

Whitepaper

KIK-V x GERDA

Werken aan databeschikbaarheid door hergebruik

**VERENIGING
DIGITALISERING ZORG
ACHTERHOEK**



Datum:

16 april 2024

Door:

GERDA (Jerry Fortuin, Ilona Oude Nijhuis en Maarten den Braber)

KIK-V (Thamar Brouwer en Joost Wildenberg)

Linksight (Maarten Everts en Martine van de Gaar)

Voor:

Datasamenwerkingen gericht op secundair datagebruik

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.

Contactpersoon: ilonaoudenijhuis@room-to.nl

Inhoudsopgave

Inhoudsopgave.....	1
Samenvatting.....	2
Aanleiding en vraagstelling.....	3
Landelijk Dekkend Netwerk.....	5
Conclusie.....	5
Bevragingen.....	6
Conclusies.....	8
Vergelijking KIK-V en GERDA.....	9
KIK-V.....	9
GERDA.....	10
Vergelijking.....	11
Governance.....	13
Hergebruik binnen KIK-V en GERDA.....	14
Techniek.....	15
Uitbreiding Datastation-as-a-Service functionaliteit.....	16
Gestructureerd vs. exploratief informatiemodel.....	18
Gebruik gezamenlijke ontologie.....	18
Snelheid van beantwoording.....	18
Volwassenheid in datagedreven werken.....	19
Aanbevelingen en vervolg.....	20

Samenvatting

Dit whitepaper is geschreven in opdracht van het Zorginstituut om vanuit 2 vraagstellingen te onderzoeken hoe toegewerkt kan worden naar databeschikbaarheid door hergebruik. Daarbij is het landelijk dekkend netwerk de ambitie. Scope van dit whitepaper betreft secundair datagebruik - hoewel er ook een wenkend perspectief voor een uniform datastation ligt voor primaire gegevensuitwisseling.

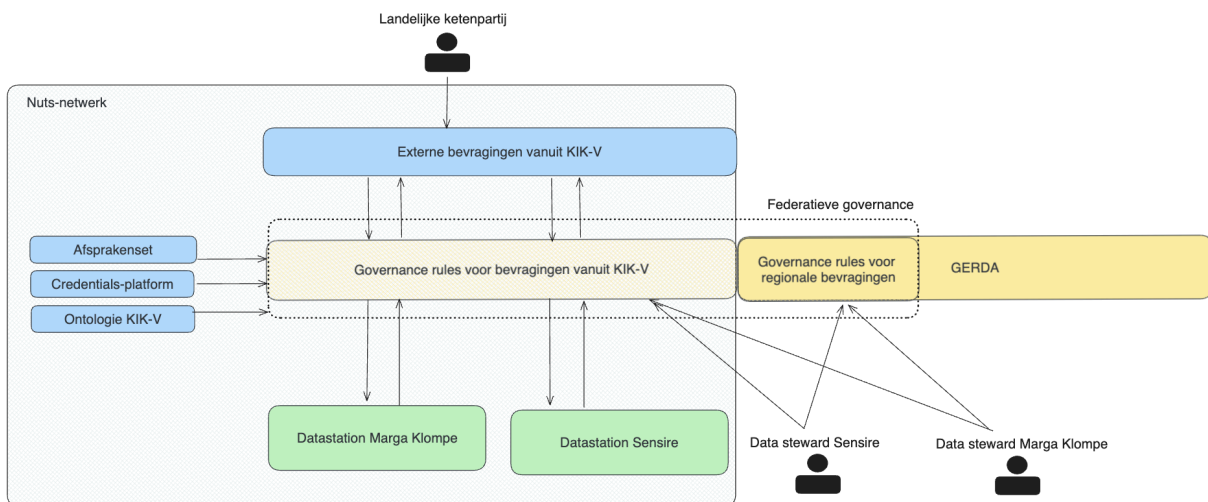
Uit een brede analyse van secundair datagebruik volgen de volgende kenmerken van bevragingen:

1. Met of zonder directe vertrouwensrelatie tussen vraagsteller en bevragde
2. De vraagsteller vraagt inzicht over een individuele organisatie of over meerdere partijen binnen de datasamenwerking (geadresseerde)
3. Vraagsteller kan gevalideerde vraag stellen of heeft exploratieve analyseruimte
4. Bevraging vraagt wel of geen data-integratie (binnen de organisatie of in de datasamenwerking)

Uit deze kenmerken volgen benodigde (herbruikbare!) functionaliteiten. Voor KIK-V en GERDA ligt dit potentieel zich op het gebied van:

- Federatieve governance: met als doel om data stewards van organisaties soevereiniteit te geven over het gebruik van data. Er zijn mogelijkheden om het federatieve gedachtegoed beide concepten ook op het gebied van governance vanuit GERDA te verbreden naar KIK-V.
- Universele DaaS: door het datastation te voorzien van rijkere functionaliteiten wordt het mogelijk om zowel het datastation als de ontologie te hergebruiken voor GERDA. Meerwaarde voor organisaties is dat hier niet dubbel op geïnvesteerd en beheerd hoeft te worden. Ook ontstaat er een FAIR basis voor het hergebruik van data voor secundaire doeleinden.

Een grove schets van een proof of concept ziet er als volgt uit:



Aanleiding en vraagstelling

De Vereniging Digitalisering Zorg Achterhoek heeft in de afgelopen twee jaar resultaten bereikt, op terrein van (1) data-uitwisseling en (2) databeschikbaarheid in de (regionale) zorg. Op een aantal van deze terreinen wordt nauw samengewerkt met het Zorginstituut Nederland (ZIN), het programma KIK-V en de directie Langdurige Zorg binnen het ministerie VWS.

Concrete resultaten zijn:

- primaire gegevensuitwisseling in de vorm van de verpleegkundige eOverdracht (ziekenhuis - VVT en VVT - VVT) (te verwachten rond de zomer '24);
- secundaire gegevensuitwisseling in het kader van KIK-V (data op orde, implementatie uitwisselprofielen en datastation as a service) bij de twee koplopers in de regio Sensire en Marga Klompé. Met als wenkend perspectief vervolg implementatie in 2024 bij de overige vier VVT organisaties (Careaz, Azora, Markenheem, Gouden Leeuw Groep).
- samenstellen en beproeven Handreiking InZicht - KIK-V in samenwerking met Min. VWS directie LZ, programma KIK-V en Bureau InZicht/eOverdracht.
- secundaire domeinoverstijgende databeschikbaarheid en datagebruik voor (beleids)onderzoek, monitoring, (richtlijnen)toetsing, op basis van Privacy Enhancing Technology (PET) in de vorm van de virtuele datawerkplaats GERDA met gebruik van de Secure Multi-Party Computation dienstverlening van Linksight.

De Vereniging Digitalisering Zorg Achterhoek heeft daarmee in de regio de basis gelegd voor een federatief netwerk (zie ook netwerkperspectief) voor (domeinoverstijgend) primaire en secundaire databeschikbaarheid, -uitwisseling en -gebruik.

