

PANDEMISCH PARAAT door middel van herbruikbare data

Kennissynthese over stakeholders en databronnen
voor FAIR data en pandemische paraatheid

TNO 2024 R11031 – 24 oktober 2024

PANDEMISCH PARAAT door middel van herbruikbare data

Kennissynthese over stakeholders en databronnen voor
FAIR data en pandemische paraatheid

Auteurs	Prof. Dr. J. (Jildau) Bouwman, Dr. A. (André) Boorsma, C.A. (Cornelis) Bouter, Msc, Dr. S. (Sanne) Kuijper, R. (Remon) Dulos, ing., D.D. (David) de Best, Msc
Rubricering rapport	TNO Publiek
Titel	TNO Publiek
Rapporttekst	TNO Publiek
Aantal pagina's	54 (excl. voor, achterblad en bijlages)
Aantal bijlagen	5
Opdrachtgever	ZonMw
Projectnaam	ZonMw-FAIR dataset Pandemisch Paraatheid
Projectnummer	060.58978

Alle rechten voorbehouden

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van TNO.

© 2024 TNO

Executive summary

During the COVID-19 pandemic, tens of thousands of people died from the disease, and there was significant societal disruption. Even now, long-COVID patients are experiencing the prolonged effects of the illness. Society and the economy are still feeling the far-reaching consequences of the pandemic. Insights from data, such as medical records, vaccination data, and behavioral data, are essential to prevent a new pandemic or to control an existing one. The need for reusability of static, historical, retrospective data, but especially for accurate, current data or even real-life data, is of great importance for research purposes and policy decisions before, during, and after a pandemic.

On behalf of the Ministry of Health, Welfare and Sport (VWS), ZonMw is implementing the Knowledge Programme on Pandemic Preparedness. Part of this program is the question of how the availability and reusability of data can support pandemic preparedness. In this context, ZonMw has commissioned TNO to compile a knowledge synthesis that includes an inventory, analysis, and recommendations regarding the stakeholders and data sources that are important for the reusability of data for pandemic preparedness. The central question in this report is whether relevant data sources from healthcare, policy, and research are efficiently available for reuse, should a new pandemic strike the Netherlands.

The knowledge synthesis presented here consists of an overview of the stakeholders of health data and other relevant data sources and their importance in data availability in the event of a pandemic. This report also provides an overview of the data sources that are important for research and policy on pandemic preparedness and gives an indication of the availability and reusability of these data sources. Finally, recommendations are made to improve the availability and reusability of data for pandemic preparedness.

The main conclusion of this report is that the availability and especially the reusability of data from the examined data sources are not up to standard. This means that the Netherlands is not pandemic-prepared in terms of data organization. Various crucial data sources (e.g., social and societal data and data from long-term care) and stakeholders (e.g., patient organizations) receive insufficient attention and financial support to make data pandemic-prepared. The complexity of the task to make data available and reusable, the lack of data stewards, and the right incentives play a role in this. Data accessibility is also hindered by data-holding parties due to fear of losing control over the data. Additionally, legislation (GDPR) is often cited as an obstacle to data availability. Finally, the sources found that are relevant for pandemic preparedness mainly contain retrospective data, but few data sources provide current data, which is crucial for rapid responses during a pandemic.

The application of the FAIR principles is an important and efficient way to make relevant data available and reusable, thereby helping to prepare for a pandemic. The FAIR principles are a set of guidelines developed to facilitate the availability and reusability of data. This applies to both manual and automated data processes. In FAIR terms: data must be both human-readable and machine-readable. In general, the investigated data sources are poorly (automatically) findable and reusable. Although the FAIR principles are known to many of the investigated stakeholders, there is still much misunderstanding about what FAIR really means in practice and what is concretely needed to make data FAIR. The investigated tools to create FAIR data are not user-friendly enough and are often at too low

a TRL level. An advantage of FAIR is that it requires the use of standards but does not impose which standards must be used. This gives users a lot of flexibility. However, this sometimes leads to problems when different standards are used within a community or domain. A FAIR Implementation Profile (FIP) solves this problem by making a set of agreements with various stakeholders. For this report, a preliminary high-level FIP has been developed for pandemic preparedness, which needs to be expanded with FIPs for various required data types.

Another finding from this report is that there is currently no central portal in the Netherlands where all relevant data for Pandemic Preparedness can be found. Several initiatives such as Health-RI, HDAB-NL, and ODISSEI are now developing data portals. If FAIR is properly implemented in these initiatives and if they represent the data relevant to pandemic preparedness, then the Netherlands will be much better prepared for the next pandemic. The focus of these initiatives has so far been mainly on retrospective data. The release of current data, which is particularly interesting for current policy decisions, is being worked on in several other projects such as GERDA, Heracles, and Plugin. These infrastructures are a good complement to the portals currently being developed by Health-RI and HDAB-NL.

