

Whitepaper

Regionale Datawerkplaats

Een samenvatting van de wensen van potentiële gebruikers van een regionale datawerkplaats



Datum:

6 december 2022

Door:

Werkgroep GERDA (Ilona Oude Nijhuis, Lotte van Dieren, Maarten den Braber en Teun Koldewey)

Als onderdeel van het werkpakket Data science & Analytics.

Met dank aan:

Anne-Claire Joon (data scientist Fundis Holding / Zoetermeer 2025), Caroline Timmerman (onderzoeker, GGD Noord- en Oost-Gelderland), Daniel Kapitan (freelance data scientist), Danielle Bekkering (Healthy Ageing Network Noord-Nederland), Dave Hamersma (data scientist, Sensire), Dennis Diederickx (beleidsmaker Maasstad Ziekenhuis), Els Roorda (data scientist Q-Consult Zorg, PhD Population Data Infrastructure LUMC), Esther Huisman (epidemioloog, GGD Drenthe), Henriëtte Hoeven (onderzoeker, GGD Noord- en Oost-Gelderland), Jan-Willem Tiemessen (BI & RPA, Sensire), Jochen Mierau (wetenschappelijk directeur, Lifelines), Jori Hoendervanger (data scientist, Zilveren Kruis), Margriet Jonker (beleidsmaker, GGD Drenthe), Mark Callaars (informatiemanager Raedelijk), Marieke Oskam (Veluwe Gezond, Raedelijk), Michiel Brinksma (kwartiermaker data, GGD Drenthe), Noortje Wolters (beleidsmaker, Veiligheidsregio Noord- en Oost-Gelderland), Noortje de Haan (eigenaar adviesbureau Bits of Care), Peter Paul Essers (projectleider, Zeeuwse Zorg Coalitie), Raun van Ooijen (assistent professor UMCG, lead DataPoort), Rogier Kessel (data scientist, Amsterdam Data Collective), Ronald Wierstra (regiomanager gemeenten, De Friesland Zorgverzekeraar), Saskia Borg (Veluwe Gezond, Equalis afdeling Strategy & Modelling), Shaen Hazenbrink (BI, Slingeland Ziekenhuis), Shirley Kalkers (Specialist Informatievoorziening, GGNet), Teun Koldewij (data analist, Marga Klompé), Thomas Hengreen (data scientist, Amsterdam Data Collective), Valentin Neevel (Beleidsadviseur ministerie VWS), Yvanka Klein Holte (verpleegkundig onderzoeker, Sensire)

Voor:

Domeinoverstijgende samenwerkingsverbanden in (andere) regio's, met name werkgroepen die zich bezig houden met data(analyse).

Over dit document

Het staat de lezer vrij om dit document te delen met vermelding van bron. We stellen het op prijs als je ons even laat weten als je dit document gebruikt en wat jouw inzichten zijn.

Inhoudsopgave

1 Aanleiding	2
2 Methode	3
3 Scope regionale datawerkplaats	4
3.1 Gebruikers	4
3.2 Soort onderzoek	6
3.3 Onderzoeksvragen	7
4 Benodigde data in de datawerkplaats	10
4.1 Databronnen	10
4.2 Breedte/historie	11
4.3 Aggregatieniveau	11
4.4 Actualiteit	11
5 Functionaliteiten van de datawerkplaats	12
5.1 Tooling	13
5.2 Delen van inzichten	14
5.3 Output opties	15
6 Mogelijke obstakels en aandachtspunten	16
6.1 Wet- en regelgeving	16
6.2 Datakwaliteit en beschikbaarheid metadata	17
6.3 Governance	17
6.4 Visie, strategie en sturing	17
6.5 Techniek, tijd, middelen en competenties	18
6.6 Ethische en sociale aspecten	18
7. Conclusie	19

1 Aanleiding

In Nederland is het vertalen van gezondheidsdata naar inzichten en informatie een uitdaging door de wijze waarop de (preventieve) gezondheidszorg georganiseerd is. Zorgwetten zijn verdeeld over diverse domeinen, met een ander uitvoeringsniveau (landelijk of regionaal) en verschillende verantwoordelijke en/of uitvoerende partijen. Daarnaast zijn voor een integrale kijk op gezondheid niet alleen medische gegevens relevant, maar ook sociaal-economische gegevens of informatie over de leefomgeving moeten worden meegenomen.

Om te komen tot integrale gezondheidsbevordering is het nodig om inzichten op persoonsniveau over deze wetten, organisaties en disciplines heen inzichtelijk te maken, door domeinoverstijgend data bij elkaar te brengen. In de Achterhoek wordt vanuit 8RHK Gezond en Population Health Data NL een handreiking gedaan over de wijze waarop een geïntegreerde regionale data-infrastructuur voor gezondheid duurzaam kan worden ingericht.

Dit whitepaper volgt op het whitepaper '[Geïntegreerde Regionale Data-infrastructuur Achterhoek](#)', dat een drietal ontwerpkeuzes beschrijft om een keuze te maken uit een zestal technische mogelijkheden om domeinoverstijgend data samen te brengen. Dit whitepaper focust zich op de stap ná het samenbrengen van de data: de regionale datawerkplaats. De scope van een regionale werkplaats, de benodigde data in en functionaliteiten van een datawerkplaats, maar ook mogelijke obstakels voor de realisatie van een datawerkplaats worden in dit whitepaper beschreven. In dit whitepaper is gekozen voor een sterke interactie met de beoogde gebruikers van een regionale datawerkplaats en kent een sterk toegepaste insteek.

Voor het bepalen van de optimale inrichting van een regionale datawerkplaats hebben we in deze whitepaper onderstaand onderscheid gemaakt naar diverse soorten gebruikers.

1. **Inzicht-vragers**

De type 1 gebruiker heeft geen affiniteit met IT en maakt graag gebruik van standaard gedefinieerde rapportages en dashboards waarvan de kwaliteit geborgd is.

2. **Informatie-onderzoekers**

De informatie-onderzoeker ziet data als een bouw pakket. Data is vooraf geschoond, gedefinieerd en gemodelleerd, maar het gebruik is naar eigen inzicht en informatiebehoefte.

3. **Data scientists**

De type 3 gebruikers weet pas wat de informatiebehoefte is op het moment van het gebruik van data. Hierdoor is vooraf gedefinieerde data vaak onbruikbaar en is zoveel mogelijk onbewerkte, ongefilterde en zelf te bewerken data gewenst.

Informatie-onderzoekers en data scientists zijn de primair beoogde gebruikers van de regionale datawerkplaats (ten dienste van de inzicht-vragers) en staan daarmee centraal in deze whitepaper.