Dit whitepaper is uitgewerkt voor het Zorginstituut Nederland op basis van de volgende vraagstellingen:

1. In de Achterhoek is ten behoeve van de virtuele data werkplaats GERDA reeds een data governance model en afspraken set ontwikkeld. De vraagstelling is als volgt:
 - a. Welke elementen uit het 'GERDA' governance model (incl. ethisch protocol) zijn herbruikbaar t.b.v een 'generiek (regionaal) data governance model', waarmee ook toepassingen als KIK-V worden ondersteund?
 - b. Welke aanpassingen/toevoegingen moeten er komen aan het 'GERDA' governance model t.b.v een 'generiek (regionaal) data governance model', waarmee ook toepassingen als KIK-V worden ondersteund?
2. In hoeverre en onder welke condities is het in het kader van het KIK-V programma ontwikkelde Credentialsplatform (mag de vraagsteller de vraag stellen?) en de KIK-starter herbruikbaar voor toepassing(en) in de regio?

In dit whitepaper worden de eerste 3 lagen van het 5-lagen interoperabiliteitsmodel van Nictiz - organisatie, proces en informatie - gevat onder 'Governance' (bovenstaande vraag 1) en de onderste 2 lagen - applicatie en IT infrastructuur - gevat onder 'Techniek' (bovenstaande vraag 2). Dit whitepaper kiest een functionele beschrijving van bevragingen, governance en techniek.

De reikwijdte van dit whitepaper omslaat secundair datagebruik. Op termijn is een versmelting tussen primair en secundair datagebruik te verwachten. Dit onderstreept nogmaals het belang van een herbruikbare inrichting rondom datagebruik.

Voordat dit whitepaper bovenstaande vraagstellingen kan beantwoorden, worden deze tegen het licht van de ontwikkeling 'Landelijk Dekkend Netwerk' geplaatst om nut en noodzaak van hergebruik te benoemen. KIK-V en GERDA zijn nader uitgewerkt om de functionele overlap en verschillen te kunnen identificeren.

Landelijk Dekkend Netwerk

In de [kamerbrief vervolg landelijk dekkend netwerk van infrastructuur](#) (d.d. 24 januari 2024) geeft de minister aan wat het vervolg is op de gekozen strategie die bestaat uit twee onderdelen:

- *Een gedistribueerd communicatienetwerk voor (geprotocolleerde) overdracht van gegevens door één-op-één communicatie tussen zorgverleners én om data beschikbaar te kunnen stellen voor andere doeleinden.*
- *Een data-centrische oplossing, bestaande uit gekoppelde dataplatformen, die het gebruik van data scheidt van de functionaliteit en databeschikbaarheid voor primair en secundair gebruik, alsmede gezamenlijke dossiervorming in de context van netwerkzorg, faciliteert.*

De minister geeft in zijn kamerbrief aan dat het doel is om van gegevensuitwisseling door te groeien naar databeschikbaarheid. De standaardisatie van taal en techniek gaat niet snel genoeg om de toenemende vraag naar toepassingen voor databeschikbaarheid te faciliteren. Diensten voor onder andere het ontsluiten van de data-, netwerk- en integratie diensten (o.a. verzamelen, valideren, samenvoegen en verrijken van data) worden nu voor elke toepassing separaat en opnieuw ontwikkeld en uitgevoerd.

De minister zet in op het verbinden van bestaande infrastructuur. Er is behoefte aan integratie- en netwerkdiensten om versnelling te realiseren. Daarbij wordt gestreefd naar publieke voorzieningen waarbij (her)gebruik van kennis, ervaring en functionaliteit bij bewezen initiatieven voorop staat. De wens is om deze te harmoniseren en integreren. Regionale en sectorale initiatieven die aantoonbaar schaalbaar kunnen na succesvolle beproeving landelijk worden geïmplementeerd.

Conclusie

De komende jaren gaat er veel gebeuren ten behoeve van databeschikbaarheid om de zorg te verbeteren. Als hier geen hergebruik van functionaliteiten aan ten grondslag ligt, ontstaat er in korte tijd een kostenintensief, niet-interoperabel landschap. Zoals eerder gesteld, is de reikwijdte van dit whitepaper 'slechts' secundair datagebruik. Op termijn is een versmelting van primair en secundair te verwachten, die de noodzaak tot hergebruik nog verder onderstreept.

Volgend hoofdstuk schetst de verschillende soorten bevragingen om een beter beeld te krijgen van de huidige en toekomstige situatie. Hieruit volgt ook een beter beeld van de mogelijkheden voor hergebruik in een bredere context dan alleen KIK-V en GERDA.

Bevragingen

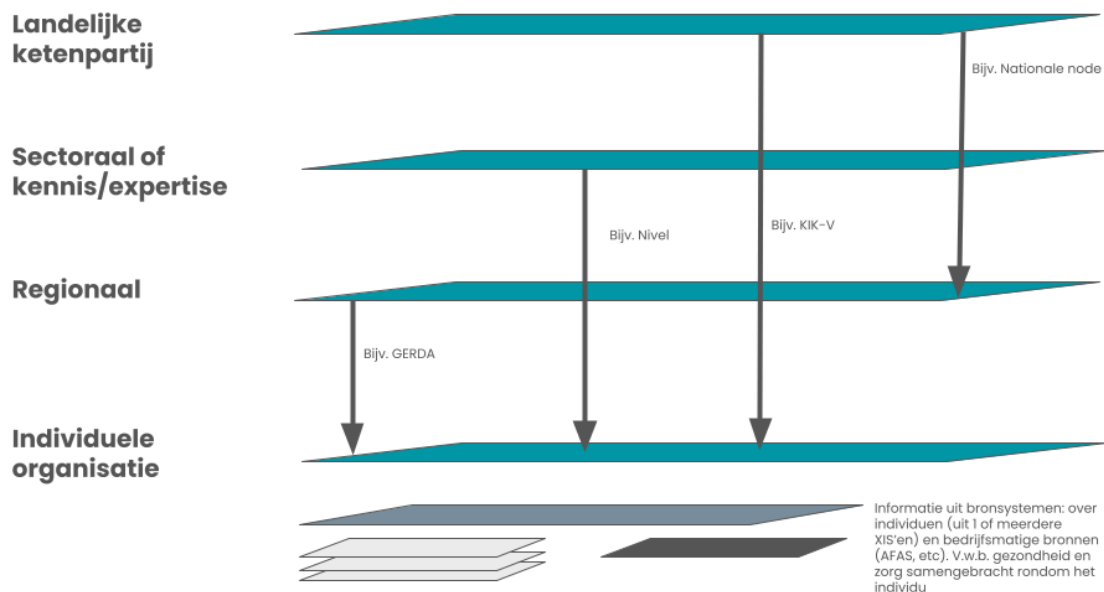
In dit hoofdstuk werken we bevragingen in het kader van secundair gebruik van data verder uit. Aan het einde van dit hoofdstuk resulteert dit in een set kenmerken. Feitelijk beschrijft dit hoofdstuk de benodigde functionaliteiten ten behoeve van bevragingen.