The main recommendation in this report is to establish a 'crisis' team for pandemic preparedness that focuses on improving the availability and reusability of data. This team should be composed of various experts such as legal experts, data stewards, and domain experts. The 'crisis' team should primarily facilitate the access to relevant data for pandemic preparedness by creating and maintaining a network of FAIR Implementation Profiles (FIPs). Furthermore, the 'crisis' team should actively seek out new relevant data sources and coordinate these well with similar international initiatives. Additionally, this team will be responsible for finding solutions to ethical/legal issues surrounding data availability and should collaborate with the obstacle-removal process¹ wherever possible.

Furthermore, the Ministry of Health, Welfare and Sport (VWS) is advised to develop a clear vision on data reusability for pandemic preparedness. This report can serve as a basis for this. This vision should clarify the roles and responsibilities of the parties managing data sources suitable for both the pandemic and non-pandemic phases. Additionally, the vision should provide clarity on the sources of funding, identify which data are essential for pandemic preparedness, and outline how legislation and regulations for data availability should be optimized. VWS should also take further leadership in developing a FAIR data infrastructure aimed at making data available for pandemic preparedness, including both retrospective data and real-time data suitable for policy-making. This can build upon and expand existing data infrastructures such as those of Health-RI and ODISSEI. For current data and real-time data, existing projects like the federated architectures of GERDA and Heracles could be utilized. Furthermore, the reusability of data and data linkages should be actively promoted. It is also essential to connect initiatives from key stakeholders, such as the GGD IV system to these infrastructures.

If data from various parties become functionally reusable in the future, it will stimulate research and thereby improve policy and care for infectious diseases, potentially preventing an infectious disease from developing into a pandemic. This would make the Netherlands optimally pandemic-prepared in terms of data. The recommendations in this report can be used by ZonMw to program research on pandemic preparedness, encourage other parties to further develop data sources, and provide policymakers with guidance on what is needed to

¹ <https://www.health-ri.nl/en/news/obstakel-verwijder-traject-health-ri-van-obstakels-naar-concrete-oplossingen-voor-de-komende>

make data specifically reusable for pandemic preparedness. Although the recommendations in this report are made in the context of pandemic preparedness, they are also important for the development of an efficient healthcare system, one of the core points of the Ministry of Health, Welfare and Sport (VWS).

Managementsamenvatting

Tijdens de COVID-19 pandemie zijn er tienduizenden mensen aan de ziekte overleden en is er sprake geweest van grote maatschappelijke ontwrichting. Nu nog ondervinden long-COVID patiënten de langdurige gevolgen van de ziekte. De samenleving en de economie ondervinden nog altijd de verstrekkende gevolgen van de pandemie. Inzichten uit data, uit bijvoorbeeld gegevens van medische dossiers, vaccinatiegegevens en gedragsgegevens zijn essentieel om een nieuwe pandemie te voorkomen of een bestaande pandemie te beteugelen. De behoefte aan herbruikbaarheid van statische, historische, retrospectieve data maar vooral van accurate, actuele data of zelfs real-life data is van groot belang voor onderzoeksdoeleinden en beleidsbeslissingen voor, tijdens en na een pandemie.

In opdracht van het ministerie van VWS voert ZonMw het Kennisprogramma Pandemische Paraatheid uit. Onderdeel van dit programma is de vraag hoe de beschikbaarheid en herbruikbaarheid van data, pandemische paraatheid kan ondersteunen. In dit kader heeft ZonMw TNO opdracht gegeven een kennissynthese op te stellen met daarin een inventarisatie, analyse en aanbevelingen met betrekking tot de stakeholders en databronnen die van belang zijn voor herbruikbaarheid van data t.b.v. pandemische paraatheid. De centrale vraag in dit rapport is of relevante databronnen uit de zorg, beleid en onderzoek efficiënt herbruikbaar zijn mocht een nieuwe pandemie Nederland treffen.

De hier gepresenteerde kennissynthese is opgebouwd uit een overzicht van de stakeholders van gezondheidsdata en van andere relevante databronnen en hun belang bij data beschikbaarheid in geval van een pandemie. Ook geeft dit rapport een overzicht van de databronnen die van belang zijn voor onderzoek naar en beleid voor pandemische paraatheid en geeft het ook een indicatie van de beschikbaarheid en herbruikbaarheid van deze databronnen. Tenslotte worden er aanbevelingen gegeven om de beschikbaarheid en herbruikbaarheid van data te verbeteren ten behoeve van pandemische paraatheid.