2 Methode

Om zicht te krijgen op gebruikersgroepen, hun onderzoeksvragen en de benodigde functionaliteiten van een regionale data-werkplaats, is gestart met een verkennend interview met een tweetal experts op het gebied van domein-overstijgend onderzoek. Dit heeft geleid tot een overzicht van relevante thema's en onderwerpen om rekening mee te houden bij de inrichting van een regionale werkplaats.

Input op deze thema's en onderwerpen is opgehaald in twee digitale gebruikerssessies, waarvoor informatie-onderzoekers en data-scientists uit de regio zijn uitgenodigd. Omdat domein-overstijgend onderzoek in de Achterhoek nu nog slechts beperkt mogelijk is (omdat de benodigde infrastructuur nog niet beschikbaar is) zijn ook kennishouders uit andere regio's en van (landelijke) instituten en organisaties aan de sessies toegevoegd.

Tijdens de (digitale) sessies werd GERDA kort geïntroduceerd, gevolgd door een aantal inspirerende voorbeelden van domein-overstijgend onderzoek. Vervolgens werden besproken (1) de scope, (2) de benodigde data en (3) de gewenste functionaliteiten van een regionale data-werkplaats. Na afloop van de sessies zijn alle deelnemers gevraagd om een vragenlijst over deze thema's (en bijbehorende onderwerpen) in te vullen. Daarnaast is dezelfde vragenlijst gedeeld met potentiële (toekomstige) gebruikers van een regionale data-werkplaats via LinkedIn. De resultaten van de sessies en de vragenlijst worden besproken in de volgende drie hoofdstukken van dit whitepaper.

3 Scope regionale datawerkplaats

Dit hoofdstuk beschrijft kenmerken en wensen van de gebruikers van de regionale data-werkplaats, het soort onderzoek dat ze doen en welke (regionale) domeinoverstijgende onderzoeksvragen er leven.

3.1 Gebruikers

Zoals in de aanleiding beschreven onderscheiden we 3 typen gebruikers van een regionale data-werkplaats. *Inzicht-vragers* zijn de grootste groep (in absolute aantallen, ten opzichte van onderzoekers en data-scientists). Zij hebben relatief weinig affiniteit met IT en willen voornamelijk gebruik maken van standaard rapporten en dashboards met kwalitatief hoogwaardige informatie. Zij worden hierin binnen hun organisaties gefaciliteerd door *informatie-onderzoekers* en *data scientists*. Omdat informatie-onderzoekers en data scientists de mensen zijn die hands-on met data aan de slag zijn in hun organisatie en de aangewezen personen zijn om met domeinoverstijgende data in de regionale datawerkplaats aan de slag gaan, richten we ons in dit whitepaper op deze twee groepen gebruikers. Hieronder worden kenmerken van deze twee typen gebruikers verder uitgewerkt.

Informatie-onderzoekers

De informatie-onderzoeker ziet data als een bouw pakket. Data is vooraf geschoond, gedefinieerd en gemodelleerd, maar het gebruik is naar eigen inzicht en informatiebehoefte.

- a. Niet tevreden met standaard rapportages, gebruikt data uit het systeem voor analyse
- b. Wil zelf data kunnen combineren
- c. Zoekt naar verklaringen voor trends en cijfers
- d. Groter wordende groep, verwachting is dat deze trend zich doorzet
- e. Aanvullende behoeftes t.o.v inzicht-vragers
 - i. Geïntegreerde data 'warehouses'
 - ii. Software om data te kunnen analyseren/zelf rapportages te kunnen maken zonder dat ICT of (uitgebreide) statistische kennis vereist is (intuïtief, flexibel en met standaard functies)
- f. Voorbeelden:
 - i. Super users met hun eigen afgebakende onderzoeksomgeving, zoals financieel adviseurs, innovatiemanagers en verpleegkundig informatiekundigen (VIK), CNIO/CMIO. Meer "data minded" - zouden in andere situaties bijv. Excel gebruiken.
 - ii. (Wetenschappelijk) onderzoekers, bijv. epidemiologen en beleidsonderzoekers. Veelal toetsen van vooraf opgestelde hypothesen, identificeren van afwijkingen en duiding geven/verklaren.

Data scientists

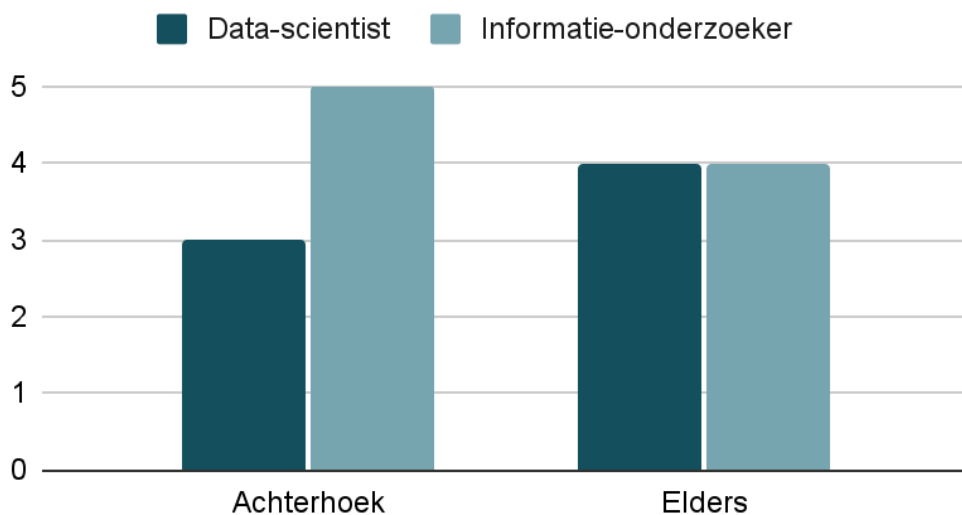
De data scientist werkt met beschikbare data om te kijken naar de (on)mogelijkheden rondom het beantwoorden van vragen op basis van beschikbare data. Hiervoor is vooraf gedefinieerde data vaak te beperkend, en is zoveel mogelijk onbewerkte, ongefilterde en zelf te bewerken data gewenst.

- a. Gebruikt grote hoeveelheden data, zonder vooraf exact te weten wat hij/zij tegen gaat komen

- b. Laat de data zelf met de antwoorden komen en ontdekt zo verrassende patronen
- c. Een groeiende groep
- d. Aanvullende behoeftes t.o.v informatie-onderzoekers
 - i. Niet gestructureerde informatie (bijvoorbeeld tekst en/of webdata)
 - ii. Zelf kunnen koppelen en verrijken van data
 - iii. Geavanceerde statistische data analyse en/of data science software/tools/scripts
- e. Voorbeelden:
 - i. Ontwikkelaars van modellen en algoritmes (bijv. voorspelmodel, risicomodel en/of clusteranalyse).
 - ii. Gebruikers van geavanceerde analyse-technieken (bijv. regressie-analyses etc.) en bredere datasets met meerdere indicatoren, zonder dat altijd van tevoren zeker is vast te stellen dat deze factoren van invloed zijn.
 - iii. Gebruikers die op zoek gaan naar mogelijke aanvullende (interne en/of externe) datasets, buiten de vooraf bekende datasets.