We zien de volgende soorten bevragingen op het gebied van inzichtbehoefte en vertrouwensrelatie:

- Bevragingen binnen de datasamenwerking met een gezamenlijke inzichtbehoefte en directe vertrouwensrelatie. GERDA is hiervan een voorbeeld. Partijen met een gezamenlijke informatiebehoefte starten een datasamenwerking die zich richt op het samenbrengen en analyseren van data om hiermee een gezamenlijk inzicht te creëren.
- Bevragingen binnen de datasamenwerking zonder gezamenlijke inzichtbehoefte en daarmee ook indirecte vertrouwensrelatie (deze bevragingen kenmerken zich door een grotere afstand tussen inzichtvrager en de data aanbieder). KIK-V is hiervan een voorbeeld. Tussen de individuele zorgorganisatie en de landelijke ketenpartij is geen directe relatie

Tevens zien we interacties tussen de verschillende aggregatieniveaus. In onderstaand plaatje zijn verschillende soorten bevragingen van het ene aggregatieniveau aan het andere aggregatieniveau getoond.

Bevragingen op het gebied van secundair datagebruik



Hieronder worden deze bevragingen verder uitgewerkt.

Bevragingen van landelijke ketenpartijen aan individuele zorgorganisaties

Voorbeeld: KIK-V

- Voor het bevragen van individuele zorgorganisaties.
- Het betreft gevalideerde vragen die een directe koppeling hebben met verplichtingen van de individuele zorgorganisaties. Deze zijn vastgelegd in een afspraken-set. Hoewel nu niet op de ontwikkelagenda van KIK-V - op termijn mogelijk ook bevragingen die wel gericht zijn op een regio en mogelijk ook data-integratie vragen (bijvoorbeeld, wat is de capaciteit in de acute zorgketen van de regio Achterhoek?).
- Data-integratie is enkel een vraagstuk binnen de muren van de eigen organisatie.
- De operationele datasamenwerking kenmerkt zich door afstand tussen inzichtvrager en bevrage. De ketenpartij die een inzicht vraagt aan het Datastation-as-a-Service (DaaS) van de individuele organisatie.
- Toekennen en controleren van rechten van landelijke inzichtvragers en bevrage. leunt nu nog op Credentialsplatform. Belangrijker is de infrastructuur waar vragen van vertrouwde organisaties op termijn met behulp van. Nuts-netwerk worden verwerkt.

Bevragingen van regio aan de individuele organisaties

Voorbeeld: [GERDA](#)

- Voor het bevragen van aangesloten organisaties in een datasamenwerking (= regio).
- Vraagstukken zijn gekoppeld aan regionale thema's en analyse is exploratief (waar zitten brandhaarden?) van aard. De onderzoeksruimte wordt in een gezamenlijke governance door data stewards middels governance rules bepaald op basis van wet- en regelgeving. Op termijn ook meer gerichte vragen, bijvoorbeeld bij het bewaken van ontwikkelingen en IZA/GALA-doelstellingen. Gevraagde inzichten krijgen daarnaast steeds meer een voorspellend/verklarend karakter (en vragen bijpassende data science expertise en tooling).
- Vraagstukken vragen per definitie data-integratie tussen de aangesloten organisatie voor de beantwoording.
- De operationele datasamenwerking bestaat uit een *learning community* van data stewards en data scientists die werken aan inzichten voor inzichtvragers.
- Toekennen en controleren van rechten van data scientists en data stewards volgt uit gezamenlijke verwerkersovereenkomst - in de eigen invloedssfeer. Op termijn mogelijkheid voor toestemming middels Nuts-netwerk van bevragers die van het gehele netwerk (bijvoorbeeld door koppeling met andere regio's en sectorale beroepsgroepen) informatie willen opvragen

Bevragingen van beroepsverenigingen en kennisinstellingen aan individuele organisaties

Voorbeelden: *Nivel, IKNL, LDF, etc.*

- Zowel gevalideerde als exploratieve vragen, vooral gericht op benchmarking
- Sectorale compute group (vanuit een beroepsgroep) met controleren van rechten via Nuts-netwerk (want groter aantal deelnemers over meerdere regio's c.q. landelijk).

- Inrichting kan vergelijkbaar zijn aan GERDA, waarbij de data stewards van deze organisaties gezamenlijk governance rules opstellen.
- Databeschikbaarheid op basis van datastations die ook al regionaal ingezet worden.

Bevragingen van (internationale en) landelijke partijen aan regio's

Voorbeelden: EHDS, CBS, National Node en Pandemische Paraatheid

- Nationaal bevragen van regionale netwerken (EHDS, nationaal beleid, Nationale Node, Pandemische Paraatheid). Op termijn ook gelinkt aan het Europese podium.
- Ambitie is andersom ook mogelijk te maken, waarbij landelijke partijen in regionaal netwerk (bijv. aansluiting CBS) betrokken worden

Conclusies

Alle 4 aggregatieniveaus zijn in ontwikkeling. Hierdoor ontstaat behoefte aan een herbruikbare inrichting voor secundair datagebruik. Uit de eerdere analyse komen de volgende kenmerken van bevragingen:

1. Met of zonder directe vertrouwensrelatie tussen vraagsteller en bevragee
2. De vraagsteller vraagt inzicht over een individuele organisatie of over meerdere partijen binnen de datasamenwerking (geadresseerde)
3. Vraagsteller kan gevalideerde vraag stellen of heeft exploratieve analyseruimte
4. Bevraging vraagt wel of geen data-integratie (binnen de organisatie of in de datasamenwerking)

Door op basis van deze kenmerken van bevragingen naar de ontwikkeling van benodigde functionaliteiten te kijken, ontstaat er een duurzaam ontwikkelperspectief voor alle aggregatieniveaus van secundair datagebruik.

In het vervolg van deze uitwerking analyseren we hoe hergebruik binnen twee succesvolle concepten (KIK-V en GERDA) bijdraagt aan realisatie van het landelijk dekkend netwerk voor secundair datagebruik. Daarbij zien we verschillen in de fase van implementatie en opschaling van de verschillende initiatieven. KIK-V bevindt zich in de fase waarbij structurele borging en inbedding kernthema zijn. GERDA is nog verder aan het opschalen in de regio/toepassing evenals de inbedding in de landelijke architectuur voor secundair datagebruik aan het verkennen.

Voorafgaand aan de analyse worden GERDA en KIK-V functioneel verder uitgewerkt en vergeleken.