De belangrijkste conclusie van dit rapport is dat data beschikbaarheid en vooral herbruikbaarheid van data bij de onderzochte databronnen niet op orde is. Daarmee is Nederland wat betreft data organisatie niet pandemisch paraat. Verschillende cruciale gegevensbronnen (bijvoorbeeld sociale en maatschappelijke gegevens en data uit de langdurige zorg) en belanghebbenden (bijvoorbeeld patiëntenorganisaties) krijgen onvoldoende aandacht en financiële ondersteuning om data pandemisch paraat te maken. De complexiteit van de taak om data beschikbaar en herbruikbaar te maken, gebrek aan data stewards en de juiste incentives spelen hierbij een rol. Datatoegankelijkheid wordt ook geremd door data-houdende partijen vanuit angst om controle over de data te verliezen. Daarbij wordt wetgeving (AVG) regelmatig opgevoerd als belemmering voor data beschikbaarheid. Tenslotte bevatten de gevonden bronnen die relevant zijn voor pandemische paraatheid vooral retrospectieve data maar weinig databronnen ontsluiten actuele data terwijl die cruciaal zijn voor de snelle reacties tijdens een pandemie.

De toepassing van de FAIR principes is een belangrijke en efficiënte manier om relevante data beschikbaar en herbruikbaar te maken en kan daardoor helpen om voorbereid te zijn op een pandemie. De FAIR principes zijn een verzameling richtlijnen die zijn ontwikkeld om de beschikbaarheid en herbruikbaarheid van gegevens te vergemakkelijken. Dit geldt voor zowel manuele als geautomatiseerde data processen. In FAIR-termen: data moet zowel

mens als machine-leesbaar zijn. In het algemeen zijn de onderzochte databronnen slecht (automatisch) vindbaar en herbruikbaar. Hoewel de FAIR principes bij veel van de onderzochte stakeholders bekend zijn, is er toch nog veel onbegrip over wat FAIR in de praktijk nu echt betekent en wat er concreet voor nodig is om data FAIR te maken. De onderzochte tools om FAIR data te creëren zijn niet gebruikersvriendelijk genoeg en zijn bovendien vaak van een te laag TRL niveau. Een voordeel van FAIR is dat het gebruik van standaarden vereist maar niet oplegt welke standaarden moeten worden gebruikt. Dit geeft gebruikers veel flexibiliteit. Echter leidt dit soms tot problemen als er in een gemeenschap of domein sprake is van het gebruik van verschillende standaarden. Een FAIR implementatie Profile (FIP) lost dit probleem op door met verschillende stakeholders gezamenlijk een set afspraken te maken. Voor dit rapport is een voorlopige high-level FIP ontwikkeld voor pandemische paraatheid die moet worden uitgebreid met FIP's voor verschillende benodigde gegevenstypen.

Een andere bevinding uit dit rapport is dat het momenteel in Nederland ontbreekt aan een centraal portaal waar alle relevante data voor Pandemische Paraatheid vindbaar zijn. Een aantal initiatieven zoals Health-RI, HDAB-NL en ODISSEI ontwikkelen nu data portalen. Als in deze initiatieven FAIR goed wordt doorgevoerd en als ze de voor pandemische paraatheid relevante data representeren dan zijn we in Nederland al veel beter voorbereid op de volgende pandemie. De focus van deze initiatieven is tot nu toe met name gericht op retrospectieve data. Aan de ontsluiting van actuele data die vooral interessant zijn voor actuele beleidsbeslissingen wordt binnen een aantal andere projecten zoals GERDA, Heracles en Plugin gewerkt. Deze infrastructuren zijn een goede aanvulling op de portalen die momenteel door Health-RI en HDAB-NL worden ontwikkeld.

Het belangrijkste advies in dit rapport is het installeren van een 'crisis'-team pandemische paraatheid dat zich focust op de verbetering van beschikbaarheid en herbruikbaarheid van gegevens. Dit team zou moeten worden samengesteld uit verschillende expertises zoals juridische expert(s), data steward(s) en domein expert(s). Het 'crisis'-team zou vooral het ontsluiten van relevante data voor pandemische paraatheid door middel van het creëren en onderhouden van een netwerk van FIPs moeten faciliteren. Verder zou het 'crisis'-team actief op zoek moeten gaan naar nieuwe relevante databronnen en deze goed afstemmen met vergelijkbare internationale initiatieven. Daarnaast krijgt dit team de verantwoordelijkheid om oplossingen te zoeken voor ethische/juridische problemen rondom data beschikbaarheid en moet daarvoor waar mogelijk samen werken met het obstakel-verwijdertraject².

