bij kinderen op verschillende scholen in de provincie. Doel van het onderzoeksprogramma is om de jongste generatie Groningers dezelfde kansen te bieden als de kinderen in de rest van Nederland. Dezelfde kans op het succesvol afronden van een opleiding, dezelfde kans op een leuke baan, dezelfde kans om gezond oud te worden. Gelijke kansen op een kanstrijk en gelukkig leven voor ieder Gronings kind. Uit onderstaande overzichten is te zien dat in Noord-Nederland kinderen die leven in Oost-Groningen opgroeien in relatieve armoede en kampen met gezondheidsrisico's als overgewicht.



Bron: CBS 2018

Resultaat en beoogde impact van het onderzoek

Data worden verzameld uit verschillende bronnen en geaggregeerd en er worden data verzameld door middel van onderzoek. Data uit secundaire bronnen:

-GGD: BMI, geboortedatum, postcode, motorische vaardigheid

-CITO: leerresultaten: rekenen, taal, maar ook bewegingsonderwijs Primaire dataverzameling (eigen onderzoek):

Een combinatie van:

-kwantitatieve data uit vragenlijsten (sportdeelname, welbevinden, scholing- en beroepskansen), behoefte, belemmeringen en mogelijkheden, en

-kwalitatieve data dmv focusgroepen, panels, foto's Er zitten sensitieve data-elementen inzoals gezondheid, maar mogelijk ook woonplaats. Het is de bedoeling om te gaan werken met pseudonimisering (gescheiden sleutelbestand).

### 3 Beoogd resultaat van de pilot DCC

Samen met Surf in een pilot toetsen of het dienstenportfolio van SURF bedoeld om praktijkgericht onderzoek te ondersteunen (in het bijzonder ResearchDrive en ResearchCloud) voldoet voor ondersteuning van dit complexe, uitdagende multidisciplinaire meerjarige data-intensieve samenwerkingsproject.

Daarvoor is het nodig dat de diensten (opgeleide mensen en ondersteunende middelen) ingericht, bekend en gebruikt kunnen worden. Daarbinnen ook de ontwikkeling en de organisatie van het datastewardship. Dit met als doel dat de onderzoekers binnen Tijd voor Toekomst op een veilige manier samen kunnen werken, maar ook geaggregeerde data kunnen delen met scholen, gemeente en derden. Idealiter is het zelfs mogelijk om met het instrumentarium (qualtrics/analyzer) rechtstreeks data op te slaan en vervolgens voor deelgebruikers beschikbaar te maken.

Beoogde impact: realisatie van een rijke bron van data die beschikbaar komt voor onderzoekers in het onderzoeksveld, maar ook voor iedereen FAIR toegankelijk bij publicaties van de onderliggende datasets in publieke repositories. Na evaluatie van de pilot worden de bevindingen gepubliceerd en komen zo voor het DCC-PO beschikbaar.

Mogelijk in te zetten diensten van SURF

Inzet van ResearchDrive als opslagtool voor onderzoeksdata en van relevante diensten uit de Surf ResearchCloud omgeving.

- Omdat het om samenwerking van verschillende hogescholen en universiteiten gaat wordt onderzocht of en op welke manier de SRAM-diensten van SURF hiervoor ingezet kunnen worden.
- We beginnen met de inzet van ResearchDrive. Daarvoor richten we de omgeving in ten behoeve van het project en beleggen rollen als datasteward en functioneel beheer bij onderzoeksondersteuners die deze rollen gaan vormgeven.
- Om met ResearchDrive goed te werken worden workshops/trainingen georganiseerd om de gebruikers goed te leren werken met deze omgeving.
- Er heeft een oriënterende bespreking plaatsgevonden met de accountmanager van Researchcloud I. Janmaat.
- Een aandachtspunt is het uiteindelijk FAIR krijgen van de verkregen onderzoeksdata zodat deze ook voor andere onderzoeksinstellingen en geïnteresseerden beschikbaar komen. Daarbij is de relatie tussen de werkomgeving en de te gebruiken publieke repositories zoals DATAVERSENL van belang.

Op te leveren resultaat: "Een beschrijving en evaluatie van de ingerichte datacyclus t.b.v. VRE voor het onderzoeksproject Tijd voor Toekomst."

## 4 Uitvoering Pilot

mei-juni 2021	Voorbereiding en indiening van de aanvraag voor uitvoering van een pilot DCC, betrokkenen Lector Remo Mombarg, Judith Pijnacker, Harm Wesseling, Marjan Bonink
oktober 2021	Afstemming organisatie pilot met Lector en hoofdonderzoeker
November 2021	definitieve toekenning van de subsidie en Kick-off pilot
December- juni 2022	uitvoering van het onderzoek en de pilot ondersteuning.
april 2022	presentatie van de voortgang aan alle deelnemende hogescholen in het DCC-PO (inclusief filmpje met Remo Mombarg)
Juni-juli 2022	evaluatie en eindrapportage
Augustus 2022	Einde pilot-project