De door ons uitgezette vragenlijst over de scope, de benodigde data en de gewenste functionaliteiten in een (regionale) datawerkplaats werd ingevuld door 16 respondenten. Van deze respondenten werkt de helft voor een organisatie uit de regio Achterhoek. Voor wat betreft de verdeling naar soort gebruiker zijn er onder de respondenten iets meer informatie-onderzoekers (56%) dan data scientists (44%). Dit geldt met name voor de regio Achterhoek (62,5% informatie-onderzoeker; 37,5% data scientist), onder de respondenten van buiten de regio is de verdeling half om half (zie figuur 1).

Data-scientist en Informatie-onderzoeker

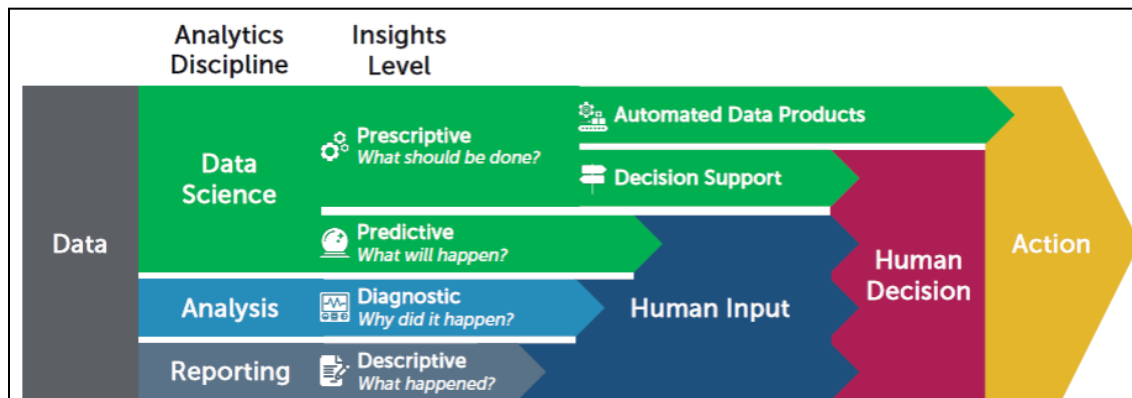


Figuur 1: verdeling naar soort gebruiker van respondenten uitgezette vragenlijst (regionale) datawerkplaats.

3.2 Soort onderzoek

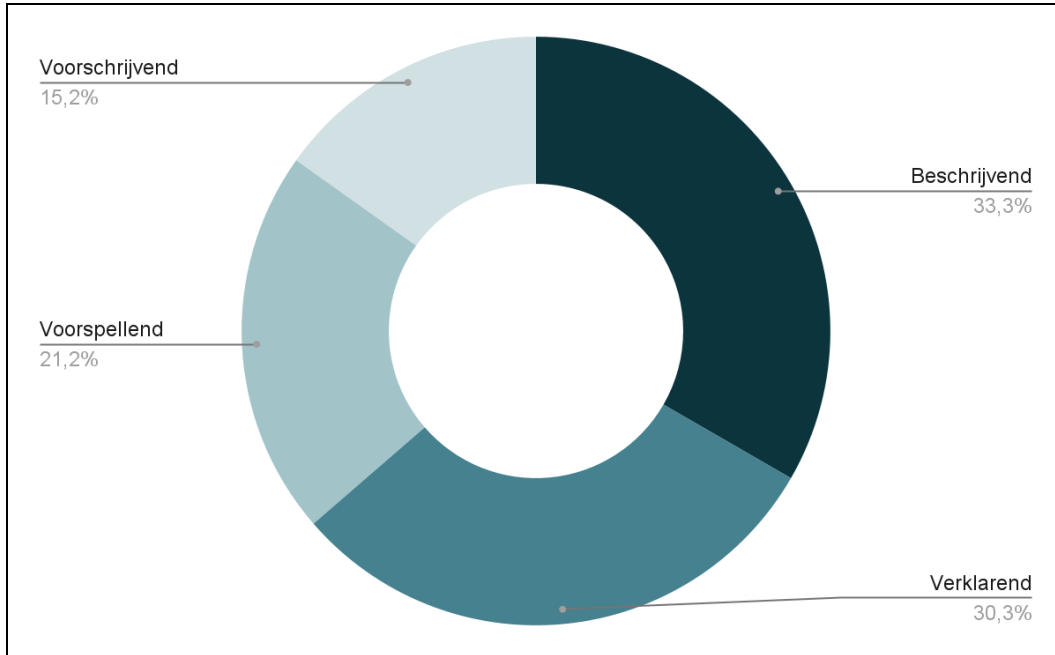
Als we het hebben over data-analyse worden vaak vier verschillende soorten onderzoek onderscheiden: beschrijvend (wat is er aan de hand), verklarend (waarom is het gebeurd), voorspellend (wat gaat er gebeuren) en voorschrijvend (wat moet er gebeuren) onderzoek. Als we deze vier soorten onderzoek nader bekijken over de as 'soort gebruiker' valt op dat de mate van menselijke input en interactie bij het opschuiven naar voorschrijvend onderzoek steeds kleiner wordt (zie figuur 2).

Bij beschrijvend, verklarend, en tijdens de eerste stappen van voorspellend onderzoek wordt automatisering vooral gebruikt voor het verzamelen en prepareren van de data. Analyse van de gegevens en duiding van de uitkomsten gebeurt nog grotendeels op basis van menselijke input en besluitvorming, door bijvoorbeeld informatie-onderzoekers. Om professionals te ondersteunen in het optimaliseren van hun besluitvorming (zowel qua tijdsinverstering als kwalitatief) werken data scientist bij meer geavanceerde voorspelmodellen en voorschrijvend onderzoek aan algoritmes voor geautomatiseerde besluitvorming, met minimale menselijke tussenkomst. Verwachting is niet dat deze modellen in de zorg 'blind' kunnen worden opgevolgd, maar een verschuiving in de balans tussen mens en machine is wel realistisch.



Figuur 2. Schematische weergave van het besluitvormingsproces op basis van data: welk discipline heeft welke rol, en wat is de resterende mate van menselijke input.

Op basis van de ingevulde vragenlijsten blijkt dat er veel variatie zit in het soort onderzoek dat de beoogde gebruikers van de regionale datawerkplaats doen (zie figuur 3). Alle vier de vormen van onderzoek worden genoemd. Dit betekent dat er al veel waarde zit in het creëren van domeinoverstijgende inzichten op basis van basisstatistiek, maar dat er ook complexere datavragen zijn. Een deel van de gebruikers blijkt een verschuiving naar meer voorspellend (in plaats van beschrijvend) onderzoek te voorzien voor de toekomst: van 21% naar 27% voor beschrijvend onderzoek en van 33% naar 27% voor voorspellend onderzoek).