VWS wordt geadviseerd om een duidelijke visie te ontwikkelen op data herbruikbaarheid voor pandemische paraatheid. Dit rapport kan hiervoor als basis dienen. Deze visie moet duidelijkheid verschaffen over de rollen en verantwoordelijkheden van de partijen die databronnen beheren, die geschikt zijn voor de pandemisch en koude fase. Tevens moet de visie duidelijkheid geven over waar de financiering hiervoor vandaan moeten komen, welke data essentieel zijn om pandemisch paraat te zijn en hoe de wet- en regelgeving voor data beschikbaarheid moet worden geoptimaliseerd. VWS zou ook verdere regie moeten nemen op de ontwikkeling van FAIR data infrastructuur die gericht is op de ontsluiting van data voor pandemische paraatheid voor zowel retrospectieve data als voor beleid geschikte actuele data. Hierbij kan worden aangesloten en uitgebreid op data infrastructuur zoals die van Health-RI en ODISSEI. Voor actuele data zou gebruik kunnen worden gemaakt van bestaande projecten zoals de gefedereerde architecturen van GERDA en Heracles. Daarbij moeten de herbruikbaarheid van data en data koppelingen actief worden gestimuleerd.

² <https://www.health-ri.nl/en/news/obstakel-verwijder-traject-health-ri-van-obstakels-naar-concrete-oplossingen-voor-de-komende>

Het is ook essentieel om initiatieven van belangrijke stakeholders, zoals het GGD IV systeem, op deze infrastructuren aan te sluiten.

Als data bij verschillende partijen in de toekomst functioneel herbruikbaar worden dan zal dat onderzoek stimuleren en daarmee het beleid en zorg voor infectieziekten verbeteren en mogelijk voorkomen dat een infectieziekte zich ontwikkelt tot een pandemie. Dit zou Nederland op gebied van data optimaal pandemisch paraat maken. De aanbevelingen in dit rapport kunnen door ZonMw worden gebruikt voor het programmeren van onderzoek naar pandemische paraatheid, andere partijen kunnen aangemoedigd worden om databronnen verder te ontwikkelen en het geeft beleidsmakers houvast wat er nodig is om data specifiek herbruikbaar te maken voor pandemische paraatheid. Hoewel de aanbevelingen in dit rapport worden gedaan in het kader van pandemische paraatheid zijn ze ook van belang voor de ontwikkeling van een efficiënt zorgsysteem, een van de kernpunten van het ministerie van VWS.

Inhoudsopgave

Executive summary	3
Managementsamenvatting	6
Afkortingen	10
Termenlijst	11
1 Achtergrond.....	12
1.1 De waarde van herbruikbaarheid van data voor pandemische paraatheid	12
1.2 Introductie van de FAIR principes.....	13
1.3 Vraagstelling.....	15
2 Methode.....	16
3 Kennissynthese data- herbruikbaarheid voor Pandemische Paraatheid.....	19
3.1 Stakeholder en databronnen analyse.....	19
3.2 Vragenlijst en interviews: Identificeren van relevante stakeholders en bronnen	20
3.3 Relevante stakeholders voor pandemische paraatheid	22
3.4 Relevante databronnen voor pandemische paraatheid.....	28
3.5 FAIRness van de databronnen	33
3.6 Evaluatie van FAIRness van de databronnen	35
3.7 Data duplicatie	40
4 FAIR Research tools & services.....	41
4.1 FAIR maken van data	41
4.2 FAIR implementatie bronnen.....	44
4.3 FAIR assessment tools	45
4.4 FAIR Implementatie Profiel	46
4.5 De FIP en FAIR assessment	46
5 Conclusies.....	48
5.1 Data beschikbaarheid	48
5.2 Organisatie van FAIR data voor Pandemische Paraatheid	49
5.3 FAIR data infrastructuur voor Pandemische Paraatheid	49
6 Adviezen en aanbevelingen	51
6.1 Algemene aanbevelingen pandemisch paraatheid.....	51
6.2 Aanbevelingen voor data beschikbaarheid	52
6.3 Aanbevelingen rondom de organisatie van FAIR voor pandemische paraatheid	53
6.4 Aanbevelingen rondom FAIR infrastructuur.....	54
7 Ondertekening.....	56
8 Bijlages.....	57
8.1 Bijlage 1 Vragenlijst.....	57
8.2 Bijlage 2 Stakeholders	60
8.3 Bijlage 3 Lijst van Databronnen voor pandemische paraatheid.....	62
8.4 Bijlage 4 PP-FIP Pandemische paraatheid FAIR implementatie profiel.....	64
8.5 Bijlage 5 lijst van aanvullende aanbevelingen	65