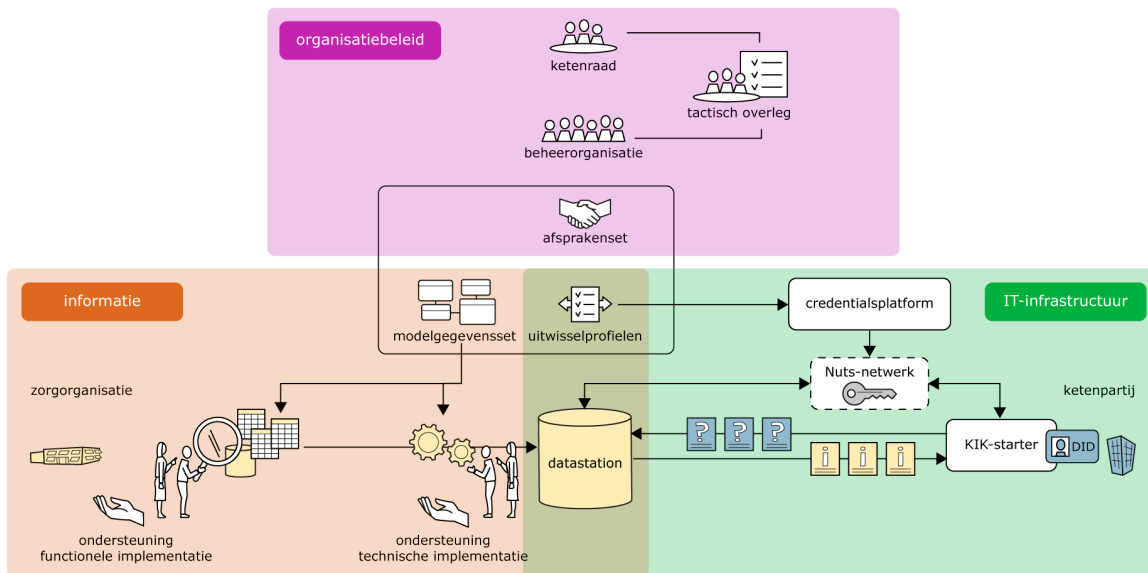
Vergelijking KIK-V en GERDA

Op basis van de analyses (zie bijlage) van KIK-V en GERDA kan het volgende overzicht gemaakt worden met betrekking tot de hoofdactiviteiten:

KIK-V

- Vertrouwensrelatie: KIK-V faciliteert bevestigingen van een landelijke ketenpartij – in de huidige fase Intramuraal Ouderenzorg – er is dus geen directe vertrouwensrelatie
- Geadresseerde: het betreft altijd vragen aan één organisatie– soms uitgesplitst per vestiging (individueel verzoek of in bulk)
- Analyseruimte: betreft altijd gevalideerde bevestigingen – met als gevolg een langer governance proces om de vraag te formuleren en formaliseren
- Data-integratie: data-integratie is enkel een vraagstuk als de bevestiging meerdere bronsystemen binnen één organisatie vraagt.

Deze kenmerken vertalen zich naar onderstaande high-level architectuur. Onder 'organisatiebeleid' worden in de ketenraad afspraken gemaakt over de bevestigingen die landelijke entiteiten mogen stellen aan individuele organisaties (zogenaamde gevalideerde bevestigingen). Deze worden vastgelegd in een afsprakenet. De individuele organisatie stelt de informatie die antwoord geeft aan de bevestigingen beschikbaar in een modelgegevensset. In de IT infrastructuur worden rechten aan vraagstellers en bevestigden toegekend in het Credentialsplatform. Via het Nuts-netwerk worden deze rechten gecontroleerd. Het datastation is de plek waar de brondata op basis van ontologie wordt ontsloten.

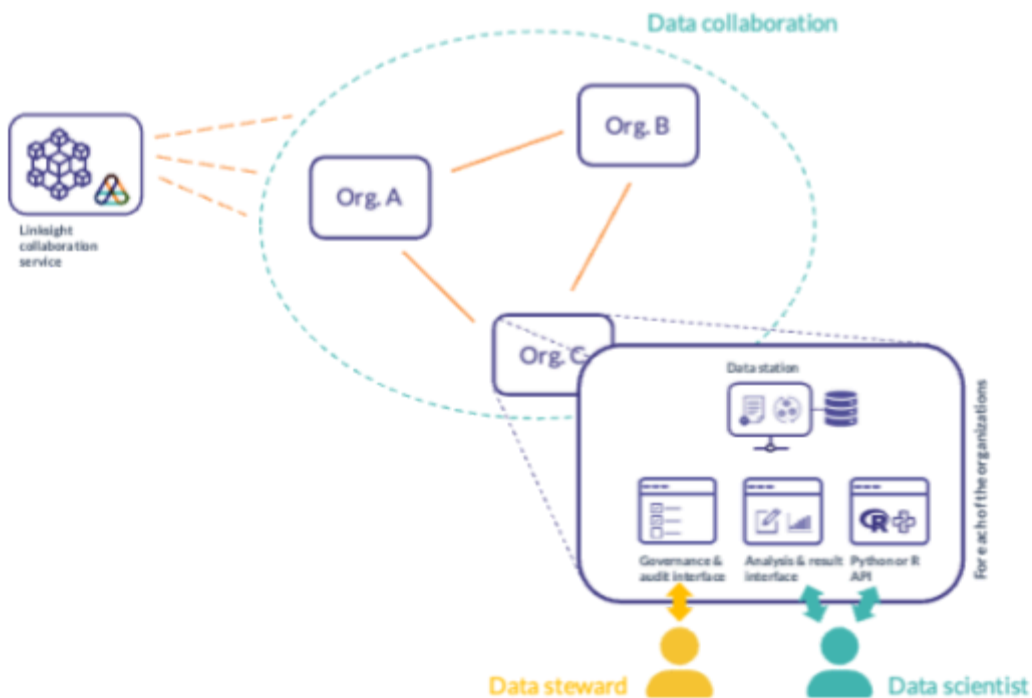


Figuur 1: Samenhang tussen de verschillende KIK-V onderdelen

GERDA

- **Vertrouwensrelatie:** GERDA faciliteert bevragingen van de regio, verenigd in een gezamenlijke datasamenwerking – er is sprake van een directe vertrouwensrelatie
- **Geadresseerde:** GERDA faciliteert bevragingen aan meerdere organisaties in de datasamenwerking
- **Analyseruimte:** Het betreft exploratieve en gerichte bevragingen
- **Data-integratie:** is noodzakelijk tussen verschillende (regionale) organisaties en mogelijk ook binnen één organisatie

Deze kenmerken vertalen zich naar onderstaande high-level architectuur. De regio start een datasamenwerking waarbij regionale partijen met elkaar aan datavraagstukken werken. In de datasamenwerking maken de data stewards onderling afspraken over het gebruik van de beschikbaar gestelde informatie door middel van het stellen van governance rules. Hierna kunnen de data scientists van de aangesloten organisaties met elkaar – binnen de gestelde afspraken – samenwerken aan datavraagstukken. De data wordt bevestigd via een data station in het eigen netwerk van de organisatie.



Figuur 2: Data scientist en data steward geplaatst op het Linkisight platform

Vergelijking

Op basis van de vraagstelling hebben we de onderdelen van KIK-V en GERDA opgedeeld in twee thema's: governance en techniek. Per onderdeel analyseren we verschillen en overeenkomsten; dit leidt tot conclusies per onderdeel die in het vervolg van dit paper worden uitgewerkt.