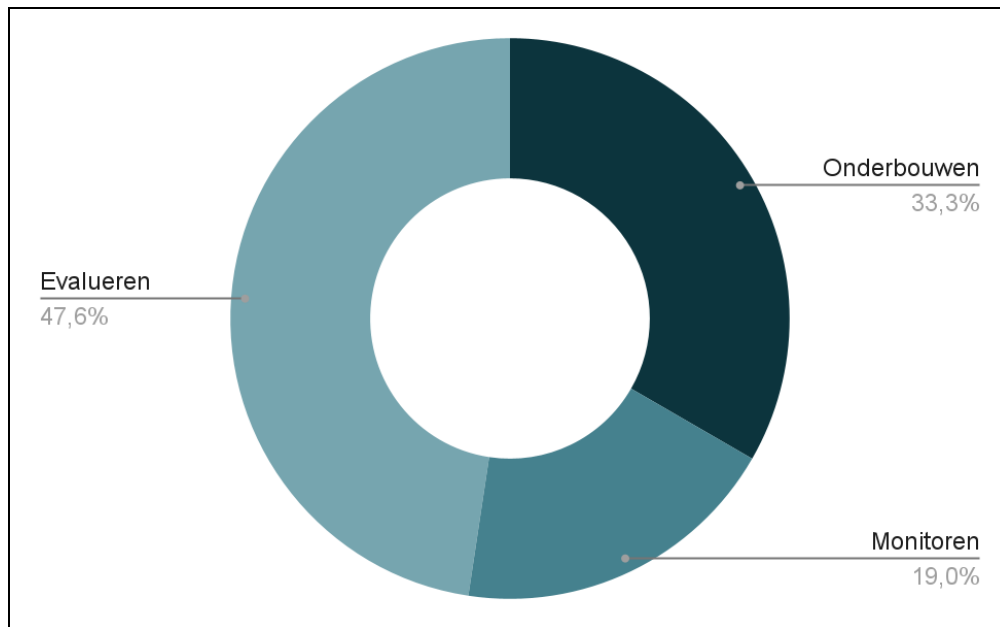
Figuur 3: Verdeling van de verschillende soorten onderzoek waar gebruikers zich in praktijk mee bezig houden.

3.3 Onderzoeksvragen

De onderzoeksvragen waar een regionale data-infrastructuur aan bij kan dragen ontstaan door de combinatie van het type gebruiker (paragraaf 3.1), het soort onderzoek (paragraaf 3.2) en de inhoud van specifieke use cases (toepassing). De combinatie van deze drie aspecten leidt tot een breed scala aan onderzoeksvragen.

Om overzicht aan te brengen in alle mogelijke soorten onderzoeksvragen hebben we de input van de respondenten op de vragenlijst geclusterd naar 'intentie'. Wat beoogt de gebruiker te bereiken? Dit leidt tot een clustering naar drie soorten onderzoeksvragen: onderbouwen (bestaande uit identificeren, prioriteren en besluitvorming), evalueren en monitoren. Deze indeling sluit ook aan op de uitwerking van de data-infrastructuur voor gezondheidsonderzoek van de Nationaal Groeifonds-aanvraag "Preventie-Promotor" van de Nederlandse Federatie van Universiteiten (NFU).

Onderstaande figuur (figuur 4) laat zien hoe vaak ieder van deze soort onderzoeksvragen genoemd is door de respondenten.



Figuur 4: Verdeling van het soort onderzoeksvragen waar gebruikers zich mee bezig houden

In bijna de helft van de onderzoeksvragen (48%) waarvoor de respondenten een regionale datawerkplaats willen gebruiken gaat het om evalueren. In 33% van de gevallen worden de gevraagde inzichten gebruikt om interventies te identificeren, prioriteren en besluitvorming te organiseren. Voor alle drie de soorten onderzoeksvragen geldt dat de respondenten aangeven dat het niet om ad-hoc analyses gaat, maar om activiteiten met een structureel karakter.

Een nadere uitwerking van de drie soorten onderzoeksvragen die we terug zien op basis van de antwoorden in de vragenlijst staan in de tabel hieronder:

Beoogd doel	Toelichting
Onderbouwen	<p>Identificeren: verdiepend bekijken van een vraagstuk, bijvoorbeeld om een beter beeld te krijgen van de doelgroep. Dit kan zowel over de tijd zijn als in vergelijking met andere groepen, regio's, etc.</p> <p>Prioriteren: leveren van feitelijke, objectieve inzichten op basis waarvan bepaald wordt welk inhoudelijk vraagstuk prioriteit heeft. Bijvoorbeeld voor het gericht inzetten van verandercapaciteit.</p> <p>Besluitvorming: onderbouwen van keuzes op basis van geïdentificeerde en geprioriteerde effecten, bijv. door het maken van één of meerdere business cases.</p>
Monitoren	Bewaken van de ontwikkeling van een indicator over de tijd, bijvoorbeeld om de effectiviteit van procesverbeteringen en interventies inzichtelijk te maken.

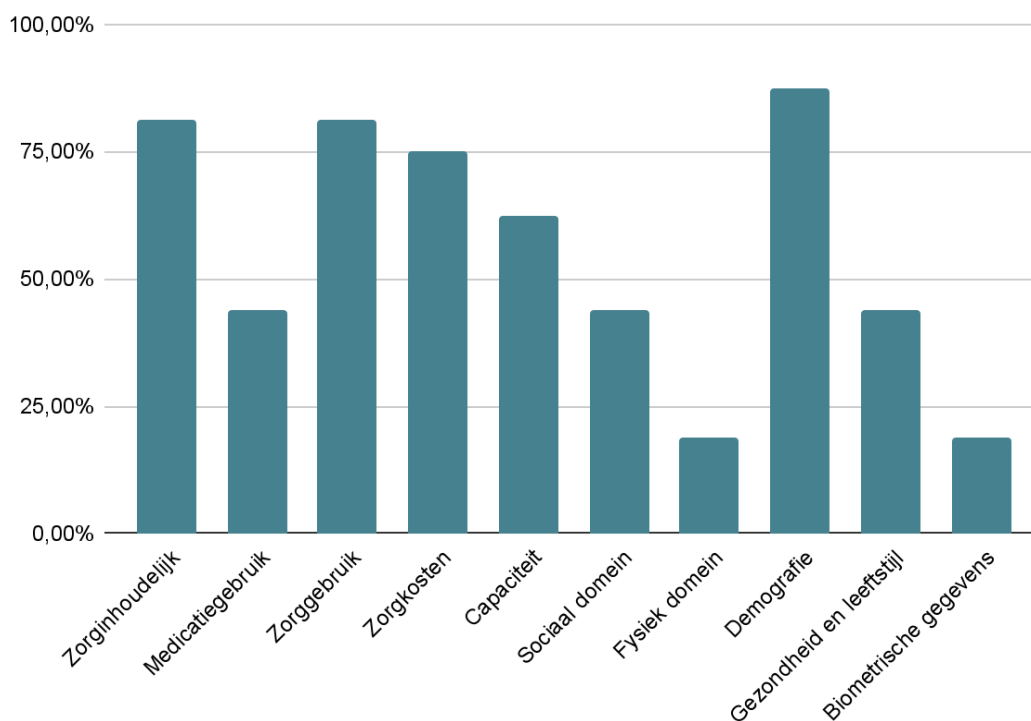
Evalueren	Evalueren van de doelmatigheid van een (zorg)keten of interventie. Dit kan gaan over (zorg)inhoudelijke vraagstukken, zoals preventiemaatregelen en doorstroom, maar ook over randvoorwaardelijke vraagstukken als capaciteitsmanagement of verzuim.
-----------	--