Afkortingen

AI	Artificiële Intelligentie
AVG	Algemene verordening gegevensbescherming
COVID-19	Coronavirus SARS-CoV-2
CBG	College ter Beoordeling van Geneesmiddelen
CBS	Centraal Bureau voor de Statistiek
DCAT	Data Catalog Vocabulary
DMP	Data Management Plan
DHD	Dutch Hospital Data
DPIA	Data Protection Impact Assessment
EHDS	European Health Data Space
EMA	European Medicines Agency
EPD	Elektronisch patiënten dossier
EZA	Expertisecentrum Zorgalgoritmen
FAIR	Findable, Accessible, Interoperable, Reuseable
FDO	FAIR Data Object
FDP	FAIR Data Point
FIP	FAIR Implementation Profile
FDR	FAIR Data Resource
GDPR	General Data Protection Regulation
GEN-AI	Generative artificial intelligence
GFF	GO FAIR foundation
GGD	Gemeentelijke Gezondheidsdienst
GHOR	Geneeskundige Hulpverleningsorganisaties in de Regio
HDAB	Health Data Access Body
Health-RI	Health- research infrastructure
ICTU	ICT-Uitvoeringsorganisatie
ICT	Information and communication technology
IKNL	Integraal Kankercentrum Nederland
LLM	Large Language Model
NIVEL	Nederlands Instituut voor Onderzoek van de Gezondheidszorg
ODISSEI	Open Data Infrastructure for Social Science and Economic Innovations
PGO	Persoonlijke gezondheidsomgeving
RDF	Resource Description Framework
RIVM	Rijksinstituut voor Volksgezondheid en Milieu
RSNN	Regulatory Science Network Netherlands
RDF	Resource Description Framework
TNO	Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek
TRL	Technology Readiness Level
URI	Uniform Resource Identifier
VWS	Ministerie van Volksgezondheid, Welzijn en Sport

Termenlijst

Data steward	Een persoon die verantwoordelijk is voor de kwaliteit en herbruikbaarheid van data en de communicatie daarover binnen een organisatie of gemeenschap
Databron	Een digitale plek waar data worden beheerd
FAIR principes	De internationale FAIR-principes zijn richtlijnen voor de manier van beschrijven, opslag, hergebruik en publicatie van wetenschappelijke (meta)data
Data beschikbaarheid	Data beschikbaarheid omvat het beschikbaar, bereikbaar en bruikbaar hebben van vastgelegde gezondheidsdata van de specifieke informatiebehoefte. Met als achterliggend doel deze in te zetten voor onderzoek, innovatie en beleid ten behoeve van preventie, zorg en welzijn.
Pandemische paraatheid	Goed voorbereid zijn op toekomstige pandemieën, als gevolg van uitbraken van micro-organismen
Provenance	De herkomst en kwaliteit van data. Dit loopt uiteen van informatie over de oorsprong van data, versies en de transformatie en bewerkingen die ze hebben ondergaan.
Meta-data	Data over data: meta-data is data die informatie bevat over andere data. Het fungeert als een beschrijving of context voor de hoofddataset en maakt het eenvoudiger om deze data te organiseren, te interpreteren, te beheren en te vinden.

1 Achtergrond

1.1 De waarde van herbruikbaarheid van data voor pandemische paraatheid.

In Amerika wordt momenteel de verspreiding van vogelgriep op veebedrijven nauwlettend in de gaten gehouden. Wetenschappers maken zich daarbij vooral zorgen over de transmissie van het virus naar mensen en overdracht tussen mensen. In April 2024 klaagden wetenschappers³ dat het Amerikaanse ministerie van landbouw niet de goede genetische data deelden die inzicht kunnen geven over de evolutie en de verspreiding van de H5N1 variant van dit virus. Deze data en andere data zijn voor dezelfde overheid nodig om beleid te voeren die mogelijk een nieuwe pandemie kan voorkomen. Vogelgriep kan potentieel leiden tot een nieuwe pandemische crisis. Niet alleen vogelgriep, maar ook andere influenza virussen, coronavirussen, onbekende pathogenen en anti-microbiologische resistentie vormen een risico voor een nieuwe pandemie. Nederland bevindt zich momenteel in de pre-pandemische fase (samen met de post-pandemische fase ook wel de koude fase genoemd). Tijdens de COVID-19 pandemie bleek hoe belangrijk snelle beschikbaarheid van accurate data was. Momenteel bevinden we ons in de post-pandemische fase van COVID-19 en is er grote behoefte is aan gezondheidsdata en andere data om inzicht te krijgen in long-COVID en de maatschappelijke gevolgen van de pandemie.