Governance			
Onderwerp	KIK-V	GERDA	Conclusie
<u>Vertrouwensrelatie en Geadresseerde</u>			
Inzicht vragen	KIK-starter stelt ketenpartijen in staat om gevalideerde bevestigingen te doen	Inzichtvragers melden zich bij data steward (bij nieuw inzicht) en data scientist (bij ontwikkeling en bestaand inzicht)	Binnen GERDA melden inzichtvragers zich bij data stewards en is een KIK-starter platform niet noodzakelijk. Op termijn is een generieke bevestigingsmodule zinvol voor

			inzichtvragers buiten de regionale datasamenwerking.
Toekennen en publiceren van rechten	Credentialsplatform geeft rechten uit een vraagstellers en bevrageden	Data scientists krijgen rechten voor analyses o.b.v. governance rules (door data stewards bepaald); Elke organisatie geeft eigen data steward en data scientist rechten	Voor vragen van buiten de regio (sectoraal of landelijk) is het gewenst om ook rechten te kunnen toekennen en publiceren voor partijen buiten de compute-groep. Daarbij verwachten we op termijn een generieke functie die datawerkplaatsen kunnen aanroepen.
<u>Analyseruimte en Data-integratie</u>			
Vraag specificeren	In het ontwikkelplatform en publicatieplatform worden afsprakensets beschikbaar gesteld die volgen uit de gemaakte afspraken	Data-vragen zijn exploratief of herhalend van karakter gekoppeld aan doelbinding - die geresulteerd heeft in een separate compute group (door data steward bepaald)	Ontwikkel- en publicatieplatform zijn op termijn wenselijke oplossing voor hergebruik voor GERDA. Mogelijkheden voor exploratieve vragen zijn op termijn mogelijk wenselijk voor KIK-V. KIK-V functioneert op basis van vastgestelde ontologie.

Techniek			
Onderwerp	KIK-V	GERDA	Conclusie
<u>Vertrouwensrelatie en Geadresseerde</u>			
Controleren van rechten	Nuts-netwerk controleert de rechten van vraagstellers en bevrageden	Linksight software controleert rechten en bevragingen binnen de governance rules	Voor het controleren van identificatie en rechten van personen en organisaties moet GERDA op basis van Nuts bovengenoemde generieke functie kunnen aanroepen. Primaire toepassing lijkt bij bevragingen buiten GERDA (maar kan ook user en organisatiemanagement binnen GERDA op termijn versimpelen).
<u>Analyseruimte en Data-integratie</u>			
Toegankelijk maken van data	Binnen KIK-V stellen de aangesloten partijen data beschikbaar vanuit hun eigen omgeving voor (vooraf) gespecificeerde vragen op basis van een ontologie	In GERDA stellen de aangesloten partijen data beschikbaar vanuit hun eigen omgeving voor exploratieve vragen zonder vooraf vastgesteld informatiemodel	Zowel GERDA als KIK-V gebruiken data uit onderliggende lokale bronsystemen van individuele organisaties. Dit biedt mogelijkheden om dit te stroomlijnen, middels één gezamenlijk technische inrichting op basis van een Datastation-as-a-Service en (deels) een uitbreiding van de gebruikte ontologie Beide concepten gaan uit van data soevereiniteit bij individuele organisaties.

De volgende twee hoofdstukken gaan nadrukkelijk in op de beantwoording van beide onderzoeksvragen. Daarbij worden de thema's Governance en Techniek nader verdiept, maar worden bovenal aanbevelingen gedaan hoe te werken aan een hergebruik van functionaliteiten.

Governance

Binnen GERDA melden regionale inzichtvragers zich bij data stewards zodat er vervolgens gewerkt kan worden aan een inzicht (nadat de governance rules en doelbinding zijn bepaald). De inzichtvragers werken dus niet in GERDA - de data stewards en data scientists nadrukkelijk wel. Hierdoor is er geen noodzaak voor een 'voorkant' voor inzichtvragers - zoals binnen KIK-V gefaciliteerd met de KIK-starter. Indien er bevestigingen aan GERDA van buiten de datasamenwerking gaan komen - dat is nu nog niet het geval, maar is wel de verwachting - gaat die behoefte er wel komen. In dat geval is een generieke bevestigingsmodule zinvol voor inzichtvragers buiten de regionale datasamenwerking.

Beide concepten omarmen voor wat betreft data-opslag en applicatieve functies op federatief gedachtegoed. In GERDA is ook de governance gefedereerd ingericht door in de datasamenwerking gezamenlijk governance rules op te stellen en hierover te stemmen. Analyse door data scientists kan enkel plaatsvinden als alle data stewards in de compute group consensus hebben over de geldende analyse-afspraken. Dit faciliteert soevereiniteit voor de data stewards namens hun organisatie op het gebied van gebruik van de data voor analysedoeleinden. Onderdeel van de analyse-afspraken zijn ook instellingen voor geautomatiseerde checks op onbedoelde onthullingen, ook wel statistical disclosure control (SDC) genoemd, zie kader.

Statistical Disclosure Control (SDC)

Bij zowel KIK-V als GERDA kunnen in principe alleen *aggregerende* vragen gesteld worden. Dat betekent dat alleen vragen gesteld kunnen worden waar bijvoorbeeld om de berekening van een gemiddelde of een maximum gevraagd wordt. Dat biedt al enige vorm van bescherming tegen onbedoelde onthulling. Echter, als het resultaat van een vraag maar over een kleine groep mensen gaat, dan kan toch nog wel wat geleerd worden over deze groepen mensen. Om dat te voorkomen kan geautomatiseerde statistical disclosure control (SDC) ingezet worden om het risico hierop te verkleinen. Een voorbeeld van een eenvoudige SDC controle is een check dat voor een resultaat data van minstens 20 personen is meegenomen. Dat deze check gedaan moet worden met welke grens (in dit geval 20), en wat de vervolgactie is, is weer een voorbeeld van een SDC governance regel zoals die binnen GERDA ingesteld kan worden.

In een ander traject is verkend hoe de SDC techniek uit het Linksight product ook voor KIK-V bevestigingen gebruikt kan worden om het risico op onbedoelde onthulling te verkleinen. Het resultaat was een proof-of-concept waarmee aangetoond is hoe een "onthullingsmitigatiemodule" met SDC methodes geplaatst kan worden tussen een DaaS en de onderliggende database (RDF triple store).

Tevens volgt er uit de samenwerking GERDA behoefte aan gezamenlijke afspraken over gevraagde inzichten, evenals de ethische aspecten die met deze beantwoording samenhangen. In KIK-V lijken ethische vraagstukken minder aan de orde.

Binnen GERDA is er sprake van data-integratie op verticaal gepartitioneerde datasets (domeinoverstijgende data-analyse). Het betreft veelal andere soorten bronnen (o.a. gezondheids-, ziekenhuis-, wijkverpleging- en huisartsen-data) waardoor de noodzaak van uniforme notatie 'over deze bronnen heen' niet primair noodzakelijk is. De data scientists hebben inzage in elkaars databeschrijvingen. Binnen KIK-V is de afspraken set een resultante van een centraal georganiseerde ketenraad. Hierna is controle op de afspraken set en ontologie 'slechts' een kwestie van technische implementatie.

KIK-V werkt vanuit een ontologie die via een OTAP beschikbaar gesteld wordt. Hoewel GERDA minder noodzaak heeft - en ook zal houden gezien haar type bevragingen - is het wel nuttig om de ontologie van KIK-V te kunnen gebruiken. Denklij is om liever landelijk gedefinieerde velden te gebruiken dan zelf velden te definiëren.

Binnen KIK-V is er al eerder noodzaak geweest om ook op het gebied van rechten gebruik te maken van een federatief netwerk. Daarom zijn bevragingen door middel van controle van rechten in het Nuts-netwerk beproefd. Ook GERDA voorziet op termijn die noodzaak (als er bevragingen buiten de datasamenwerking gaan komen).