4 Benodigde data in de datawerkplaats

Om domeinoverstijgende vraagstukken op te pakken is logischerwijs een breder spectrum aan data nodig. In aanvulling op interne bronnen is de beschikbaarheid van externe (open) data gewenst. Ook vragen verschillende soorten onderzoek (van beschrijvend tot voorschrijvend) een andere breedte, aggregatieniveau of actualiteit van de in de datawerkplaats beschikbare data. De gewenste data en datamenmerken zijn in de vragenlijst uitgevraagd, waarvan de antwoorden hieronder verder zijn uitgewerkt.

4.1 Databronnen

In de antwoorden zien we de wens terug voor een breed spectrum aan data om de diverse onderzoeksvragen die leven te kunnen beantwoorden (figuur 5). Ruim driekwart van de respondenten (75-88%) geeft aan in de datawerkplaats graag gebruik te willen maken van zorginhoudelijke data (o.a. diagnose, ziektebeelden, voorgeschiedenis), data over zorggebruik (o.a. DBC, fysiotherapie, GGZ), (zorg)kosten (o.a. declaratie gegevens) en demografische gegevens (o.a. leeftijd, geslacht, SES). Daarnaast noemt ongeveer de helft (44-63%) de wens voor gegevens over capaciteit van instellingen (voor wat betreft plekken, middelen en personeel), medicatiegebruik en data vanuit het sociaal domein en het domein publieke gezondheid. Een kleiner deel (zo'n 20%) verwijst naar databronnen uit het fysieke domein en/of door patiënten zelf vastgelegde gegevens.



Figuur 5: Gewenste domeinen van waaruit data benodigd is om relevante onderzoeksvragen te beantwoorden onder gebruikers.

4.2 Breedte/historie

De respondenten zijn eensgezind over dat historische data in de datawerkplaats beschikbaar moet zijn. Dit is onder andere nodig voor trends en het doen van voorspellingen. De hoeveelheid historie waarover de respondenten willen kunnen beschikken varieert van minimaal 1 tot maximaal 15 jaar terug, waarbij een periode van 3-5 jaar het meest genoemd wordt. De specifieke onderzoeksvraag bepaalt hoeveel historie relevant is, bijvoorbeeld op basis van de incidentie/prevalentie van bepaalde ziektebeelden.

4.3 Aggregatieniveau

Ook voor wat betreft de voorkeur voor het niveau waarop de brondata vastgelegd moet zijn geven de respondenten aan dat dit afhankelijk is van de specifieke onderzoeksvraag. Vastlegging op individueel niveau heeft bij de meeste respondenten de voorkeur (12 keer genoemd) over groepsniveau (7 keer genoemd). Bij het opschuiven naar voorspellen en voorschrijven neemt het belang van een zo laag mogelijk aggregatieniveau steeds verder toe.

4.4 Actualiteit

Over de gewenste actualiteit van de beschikbare data in de datawerkplaats zijn respondenten minder eensgezind. De meningen zijn (ongeveer) gelijk verdeeld over de noodzaak om dagelijks, wekelijks, ieder kwartaal of ieder jaar de data verversen. De gewenste frequentie hangt ook hier weer sterk samen met de onderzoeksvraag. Als de uitkomsten gebruikt worden voor beleidsvorming is minder actuele data afdoende, bij capaciteitsmodellen is er de noodzaak om regelmatig de data bij te werken (afhankelijk van de beslisfrequentie).

5 Functionaliteiten van de datawerkplaats

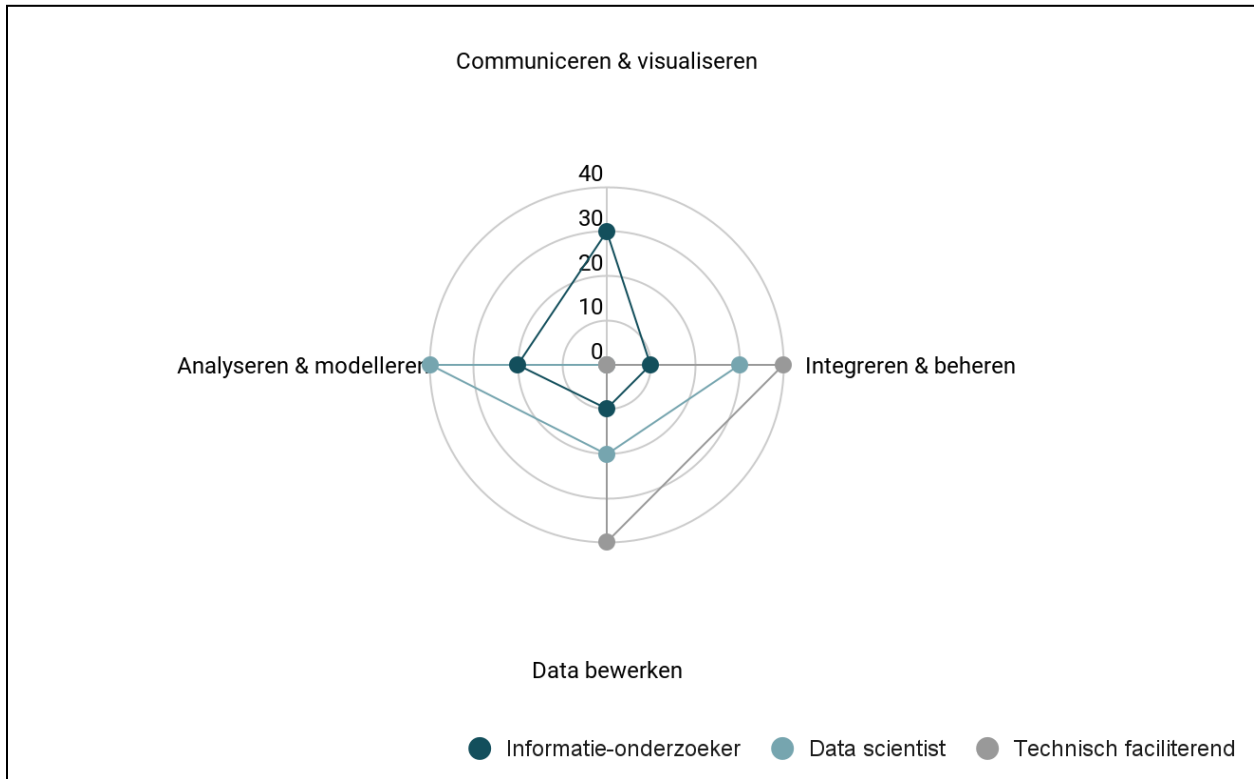
Om beschikbare domeinoverstijgende data tot waarde te kunnen maken hebben gebruikers van de regionale datawerkplaats verschillende functionaliteiten nodig voor het verwerken en analyseren van de beschikbare data en het opleveren van gezondheidsinzichten. De [data science radar](#) geeft een indruk van benodigde gebruikersrollen in een data(science)team. Daarbij valt op te merken dat in een groter team de rollen door verschillende gebruikers worden ingevuld en in een kleiner team gebruikers zich vaker meerdere rollen eigen moeten maken. Op basis van de in de data science radar beschreven werkzaamheden en rollen is voor dit whitepaper een eigen aggregatie gemaakt naar vier (minimaal) benodigde functionaliteiten voor een regionale datawerkplaats. In onderstaande tabel zijn deze functionaliteiten opgenomen, met daarnaast een toelichting op de werkzaamheden behorend tot deze functionaliteit.