Tijdens de COVID-19-pandemie zijn tienduizenden mensen overleden en heeft de samenleving grote ontwrichting ervaren. Long-COVID-patiënten kampen nog steeds met de langdurige gevolgen van de ziekte. De samenleving en de economie voelen nog altijd de impact van de pandemie. Inzichten uit data, uit bijvoorbeeld gegevens van medische dossiers, vaccinatiegegevens en gedragsgegevens zijn essentieel voor beleidsbeslissingen om een nieuwe pandemie te voorkomen of die te beteugelen. Dit type data worden ook wel retrospectieve of historische data genoemd en worden gebruikt om trends in ziektepatronen te analyseren, oversterfte in een bepaalde periode te verklaren of de effectiviteit van beleidsmaatregelen te evalueren. Behalve deze data is er ook behoefte aan actuele of real-life data die van belang zijn voor actuele beleidsbeslissingen. Met dergelijke gegevens kan de ziekenhuisbezetting tijdens een pandemie live worden gemonitord en kunnen bepaalde trends vroegtijdig worden gesignaleerd. Historische data bieden inzicht in lange-termijn trends en helpen bij het maken van weloverwogen beleidsbeslissingen terwijl actuele data zorgverleners en beleidsmakers in staat stellen om snel te reageren op veranderende omstandigheden zoals tijdens een pandemie. De beschikking over actuele capaciteitsdata kan ook bijdragen aan efficiënte bedrijfsvoering en daarmee aan de efficiënte inzet van zorgpersoneel. De juiste data kan de overheid helpen beleid te ontwikkelen om een pandemie te voorkomen of de effecten van een pandemie op de samenleving te beperken.

Tot zover de theorie. De praktijk van data hergebruik is weerbarstig, met als meest aansprekend voorbeeld de commotie die ontstond rondom de geobserveerde oversterfte in de corona jaren. Voor het beantwoorden van vragen over de onverklaarde oversterfte

³ [Scientists say USDA is sharing too little data too slowly on H5N1 flu | CNN](#)

moesten data worden gekoppeld die waren verzameld door respectievelijk de GGD, RIVM en het CBS. Vanwege vooral (ervaren) problemen met de privacy wetgeving was het niet mogelijk om een snelle analyse uit te voeren. Een kleine drie jaar na de motie van Omzigt⁴ waarin werd opgeroepen tot onderzoek naar de onverklaarde oversterfte, is recent het rapport⁵ verschenen waarin het Nivel verschillende van de onderzoeksvragen beantwoord.

Behalve problemen met privacy wetgeving had de analyse ook sneller kunnen worden uitgevoerd als verschillende systemen waarin deze data werden verzameld beter op herbruikbaarheid waren ingericht. De hoofdonderzoeker van het Nivel rapport gaf in een gesprek aan dat de bevindingen uit het rapport worden gewantrouwd. Daarbij wordt vooral gewezen op de mogelijke onbetrouwbaarheid van de gebruikte data. Het herleidbaar vastleggen bij de bron met relevante meta-data (informatie over de data die beschrijft hoe, wanneer en door wie de data zijn verzameld) kan leiden tot meer transparantie over de data en daarmee tot meer vertrouwen.

De uitdagingen rondom beschikbaarheid en herbruikbaarheid van pandemische data geldt niet alleen binnen landen, maar ook over landsgrenzen heen. In alle gevallen geldt dat de data zelf op een gestandaardiseerde manier moet worden beschreven en met voldoende meta-data, om herbruikbaar te zijn.

In Nederland voert ZonMw in opdracht van het ministerie van VWS het Kennisprogramma Pandemische Paraatheid uit. Onderdeel van dit programma is de vraag of Nederland pandemische paraat is op het gebied van data: zijn de voor pandemische paraatheid relevante databonnen beschikbaar en herbruikbaar? In dit licht heeft ZonMw onderzoeksorganisatie TNO benaderd voor het opstellen van een kennissynthese met betrekking tot stakeholders en databronnen die van belang zijn voor een data herbruikbaarheid voor beleid, innovatie en onderzoek naar pandemische paraatheid. Het beschrijven en organiseren van deze data voor pandemische paraatheid kan gedaan worden volgens de FAIR principes. Dit zorgt uiteindelijk voor betere data beschikbaarheid en herbruikbaarheid van data. Indien gedurende de COVID-19 pandemie data van verschillende bronnen volgens de FAIR principes beschikbaar waren geweest had dat waarschijnlijk innovaties kunnen versnellen en tot betere en efficiëntere beleidskeuzes kunnen leiden.

1.2 Introductie van de FAIR principes

FAIR kan het best worden beschreven als een raamwerk van principes voor het beschikbaar en herbruikbaar maken van data. Het acroniem FAIR staat voor de vier principes Findable, Accessible, Interoperable and Reusable. In het raamwerk zijn onder de vier principes een aantal richtlijnen uitgewerkt die zorgen voor het verbeteren van beschikbaarheid en herbruikbaarheid van data, datasets en andere digitale objecten. De FAIR principes leggen een nadruk op het beschikbaar en (her)bruikbaar maken van data voor mensen, maar ook op de toepassing van data in geautomatiseerde processen. Data wordt dan vindbaar en leesbaar voor bijvoorbeeld het trainen en toepassen van artificiële intelligentie en dus voor machines. In dit rapport worden hiervoor de termen ‘mens-leesbaar’ en ‘machine-leesbaar’ gebruikt.