Hergebruik binnen KIK-V en GERDA

Voorstel is om te beproeven hoe een gevalideerde bevraging van KIK-V aansluit op de governance-inrichting van een bevragingen in GERDA. Idee is om de zorgorganisaties Marga Klompé en Sensire individueel te laten bevragen en een separate set governance rules voor KIK-V vanuit het Linksight-product inrichten conform die bevraging. Deze governance rules zullen volgen uit de afspraken set en geoperationaliseerd worden in de governance rules KIK-V. De landelijke afspraken zijn leidend voor deze bevragingen. Hierdoor zullen deze governance rules 'slechts' geïmplementeerd worden en inzage gegeven worden aan de data stewards.

Formaliseren van de gevalideerde vraag vindt daarbij nog steeds plaats bij KIK-V. Naar verwachting ontstaat er een herbruikbare set governance rules voor gevalideerde bevragingen vanuit KIK-V en voor de zorgorganisaties een herbruikbare en herkenbare inrichting voor secundair datagebruik. Dit vraagt een losgekoppelde governance functionaliteit in het Linksight-product en idealiter controle van rechten op basis van Nuts-netwerk. Ter illustratie, enkele voorbeelden van deze governance rules zijn:

- Enkel de toegestane ketenpartij mag gevalideerde bevragingen initiëren
- Van de toegestane ketenpartij wordt de digitale handtekening van de organisatie en betreffende rol van de persoon gecontroleerd (geëffectueerd door NUTS-netwerk)
- Bij een KIK-V bevraging moet de datastructuur voldoen aan de ontologie passend bij die bevraging (geëffectueerd door Daas)
- Op de resultaten wordt geautomatiseerde SDC toegepast met nader te bepalen instellingen
- Resultaten worden enkel teruggegeven als de N voor de bevroegde scope groter is dan vast te stellen aantal (bijvoorbeeld N groter dan 20)
- Resultaten worden vrijgegeven na een manuele outputcontrole door de bevroegde organisatie
- De set met regels is 1 jaar geldig en moet daarna opnieuw goedgekeurd worden door de data steward van KIK-V*

* Alle data stewards van de organisaties valideren na 1 jaar de audit log op eigenlijk gebracht en toepasselijk van de gestelde rules.

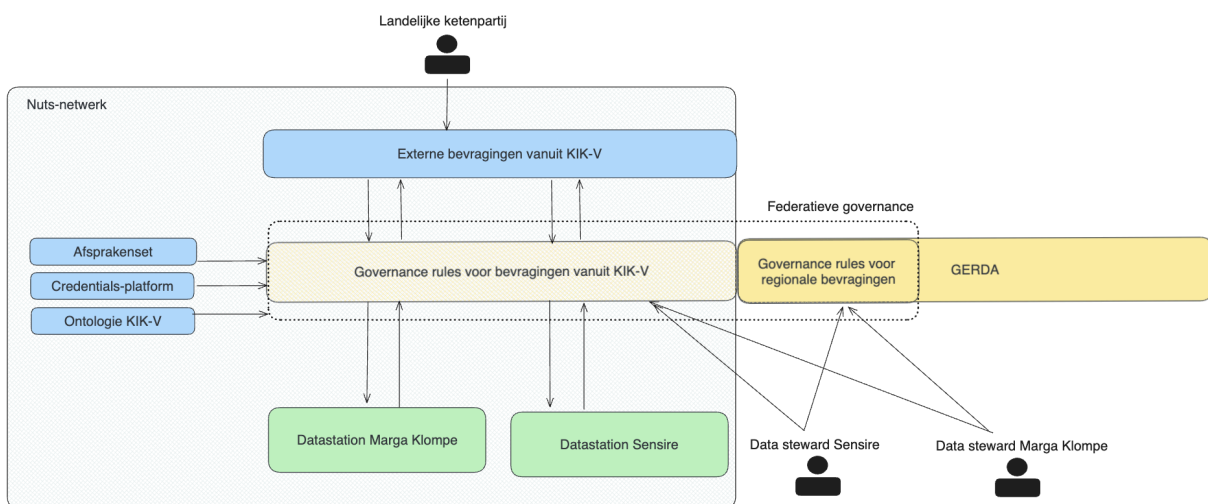
Voor het opzetten van een schaalbare federatieve governance zal het framework in het [whitepaper van het CoE-DSC](#) als basis gehanteerd worden.

Aandacht is nodig hoe bevestigingen van buiten de datasamenwerking opgevraagd kunnen worden. Daarvoor zit nu functionaliteit in de KIK-starter, maar dit zou ook op termijn een generieke bevestigingsmodule kunnen worden in/aan het Linksight-product voor bevestigingen buiten de datasamenwerking. Tevens zal er nagedacht moeten worden over de rol van de data steward vanuit KIK-V. Denklijjn is dat de vertrouwenslaag op basis van het Nuts-netwerk wordt ondersteund.

Hierdoor ontstaat er een universele rule engine waarin governance rules (onder welke condities mag de data geanalyseerd worden? Welke ontologie is van toepassing voor deze bevestiging?) geïmplementeerd worden. Op termijn is denkbaar dat in deze rule engine ook rules over credentials, ongewenste onthulling (o.a. statistical disclosure control) en datakwaliteit een plek krijgen. Daarbij zijn er landelijke rules - die simpelweg 'overorven' worden - en rules waar afstemming over noodzakelijk is. Hoe deze afstemming zelf in te regelen is ook nog een aspect waar verkenning en experimenten noodzakelijk zijn, met name met oog op schaalbaarheid. Een eerste aanzet over hoe de governance op te schalen bij groter wordende partijen in een datasamenwerkingen is beschreven in een recent [whitepaper van het CoE-DSC](#).

Meerwaarde is dat de governance afspraken herbruikbaar worden en individuele organisaties niet voor bevestigingen uit elk aggregatieniveau hun eigen inrichting en afspraken kennen. Hierdoor ontstaat er een herbruikbare federatieve governance, die data stewards vanuit hun organisatie ook op het gebied van governance soeverein maakt.

Een eerste schets van de opzet ziet er als volgt uit:



Techniek

In de voorgaande onderdelen van dit whitepaper hebben we gekeken naar de bevragingen en de governance van GERDA en KIK-V. In dit hoofdstuk onderzoeken we welke mogelijkheden er zijn om (onderdelen) van de techniek uit beide oplossingen te hergebruiken.