Functionaliteiten	Toelichting
Integreren & beheren	Deze groep richt zich op het ontsluiten & integreren van data en het optimaliseren van het data transformatieproces. Dit gebeurt o.a. door het bouwen van pijplijnen om data uit de bronnen te ontsluiten en door technisch beheer van de regionale datawerkplaats (o.a. versiebeheer en testwerkzaamheden).
Data bewerken	Deze groep richt zich op databewerking, zodat ruwe data geschoond wordt en klaar gemaakt voor verdere analyse. Het doel is om een valide, goed opgebouwde (gemodelleerde) dataset op te leveren die gebruikt kan worden voor analyse.
Analyseren & modelleren	Deze groep richt zich op het willen exploreren en doorgronden van de dataset, met behulp van statistische technieken. Daarvoor gebruiken zij minder of meer geavanceerde analyses (van basisstatistiek in de vorm van prevalenties en kruistabellen tot voorspellende modellen, clusteranalyse, neurale netwerken etc).
Communiceren & visualiseren	Deze groep richt zich op de dialoog met de informatie-vragers. Pro-actief en reactief gaat deze groep de dialoog aan met alle stakeholders om data te laten helpen in het nemen van business keuzes. Zij begrijpen dat (en weten hoe!) een plaatje meer zegt dan duizend woorden.

In de context van de regionale datawerkplaats in de Achterhoek onderscheiden we zoals eerder gezegd twee soorten gebruikers: informatie-onderzoekers en data scientists. Daarnaast constateren we dat er een extra rol ontstaat met een wat meer technisch profiel, gericht op integratie & beheren en/of data bewerking. Bij data scientists valt de rol 'Data bewerken' vaak binnen hun werkzaamheden. Bij informatie-onderzoekers is dit soms wel het geval (bijvoorbeeld voor BI-specialisten) en soms niet (bijvoorbeeld voor (beleids)onderzoekers). Integreren & beheren en/of data bewerking ligt in die

gevallen op een andere plek in de organisatie of bij een externe (BI) partij. Deze rollen zijn minder gericht op het creëren van inzichten voor inzicht-vragers, maar hebben vooral een faciliterende rol in het datatransformatieproces.

In figuur 6 zijn de twee gebruikersgroepen en de technisch faciliterende rol geplot op de vereenvoudigde en door ons naar functionaliteiten vertaalde variant van de data science radar.



Figuur 6: Informatie-onderzoekers, data scientists en technisch faciliterende rollen ingedeeld in de gedefinieerde functionaliteiten in een data(science)team.

5.1 Tooling

In de vragenlijst is aan de respondenten gevraagd of zij voor de vier gedefinieerde functionaliteiten (integreren & beheren, data bewerken, analyseren & modelleren en visualiseren & communiceren) voorkeur hebben voor specifieke tooling. In onderstaande tabel staan de verkregen inzichten.

Functie	Tooling van voorkeur	Toelichting
Integreren & beheren	Embedded analytics (naar dossiers); APIs; Databasekoppeling (naar eigen omgeving)	Behoeftte rondom data-integratie met inkomende bronnen is niet nagevraagd in vragenlijst ivm eerdere keuze voor MPC. Outputfunctionaliteiten zijn uitgewerkt in paragraaf 5.3
Data bewerken	Datavault	Faciliterend in dataschoning en datamodellering (bijv ten behoeve van informatiestandaarden)
Analyseren & modelleren	R/Python, SSAS Tabular, Stata	Voorkeur voor SQL-based tooling om zelf specificaties (o.a. Het plaatsen van filters, specifieke ad hoc-onderzoeksvragen) te kunnen formuleren in analyse & onderzoek. De tooling moet daarnaast het soort onderzoek faciliteren (van beschrijvend tot voorschrijvend)
Communiceren & visualiseren	Power BI, Excel	Wens voor 'stand-alone' tool voor visualisatie, gescheiden van data-integratie en -modellering. Communiceren vraagt geen (technische) functionaliteit in de regionale datawerkplaats

5.2 Delen van inzichten

Zoals al eerder aangegeven ligt de focus van dit whitepaper op informatie-onderzoekers en data scientists als primaire gebruikers van de datawerkplaats. Zij zijn in veel gevallen aan de slag in dienst van het derde type gebruiker, de inzicht-vrager. Om gezondheidswinst te kunnen behalen is het belangrijk de resultaten van het onderzoek om te zetten in gerichte (preventieve) acties. Hiervoor is het essentieel de de opgedane inzichten vanuit de datawerkplaats te kunnen delen.