⁴ <https://www.parlementairemonitor.nl/9353000/1/j9vvi5epmj1ey0/vloep8orqza>

⁵ <https://www.nivel.nl/nl/publicatie/oversterfte-tijdens-de-covid-19-pandemie-onderzoek-onder-mensen-met-en-zonder-covid-19>

De **F** voor **Findable** (vindbaarheid) betekent dat zowel een mens als een machine kan vaststellen waar bepaalde data kan worden gevonden. Essentieel voor de vindbaarheid van data is gedetailleerde of rijke meta-data onder andere door te beschrijven welke eigenschappen data van een databron heeft. Meta-data beschrijft ook welke informatie er in de databron te vinden is, bijvoorbeeld of de databron data informatie heeft over het aantal vrije bedden, of over specifieke bloedwaarden. Hierdoor is voor anderen in te schatten of de databron relevant is, zonder dat de data ingezien hoeft te worden. Om databronnen geautomatiseerd vindbaar te maken betekent dat er gebruik wordt gemaakt van technische standaarden die een bepaald formaat voor data catalogi, een overzicht van databronnen, afdwingen.

Binnen het FAIR framework staat de **A** voor **Accessible** (toegankelijkheid). De onderliggende richtlijnen van dit principe vereisen dat het helder is voor wie, wanneer en waarvoor de data beschikbaar zijn. Deze informatie moet ook machine-leesbaar beschikbaar zijn om processen te kunnen automatiseren. Dit vereist ook passende authenticatie en autorisatie systemen. Voor gebruikers van data moet duidelijk zijn wie toegang tot de data kan krijgen (autorisatie), op welke manier en onder welke voorwaarden de toegang kan plaatsvinden. Deze informatie moet beschreven staan in een gestandaardiseerd communicatie protocol. In sommige gevallen, bijvoorbeeld als het gaat om gevoelige data kan het genoeg (lees FAIR genoeg) zijn om een e-mail adres of een telefoonnummer van een contactpersoon te noemen die meer over de toegankelijkheid van de databron kan vertellen. Deze informatie moet dan ook machine-leesbaar zijn, namelijk dat de bron niet zo maar toegankelijk is. De toegankelijkheid in FAIR vereist dat meta-data leesbaar moet blijven zelfs als de onderliggende data niet (meer) beschikbaar zijn. Het komt namelijk voor dat de meta-data van databronnen aanwezig is, maar de databron zelf vanwege verschillende redenen niet (meer) toegankelijk is of voor bepaalde partijen en/of doelstellingen. Voor de duidelijkheid, de A van FAIR betekent namelijk niet toegankelijk voor iedereen.

Het **Interoperable** (Interoperabiliteit) principe, de **I** binnen FAIR vereist dat parameters van data op een duidelijke en eenduidige manier voor mens en machine worden beschreven. Om goede interoperabiliteit te bewerkstelligen is het nodig om de data te beschrijven aan de hand van domein specifieke vocabulaires en ontologieën (zogenaamde semantiek), waardoor machines op basis van de gebruikte termen verschillende databronnen kunnen koppelen en mensen in staat zijn de data eenduidig te interpreteren.

Nadat de eerste drie principes goed ingeregeld zijn en data vindbaar en beschikbaar zijn, is het voor het **Reuseability** (herbruikbaarheid) principe, de **R** uit FAIR, van belang om ook de data daadwerkelijk herbruikbaar te maken. Hiervoor is het belangrijk dat er genoeg meta-data aanwezig zijn waarin wordt beschreven hoe en waarvoor data zijn verzameld zodat het ook helder is of de data bruikbaar is binnen de vraagstelling waarvoor de data wordt opgevraagd. In totaal zijn er 15 richtlijnen⁶ waaraan voldaan moet worden om data volledig FAIR te maken.

FAIR data is onderdeel van data management; het proces van het beheren van data. Dit proces begint bij een goede voorbereiding van software en hardware die benodigd zijn voor het vastleggen en behouden van data nodig voor onderzoek, beleid of gezondheidszorg. Data management behelst ook het maken keuzes van technische en semantische standaarden. De FAIR principes faciliteren dit proces.

⁶ <https://www.go-fair.org/fair-principles/>

1.3 Vraagstelling

De centrale vraag binnen het hier gepresenteerde onderzoek is of een selectie van relevante databronnen in de zorg, beleid en onderzoek herbruikbaar zijn, mocht een nieuwe epidemie of pandemie Nederland treffen. Bijvoorbeeld, voldoet data in de context van pandemische paraatheid aan FAIR, en zijn data daarmee pandemisch paraat? In de kennissynthese die volgt gaan we specifiek in op de volgende vragen; welke data zijn essentieel voor pandemisch paraatheid? In hoeverre voldoen deze data algemeen aan de FAIR principes? Welke stakeholders zijn of moeten betrokken worden bij het verzamelen van relevante data voor een nieuwe pandemie? Welke stakeholders zijn gebruikers van deze data? Welke obstakels staan data hergebruik in de weg? Hoe moeten de FAIR principes worden toegepast in de context van pandemische paraatheid? En wat moet er gebeuren om de herbruikbaarheid van data te verbeteren?