Thamar Brouwer heeft namens het Zorginstituut Nederland een kort whitepaper geschreven: "[Herbruikbaarheid en schaalbaarheid platformen KIK-V x GERDA](#)". In dit whitepaper is gekeken naar de het Credentialsplatform, KIK-Starter en Datastations en zijn een aantal aanbevelingen en scenario's geschetst. Op basis van dit whitepaper werken we voor techniek de volgende onderdelen verder uit in dit document: (1) uitbreiding DaaS functionaliteiten en (2) gebruik Semantic Web Technologie. In het whitepaper komt ook het NUTS-netwerk aan de orde; dat is in de voorgaande hoofdstukken (Bevragingen en Governance) al aan de orde geweest

Uitbreiding Datastation-as-a-Service functionaliteit

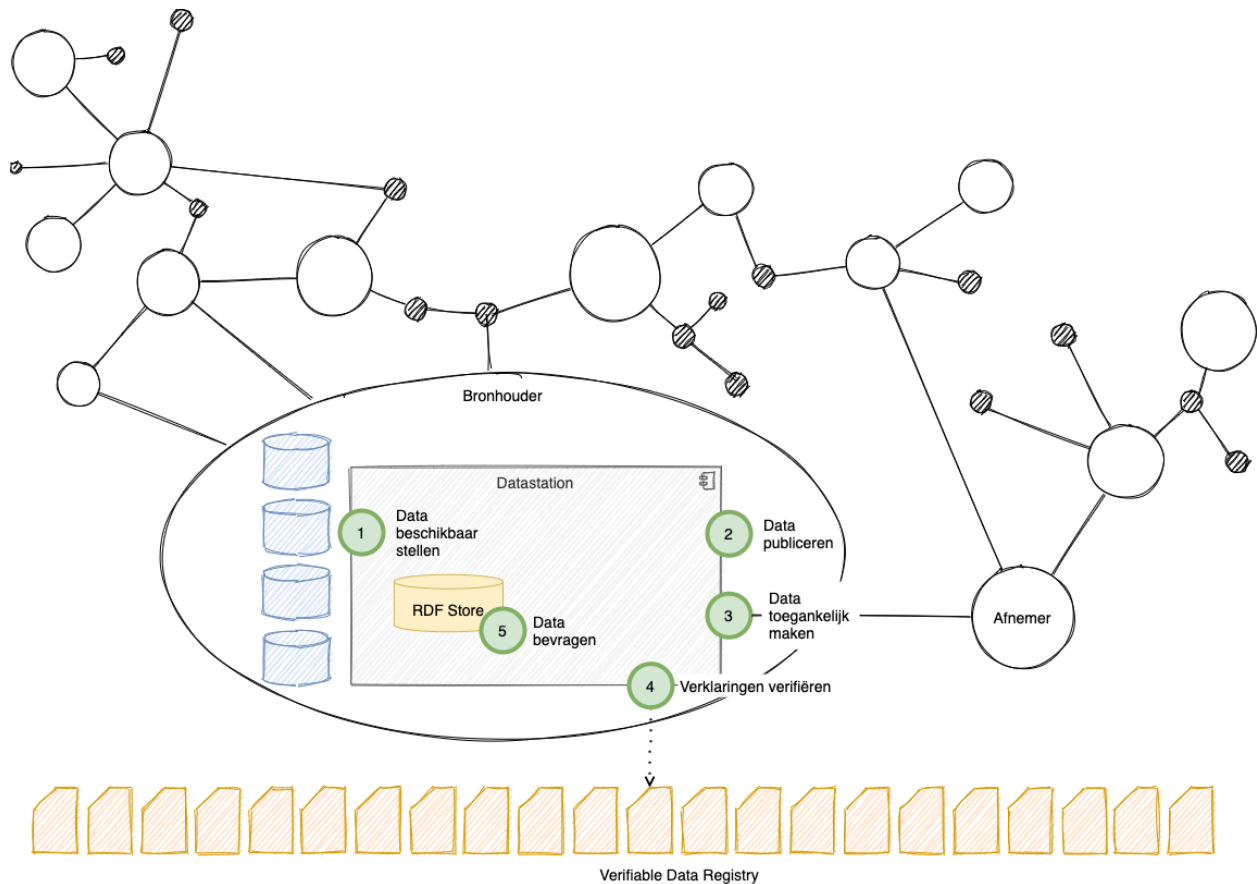
KIK-V en GERDA leunen zowel qua governance als qua techniek sterk op federatief gedachtegoed. In beide gevallen uit zich dit door te leunen op een IT Infrastructuur waarbij de data middels datastation bij de bron (bevraagde) geraadpleegd kan worden.

Binnen GERDA en KIK-V ligt de focus op 'hergebruik' van bestaande data (ook wel: secundair datagebruik). De data die wordt hergebruikt is afkomstig uit de 'primaire data' van bronsystemen binnen de organisatie (bijv. systemen voor personeelsadministratie, financiële administratie of patiënten/dossier-informatie). Binnen de opzet van een gefedereerde data-infrastructuur is het uitgangspunt 'data bij de bron' en replicatie van data te voorkomen.

Voor het beschikbaar stellen van data binnen KIK-V wordt gebruikgemaakt van een FAIR datastation. Deze wordt onder de naam 'Datastation-as-a-Service' (DaaS) geleverd door verschillende leveranciers, zoals NLcom, SDBGroep, Bince, Axians/OCS of Esculine/Tribe. Twee laatstgenoemde zijn samenwerkingsverbanden. Onderdeel van deze DaaS zijn het geautomatiseerd beschikbaar stellen van informatie via verschillende koppelvlakken met bronsystemen.

Binnen GERDA is (nog) geen DaaS aanwezig, maar wordt data beschikbaar gesteld via de Linksight nodes. Dat gebeurt op dit moment veelal nog handmatig via exports uit bronsystemen met bijvoorbeeld CSV bestanden die (via een lokale upload-functie) beschikbaar worden gesteld aan de lokale Linksight node die het vervolgens versleuteld weer beschikbaar maakt binnen het MPC-netwerk.

De architectuur van een DaaS is gebaseerd op de architectuur van een FAIR datastation, zoals gespecificeerd door ZIN (zie *bijlage 3 bij Meervoudig Onderhandse Aanbesteding: Datastation as a Service van Programma KIK-V, Zorginstituut Nederland*)



Een DaaS kan als centraal knooppunt opereren binnen een organisatie om data beschikbaar te stellen voor meerdere toepassing van secundair datagebruik. Op deze manier is er binnen een organisatie slechts één platform nodig om meerdere toepassingen te kunnen bedienen. Een dergelijke uniforme toepassing van een DaaS maakt het mogelijk om data ter beschikking te stellen voor zowel bevragingen van systeempartijen (KIK-V), regionale netwerken (GERDA) en andere toepassingen.

Om een universele(re) DaaS mogelijk te maken, kan de huidige specificatie¹ van de DaaS uitgebreid worden met een functie die het mogelijk maakt om aanvullende toepassingen (zowel aan de organisatie-kant als aan de kant van de bevrager) te laten draaien binnen een DaaS. Dit kan in de vorm van een '3rd party module'. Deze module kan bijvoorbeeld een container zijn waarbij data die ontsloten wordt door de kernfunctionaliteit van de DaaS, beschikbaar wordt gemaakt aan de module, die het vervolgens weer beschikbaar maakt voor een toepassing. Voorbeeld hiervan kan een MPC-module (voor het communiceren van een DaaS met een MPC-netwerk) zijn of een FHIR module om de koppeling ook met primaire (zorg)gegevens te maken.

¹ Op basis van:

https://drive.google.com/file/d/186btmaD91b0UldzIY2fKnLYH2_KseBlk/view?usp=drive_link
https://drive.google.com/file/d/1XVOI6_UkfK1EkZtrhkZ88h5npHFDZxkM/view?usp=drive_link

Gestructureerd vs. exploratief informatiemodel

De data die wordt aangeboden binnen KIK-V is sterk gestructureerd en gebaseerd op een ontologie; de data die wordt aangeboden binnen GERDA is exploratief en niet gebaseerd op een gestructureerd informatiemodel. Het voordeel van een gestructureerd datamodel is o.a. een zeer duidelijke en vergelijkbare uitkomst tussen verschillende momenten (voor bijvoorbeeld benchmarking) waarbij de kans op interpretatieverschillen geminimaliseerd is. Het voordeel van een exploratief model is dat op zeer snelle wijze (mogelijke) verbanden en relaties binnen een dataset kunnen worden onderzocht voor betere begripsvorming of ad hoc vraagstellingen.