### 4.1 Stand van zaken onderzoek per 10 juni 2022

- De data van de eerste meting zijn verwerkt (9 van de 10 scholen).
- Data zijn via Analyzer verzameld en opgeslagen op Researchdrive. Dit is nog niet geautomatiseerd.
- De invoerschermen zijn meer afgestemd op kinderen door te werken met smileys en één vraag per scherm en waar nodig taalkundige aanpassingen. Dit is gedocumenteerd.
- De data zijn met SPSS geanalyseerd. Sommige gehanteerde schalen zijn ivm goede verwerking aangepast. Syntax is vastgelegd in SPSS.
- Datasets zijn in Researchdrive ondergebracht.
- Op basis van de data zijn factsheets gemaakt in CANVA. Belangrijk goed te definiëren wat je wil presenteren.
- Er wordt een volgende meting voorbereid die na de zomer 2022 gaat plaatsvinden
- De volledige datastroom wordt nog vastgelegd. Fokke ondersteunt hierbij.

### 4.2 Bevindingen pilot RDM

#### 4.2.1 Codeboek

Advies informatiespecialist onderzoek om een codeboek te gaan gebruiken:

Om een onderzoek gedurende en achteraf goed te kunnen begrijpen is het gebruik van een codeboek aan te bevelen. Hierin leg je tenminste de volgende zaken vast:

- methodiek en definities,
- alle variabelen en hoe deze worden gemeten,
- wijzigingen gedurende het onderzoek.

De Informatiespecialist onderzoek biedt ondersteuning, er wordt een afspraak gemaakt.

#### 4.2.2 Algemene Verordening gegevensverwerking (AVG)

- Goed inregelen van het proces “verzamen en onderhouden van de ondertekende instemmingsverklaringen” is arbeidsintensief met de nodige risico’s.
- De instemming voor het onderzoek wordt door de deelnemende school aan de ouders van de kinderen gevraagd.
- Er zijn papieren formulieren die worden gescand en in Researchdrive ondergebracht.
- Sommige scholen registreren dit in het leerling volgsysteem. In die gevallen dient er op gelet te worden dat deze lang genoeg worden bewaard (zolang als de toestemming aangetoond moet kunnen worden voor het project. Als dit niet het geval is dienen er kopieën hiervan in het projectdossier te worden opgeslagen.
- Zorgen voor voldoen aan de Privacy en security voorwaarden: pseudonimisering/anonimisering.

#### 4.2.3 Datamanagementplan

- Er wordt gewerkt volgens het DMP. Actualisering van het DMP kan nuttig zijn. De informatiespecialist onderzoek helpt hierbij.
- Opslagstructuur: Aanbevolen en gebruikt: SURF researchdrive, Sharepoint community; SURF researchcloud (niet gebruikt); afgeraden: privé-opslag: Onedrive. ICT- Adviseur onderzoek, functioneel beheerder ondersteunt.

- Onderzoekssoftware: Analyzer, Excel, SPSS, Statistische programmeertaal R, etc. ICT- Adviseur onderzoek, functioneel beheerder ondersteunt.

#### 4.2.4 Datastewardship

Invulling door de hoofdonderzoeker: Deze vervult nu vrijwel alle taken van een datasteward op domein niveau. Aan de hand van de 'data life cycle': onderzoeksopzet, DMP, uitvoeren onderzoek door verzamelen data, analyseren, bewerken, opslaan versies datasets, delen met andere onderzoekers, publiceren, archiveren. Rapportages worden handmatig gemaakt in CANVA.

Invulling door Informatiespecialist onderzoekdatastewardship:

- Procesondersteuning voor het maken van een DMP.
- Advisering over wat er aan best practices beschikbaar is.
- Advisering datamanagementplan.
- Advisering over voldoen aan AVG.
- Advisering gebruik van codeboek.

#### 4.2.5 Durf te vragen?

Vanuit de rol die SURF vervult in het ontwikkelen van een Digital competencecenter Praktijkgericht Onderzoek is ondersteuning en begeleiding gegeven aan de Pilot.

- Eerder verricht onderzoek waarbij peutermonden in kaart zijn gebracht en kunnen worden gevolgd. Hiervoor zijn vragenlijsten geprogrammeerd in een open source taal. Onderzocht is of deze code hergebruikt zou kunnen worden. Daarvoor is via het DCC praktijkgericht onderzoek hulp gevraagd. Het DCC van de RUG heeft naar de opzet van het onderzoek "tijd voor Toekomst" gekeken en nagegaan of hergebruik van de code mogelijk was.
- De code bleek eerst nog lastig te vinden (in Github), maar al gauw werd duidelijk dat het efficiënter zou zijn om de relatief eenvoudige vragenlijsten opnieuw te programmeren. Door een ontwikkelaar van de RUG is in korte tijd een prototype klaargezet.
- Nadat tijdens het ontwerpproces van de vragenlijsten bleek dat de vragenlijsten ook door kinderen zelf ingevuld moesten gaan worden op een Ipad is de doorontwikkeling gestaakt.
- Ook omdat voor het beheer en de doorontwikkeling van de code geen kennis binnen de Hanzehogeschool beschikbaar is en de code dus altijd door een derde partij onderhouden zou moeten gaan worden.
- Na bespreking van de resultaten van het ontwikkeltraject is nog gekeken naar alternatieve software zoals Mentimeter, maar ook naar mogelijkheden om binnen Analyzer de invoerschermen kindvriendelijker te maken. Conclusie was dat Analyzer goede mogelijkheden biedt door het gebruik van Smileys en door slechts één vraag per scherm te presenteren.