Aan de respondenten van de vragenlijst is gevraagd met welke inzicht-vragers de uitkomsten gedeeld moeten worden en op welke wijze deze inzage (op basis van mate van vertrouwelijkheid) goed vorm kan krijgen. In de helft van de gevallen betreft de groep inzicht-vragers een geautoriseerde groep, bijvoorbeeld beleidsmakers of zorgprofessionals in een Multi-Disciplinair Overleg. In dit geval is het van belang dat het binnen de datawerkplaats mogelijk is om specifieke groepen rechten te geven via autorisatiebeheer. Dit kan technisch of via onderlinge afspraken geborgd worden. Ruim 37% van de respondenten geeft aan dat de opgeleverde inzichten openbaar toegankelijk kunnen zijn, voor zowel de inzicht-vrager als andere geïnteresseerden. Naar inschatting betreft het hier met name geaggregeerde output, bijvoorbeeld in het kader van benchmarking, of een representatie van voornamelijk openbare data. Tot slot geeft 12,5% van de respondenten aan dat zij verwachten zowel informatieproducten op te leveren die openbaar gedeeld mogen worden, als informatieproducten die voor een specifieke groep bedoeld zijn.

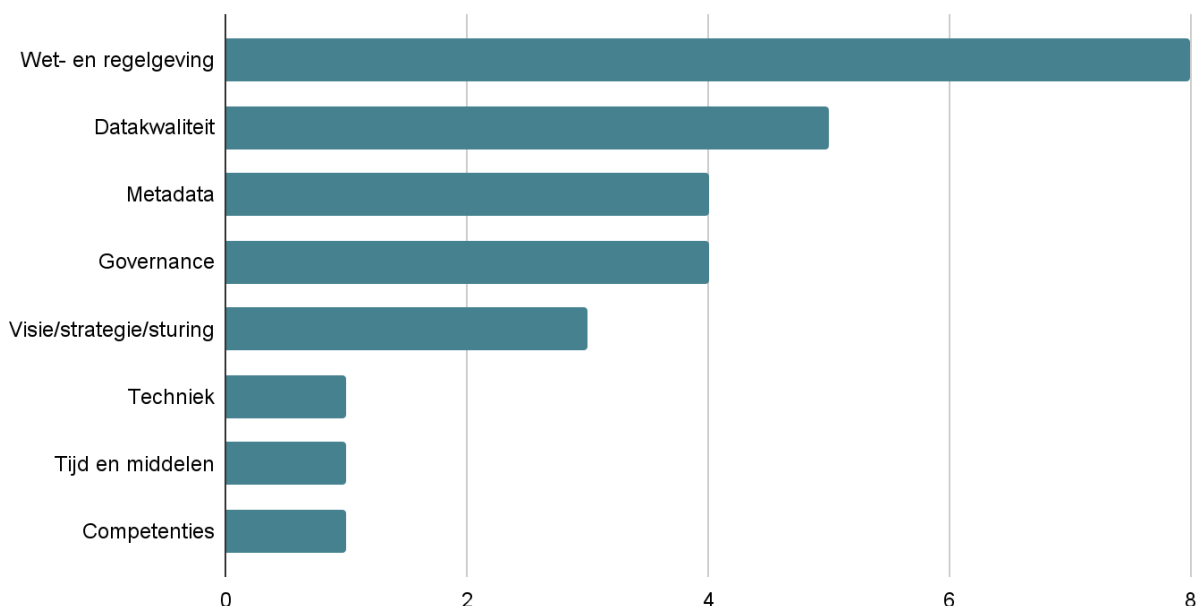
5.3 Output opties

Naast het direct delen van inzichten vanuit de regionale datawerkplaats is het soms wenselijk en/of noodzakelijk om output vanuit de datawerkplaats te kunnen exporteren, bijvoorbeeld voor nadere analyse in een eigen omgeving. In de vragenlijst is gevraagd op welke wijze de datawerkplaats gegevens hiervoor beschikbaar zou moeten maken.

Iets meer dan de helft van de respondenten (56%) geeft aan dat een mogelijkheid tot exporteren gewenst is, bijvoorbeeld om output op te nemen in de applicatie van een zorgprofessional (zoals het elektronisch cliëntdossier) en/of een interne BI/onderzoeksomgeving. De vorm van de gewenste export naar een eigen BI/onderzoeksomgeving varieert sterk; van CSV naar directe koppeling of een API. Er zijn ook respondenten (44%) die geen (dringende) wens hebben voor een mogelijkheid tot exporteren. In dit geval blijft de informatie binnen de datawerkplaats. Vraag is of deze respondenten exports niet noodzakelijk, of niet wenselijk achten. Met andere woorden, zijn er respondenten die uitdrukkelijk niet willen dat output de datawerkplaats verlaat? Dit is een aandachtspunt bij het inrichten van governance. Is dit een randvoorwaarde voor de eigen data-ontsluiting naar de datawerkplaats?

6 Mogelijke obstakels en aandachtspunten

Hoewel domeinoverstijgend data delen en analyseren in een regionale datawerkplaats mooie kansen biedt, zijn er ook randvoorwaarden nodig om binnen de regionale datawerkplaats de toegevoegde waarde voor de regio en/of deelnemende organisaties te realiseren. In de vragenlijst hebben we bij potentiële gebruikers van een regionale datawerkplaats nagevraagd welke obstakels zij voorzien bij het oppakken van domeinoverstijgend onderzoek in het algemeen, of de eigen domeinoverstijgende onderzoeksvraag in het bijzonder. Zie hieronder (figuur 7) de genoemde randvoorwaarden, van meest naar minst genoemd.



Figuur 7: De door beoogde gebruikers genoemde obstakels bij het oppakken van domeinoverstijgend onderzoek.

6.1 Wet- en regelgeving

Wet- en regelgeving werd het meest genoemd als potentieel obstakel voor domeinoverstijgend onderzoek (in de regio Achterhoek). Als voorbeelden van vraagstukken die leven worden onder andere genoemd: het mogen koppelen van persoonsgegevens, het mogen analyseren van data over de domeinen heen, het risico op het ontstaan van tot personen herleidbare informatie bij het combineren van bronnen, de instemming van patiënten/cliënten voor het gebruik van de data en privacy, en security in het algemeen. Tegelijkertijd wordt aan dit onderwerp gewerkt; bijv. binnen *Multi Party Computation* worden verschillende zaken rondom data-minimalisatie (onderdeel van de DPIA) weggenomen door middel van technische oplossingen. Op andere vlakken, bijvoorbeeld landelijk, wordt gewerkt aan het wegnemen van obstakels door (mogelijke) verandering van wetgeving, zoals in het [Obstakel Verwijder Traject Gezondheidsdata van Health-RI](#).

6.2 Datakwaliteit en beschikbaarheid metadata

De respondenten zien ook veel uitdagingen op het gebied van datakwaliteit van de in de regionale datawerkplaats beschikbare gegevens en de aanwezigheid van relevante metadata. Voor wat betreft datakwaliteit noemen zij dingen als verschil in definities tussen organisaties, mogelijke foute registratie van gegevens bij de bron, het ontbreken van benodigde data en een onvoldoende detailniveau van beschikbare data om de gewenste analyse(s) uit te kunnen voeren. Daarnaast wordt de relevantie van de beschikbaarheid van metadata helder. Voor iedere bron is het noodzakelijk om te weten met welk doel de gegevens zijn vastgelegd, wat de gebruikte definities zijn, hoe de beschikbare gegevens geïnterpreteerd kunnen worden, of er sprake is van selectiviteit van de gegevens en of er wegingsfactoren beschikbaar zijn. Inzicht krijgen in datakwaliteit en metadata pleit voor nadrukkelijke en actieve betrokkenheid van alle bij het informatievraagstuk betrokken organisaties, zodat samen een goede informatiehuishouding neergezet kan worden.