Het beschikbaar stellen van data via één (gezamenlijk) DaaS maakt het mogelijk zowel een exploratief datamodel te gebruiken, als een gestructureerd informatiemodel. Echter, voor KIK-V is een DaaS momenteel zo geconfigureerd dat de RDF-data die beschikbaar is gesteld in de DaaS geplot is op het kunnen beantwoorden van reeds gevalideerde vragen. Er is dus een toets geweest op de data die beschikbaar is. Bij meer exploratieve vragen wordt er wellicht meer verwacht van de kwaliteit en vooral de data-omvang in de RDF-datasets. Er zijn dus structurele implicaties van een breder hergebruik van de DaaS indien de vragen exploratief zijn.

Gebruik gezamenlijke ontologie

Vanuit de huidige oplossing in GERDA / Linksight (gebruik maken van CSV exports) kan onderzocht worden op welke wijze exploratieve vragen die kansen bieden voor breder gebruik (bijv. meerdere regio's of een herhalende vraag) kunnen worden doorontwikkeld naar RDF op basis van de Ontologie Nederlandse Zorg. GERDA kan meer richting het gebruik van een ontologie gaan waar dit mogelijk is. Er bestaat echter geen universele ontologie en er zal tevens ruimte moeten blijven voor het exploratieve karakter van de onderzoeksvragen.

Voor de data-uitvraag via een DaaS is op dit moment een [SparQL](#) query nodig op basis van een ontologie. De KIK-V ontologie is specifiek ontwikkeld voor de Nederlandse Zorg ("Ontologie Nederlandse Zorg") waarbij vooral de langdurige zorg centraal heeft gestaan. Wanneer een meer generieke functionaliteit van een DaaS gewenst is, die op een vergelijkbare manier werkt, is doorontwikkeling van deze KIK-V ontologie nodig.

Snelheid van beantwoording

Naast de doorontwikkeling van de ontologie is een ander aandachtspunt de snelheid van beantwoording. Een vraagstelling via een DaaS bij KIK-V gebeurt nu asynchroon. Dit kan betekenen dat in de praktijk er (ondanks de ontologie en deels geautomatiseerde databeschikbaarheid) toch dagen of weken tussen het stellen van de vraag en de beantwoording kunnen zitten. Bij KIK-V moeten antwoorden door de 'verifier/zorgorganisatie' worden vrijgegeven; terwijl er bij GERDA een

governance-netwerk is gevormd waarbij op voorhand is aangegeven welke ruimte er is voor het stellen van vragen (en in principe geen aparte bevestiging nodig is).

Bij de huidige GERDA oplossing kan alleen beschikbare data (inzichtelijk via de catalogus) worden gequeryed, maar is de beantwoording vrijwel realtime. De huidige inrichting van DaaS is voornamelijk asynchroon, wanneer de functionaliteit van de DaaS wordt uitgebreid zoals hiervoor beschreven met 3rd party modules en realtime exploratieve vragen heeft dit ook impact op de (technische) ontwikkeling van de DaaS.

Volwassenheid in datagedreven werken

Het specificeren van de vraag binnen KIK-V volgt een sterk gestructureerd proces, terwijl binnen GERDA het exploratieve karakter van vragen centraal staat. Om dit meer op elkaar af te stemmen (dichter bij elkaar te brengen) is volwassenheid met betrekking tot datagedreven werken nodig zijn van de afnemende partijen: *waarom wordt een vraag gesteld en welke verwachtingen of vereisten zijn er over de beantwoording?*

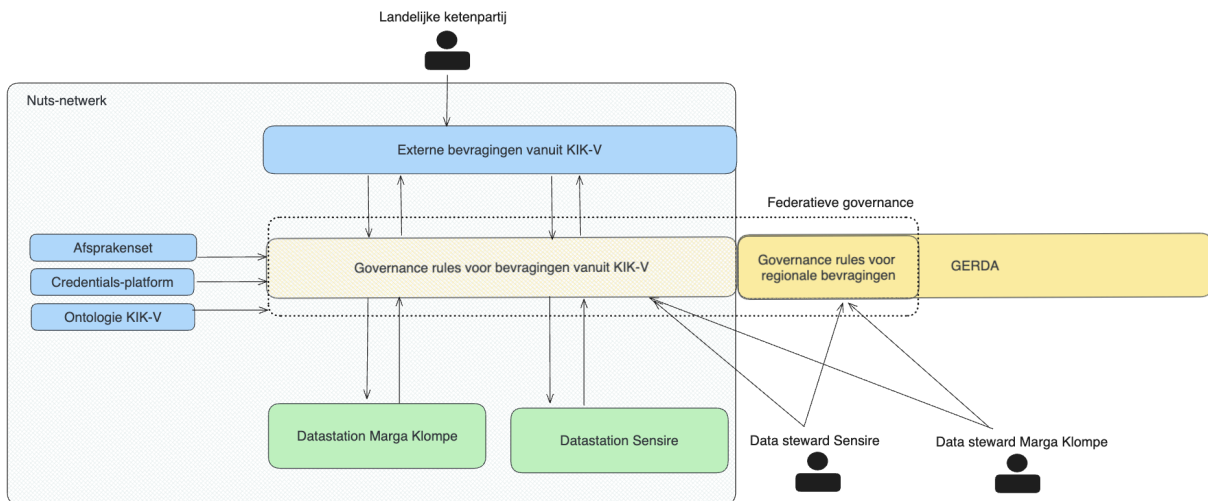
Aanbevelingen en vervolg

Uit bovenstaande analyse volgen de onderstaande aanbevelingen.

Federatieve governance

Onderzoek mogelijkheid voor ontwikkeling van federatieve governance binnen KIK-V en kan er synergie gecreëerd worden door hierbij gebruik te maken van (of integreren met) onderdelen van de governance zoals nu ontwikkeld binnen GERDA.

Globale schets ziet er als volgt uit:



Universele DaaS

Onderzoek met DaaS leveranciers mogelijkheid voor het ontwikkelen van een universele DaaS:

- Aanbieden van exploratieve en gestructureerde vragen
- Ontwikkelen *3rd party modules* binnen een DaaS voor bijv. aansluiting op MPC of FHIR
- Faciliteer de transitie (bijv van CSV naar hybride informatiemodel) vanuit RDF

Borging

KIK-V heeft een professionele basis voor een OTAP; daar kan bij GERDA ook (meer) gebruik van gemaakt worden.

Ontologie Nederland-breed

Onderzoeken mogelijkheden voor doorontwikkeling ontologie Nederland-breed om aan te sluiten op het type exploratieve vragen dat vanuit GERDA gesteld wordt. Hierbij dient ook gekeken naar de beschikbare ontologieën en de aansluiting op de ZIB transitie zoals omschreven door Nictiz.