## 5 Evaluatie pilot

Wat kan beter?

- Borgen continuïteit: gedurende het project en nadat het project is afgerond. Door tijdelijke onbeschikbaarheid van de hoofdonderzoeker ontstonden enige afstemmingsproblemen als gevolg van overdracht naar twee junior onderzoekers.
- Waar ligt de verantwoordelijkheid: binnen het lectoraat: hoofdonderzoeker bespreekt dit met de lector of binnen het kenniscentrum, bespreekpunt lector met management COE-HA.
- Tijdelijke wisseling in bezetting onderzoeksteam leverde uitdagingen op.
- intensiever contact ondersteuning en onderzoeker leidt tot betere resultaten
- Programmeerkennis kan via DCC-PO beschikbaar komen voor onderzoekers.
- Beheer van de ontwikkelde tooling is op dit moment niet vanuit SB-INF als dienst beschikbaar.

Wat ging goed?

- DMP en ondersteuning helpen bij het structureren en beter organiseren van het onderzoek en leveren meerwaarde voor onderzoek.
- De wisselwerking met ondersteuning leidt tot betere opzet en uitvoering van het onderzoek en tot meer inzicht bij ondersteuners in wat nodig is.
- De technische ondersteuning bij Researchdrive en gebruik van tools heeft goed geholpen
- Functioneel beheerders kunnen ondersteuning bieden bij beter gebruik van de tools.

### 5.1 Conclusies

De pilot heeft bijgedragen aan een intensiever contact tussen onderzoek en ondersteuning waardoor inzicht ontstond in wat er allemaal komt kijken bij goed datamanagement en waardoor uitvoering en organisatie van het datamanagement zichtbaar beter is geworden. Dit heeft voor beide kanten positief gewerkt.

De beschikbare ondersteuning en tooling voor dit onderzoek vanuit SB-informatisering voldoet voor dit project. De kennis hiervan bij het project en bij de ondersteuning is aanmerkelijk toegenomen. De mogelijkheden van Analyzer voor dataverzameling, de kennis van de beschikbare analysetools Excel en SPSS is vergroot. Het gebruik van SURF Researchdrive heeft de meerwaarde aangetoond als goede omgeving voor het beheren van onderzoeksdata gedurende het onderzoek. Er is niet toegekomen aan het toetsen of SRAM-diensten van Surf voldoen omdat de focus in deze periode lag op data die op Noordelijke basisscholen worden verzameld waar geen andere kennisinstelling bij betrokken was. Ook is nog geen praktische invulling gegeven aan het FAIR publiceren van datasets.

De keuze om de eerste lichting vragenlijsten op papier uit te delen en daarna via Analyzer te verwerken is arbeidsintensief en geeft kans op invoerfouten. Consistent gebruik van begrippen en schalen is cruciaal om de resultaten over een langere periode te kunnen volgen en analyseren.

Een goede voorbereiding van data projecten, waarbij persoonsgevoelige data worden verzameld blijft de aandacht vragen:

1. De omgang met de instemmingsformulieren van de (ouders) van de kinderen vergt het nodige organisatie-talent van de onderzoeker.
2. De manier waarop anonimiseren/ pseudonimiseren wordt ingericht luistert erg nauw, als je kinderen over meerdere scholen, meerjarig wilt kunnen volgen.

### 5.2 Aanbevelingen:

1. Denk vanaf het begin van een onderzoeksproject na over het datamanagement gedurende de gehele datalifecycle. Het Datamanagementplan kan hier goed bij helpen.
2. Bepaal welke ondersteuning nodig is. Denk daarbij expliciet na over welke taken op het gebied van datastewardship relevant zijn in welke fase van het onderzoek en organiseer dat deze taken ook ingevuld gaan worden. Maak onderscheid in domeinspecifieke taken en generieke taken. Bepaal wie binnen de onderzoeksgroep dan wel het kenniscentrum de domeinspecifieke taken gaat uitvoeren.
3. Stem de uitvoering van het Datamanagementplan regelmatig af met de ondersteuning en bepaal per onderzoeksfase welke ondersteuning en tools nodig zijn. Dit kan veranderen gedurende de uitvoering van het onderzoek. Betrek ook de adviseur onderzoekstooling hier nadrukkelijk bij.
4. Neem in het onderzoeksplan tijd en capaciteit op voor uitvoering datasteward taken en ICT-taken en budget voor het kunnen gebruiken van de gewenste software.
5. Maak zoveel mogelijk gebruik van de door Stafbureau Informatisering en SURF aangeboden (standaard)diensten voor ondersteuning van het onderzoek.
6. Er is nog ontwikkeling mogelijk door gebruik te gaan maken van de statistische programmeer taal R en de visualisatie mogelijkheden die R-Vision bieden.