6.3 Governance

De onderwerpen die genoemd worden onder “wet- en regelgeving” en “datakwaliteit en beschikbaarheid metadata” vertalen zich door naar het belang van goede governance. Zo is er de wens voor helderheid omtrent eigenaarschap van de data, het kunnen ‘terug trekken’ van gedeelde data, rechten (en plichten) die je hebt als organisatie en/of data scientist/informatie-onderzoeker binnen de datawerkplaats, het bepalen van een gouden standaard voor definities en welke verwerking mogelijk en/of wenselijk is op basis van de data. Dit geldt voor de diverse onderzoeksvragen, maar governance wordt ook genoemd voor de regionale datawerkplaats in het algemeen. Waar en hoe is het eigenaarschap, beheer en onderhoud van de datawerkplaats belegd? Verdere invulling van deze vragen is in scope van het werkpakket governance.

6.4 Visie, strategie en sturing

Om daadwerkelijk waarde te creëren binnen de regionale datawerkplaats is het van belang om helder te hebben welk doel deelname aan de datawerkplaats dient en of opgeleverde inzichten (nog) relevant zijn en meerwaarde hebben ten opzichte van wat anderen bieden. Op welke onderwerpen wordt wel of niet ingezet? Daarnaast wordt door een respondent genoemd dat het belangrijk is om onderscheid te maken tussen samenwerken tussen domeinen in het algemeen en datagedreven samenwerken tussen domeinen in het bijzonder. Praktische toepasbaarheid is essentieel.

6.5 Techniek, tijd, middelen en competenties

Praktische randvoorwaarden die tot slot worden genoemd liggen op het gebied van techniek, tijd, middelen en competenties. Sluit de beschikbare software aan bij het soort onderzoek dat gedaan wordt? Is de omgeving snel en krachtig genoeg? Waar vindt men tijd en middelen voor het samen (laten) brengen en analyseren van de data? En is er binnen de regio en/of organisatie voldoende kennis en kunde voor het optimaal gebruiken van de mogelijkheden die de regionale datawerkplaats biedt?

6.6 Ethische en sociale aspecten

Het onderwerp ethische en sociale aspecten kwam niet expliciet terug in de antwoorden die werden gegeven. Tegelijkertijd achten wij het wel belangrijk dat alle "ELSI" (*ethical, legal, societal*) aspecten aandacht krijgen in de onderzoeksgovernance van een regionale datawerkplaats. Een goede invulling hiervan vereist meer onderzoek, bijvoorbeeld aansluitend op de activiteiten zoals deze in Nederland nu plaats vinden binnen de zgn. [ELSA-labs](#).

7. Conclusie

In dit whitepaper staat de regionale datawerkplaats centraal: een middel om van gezondheidsdata naar bruikbare inzichten en informatie te gaan. Dit whitepaper focust op de benodigdheden van de datawerkplaats volgens de verschillende beoogde gebruikers. De onderzochte onderwerpen worden in elk individueel hoofdstuk uitgebreid behandeld, maar hieronder kort samengevat per onderwerp.

Scope

De respondenten voorzien een verschuiving van beschrijvend naar voorspellend onderzoek. Verder blijkt 70% van de onderzoeksvragen die spelen gericht op het identificeren van bijv. doelgroepen en vraagstukken, en evalueren van interventies en doelmatigheid.

Benodigde data

Het oppakken van domeinoverstijgende vraagstukken, in combinatie met een verschuiving naar meer voorspellend en voorschrijvend onderzoek, verandert de databehoeftte. Een breed spectrum aan data is gewenst, waaronder medisch diagnostische gegevens (inclusief voorgeschiedenis), ziektebeelden, medicatie gebruik, zorggebruik, (zorg)kosten, demografische gegevens, capaciteitsgegevens van instellingen (voor wat betreft plekken, middelen en personeel), data vanuit het sociaal domein, het domein publieke gezondheid, fysieke domein en/of door patiënten zelf vastgelegde gegevens. Deze data moet waar mogelijk historisch beschikbaar zijn, waarbij een periode van tot 3-5 jaar terug het vaakst genoemd wordt. Voor wat betreft het niveau van vastlegging van de brondata, heeft individueel niveau de voorkeur. Over de actualiteit van de data is er een minder eenduidig beeld. Zowel dagelijks, wekelijks, ieder kwartaal of jaarlijks verversen worden genoemd. Vast staat dat met het gebruik van meer complexe data-analyse technieken (o.a. clusteranalyse, regressie-analyse, decision trees) de behoefte aan meer (historische) data, actuele data en data op individueel niveau toeneemt.

Functionaliteiten

Een analyse van functionaliteiten leert dat niet enkel informatie-onderzoekers en data scientists in de regionale datawerkplaats werkzaam zullen zijn, maar ook collega's met een technisch faciliterend profiel (activiteiten gericht op data integratie & beheer en data bewerking). Verwachting is dat in de helft van de gevallen de geleverde inzichten aan een geautoriseerde groep beschikbaar gesteld moet worden. Ten aanzien van de exportmogelijkheden blijkt er een punt van aandacht: waar 56% van de respondenten vraagt om een exportmogelijkheid, geeft de rest aan het (nu) niet nodig te achten of zelfs onwenselijk te vinden.

Obstakels en aandachtspunten

Hoewel domeinoverstijgend data delen en analyseren in een regionale datawerkplaats mooie kansen biedt zijn er ook randvoorwaarden nodig om binnen de regionale datawerkplaats toegevoegde waarde voor de regio en/of deelnemende organisaties te realiseren. Het meest genoemd worden vraagstukken rondom wet- en regelgeving, datakwaliteit en beschikbaarheid metadata, wat zich vertaalt naar de noodzaak voor een goede governance, waarbij regels en afspraken zoveel mogelijk binnen de regionale datawerkplaats worden afgedwongen. Wanneer voorgaande stappen worden doorlopen wordt concreet invulling gegeven aan 'data-gedreven werken' rondom gezondheidsvraagstukken.

Wil je bijdragen aan de (door)ontwikkeling van het technische ontwerp van GERDA, laat het ons dan weten. Neem hiervoor contact op met Ilona Oude Nijhuis (ilonaoudenijhuis@room-to.nl) of Maarten den Braber (maarten@populationhealthdata.nl)