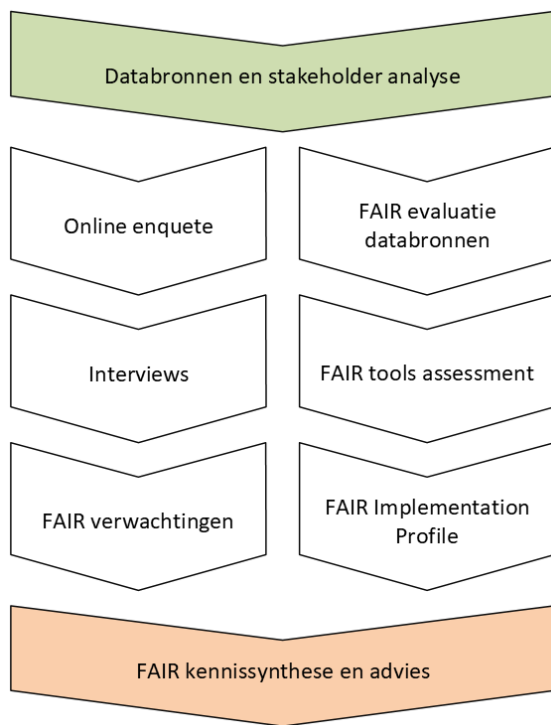
In het volgende hoofdstuk (hoofdstuk 2) wordt de methode toegelicht die is gebruikt voor de totstandkoming van deze kennissynthese. Daarna volgt een analyse van relevante databronnen en stakeholders, en bespreken we de vragenlijst en interviews die zijn gedaan in het kader van dit project (hoofdstuk 3). Hierna volgt een uiteenzetting van verschillende FAIR research tools en services (hoofdstuk 4), waarna we komen tot de conclusies (hoofdstuk 5) en tenslotte geven we een reeks aanbevelingen gericht op het verbeteren van het herbruikbaar maken data voor pandemische paraatheid (hoofdstuk 6).

2 Methode

De centrale vraag in dit rapport is of relevante databronnen uit de zorg, beleid en onderzoek, beschikbaar en herbruikbaar zijn mocht een nieuwe pandemie Nederland treffen. Om deze vraag te beantwoorden is een kennissynthese opgesteld met daarin een inventarisatie, analyse en aanbevelingen met betrekking tot de stakeholders en databronnen die van belang zijn voor de herbruikbaarheid van data t.b.v. pandemische paraatheid. Voor deze kennissynthese zijn de volgende werkzaamheden uitgevoerd (zie [Figuur 2.1](#)):

- 1) Deskresearch voor het identificeren van een selectie van relevante stakeholders en databronnen;
- 2) Uitvoeren van een enquête om de status van FAIR in het werkveld te inventariseren en onbekende stakeholders te identificeren;
- 3) Het uitvoeren van interviews en bijwonen van het symposium congres '[Oversterfte in tijden van corona](#)' voor het verkrijgen van diepgang wat betreft de ervaring met FAIR van een aantal stakeholders;
- 4) Evalueren van de FAIR status van verschillende databronnen en portals (interviews en desk research);
- 5) Inventarisatie en evaluatie van verschillende FAIR gerelateerde tools;
- 6) Ontwikkeling van een FIP voor pandemische paraatheid;
- 7) Opstellen van een lijst van aanbevelingen m.b.t. de herbruikbaarheid van data voor pandemische paraatheid.

Het uitvoeren van de kennissynthese is gestart met behulp van deskresearch voor de identificatie van relevante stakeholders en databronnen op het gebied van pandemische paraatheid. Daarbij is zoals beschreven in de opdracht, gefocust op Nederland en direct voor Nederland relevante internationale initiatieven. Hiervoor zijn onder andere websites van ZonMw, GLOPID-R, Health-RI en PubMed onder de loep genomen.



Figuur 2.1: Een schematische weergave van de uitvoering van de kennissynthese.

Om onbekende stakeholders te identificeren en de status van FAIR in het werkveld te inventariseren is het deskresearch geïmplementeerd met een online enquête.. Daarbij werden ook de verwachtingen over hoe FAIR in het werkveld kon helpen uitgevraagd. De enquête was gericht op personen of organisaties die betrokken zijn bij het verzamelen of analyseren van gegevens relevant voor infectieziekten met een pandemisch risico. De enquête omvatte vragen over de persoon of organisatie zelf, welke datasets er in bezit zijn of worden gebruikt, en waar deze datasets zijn opgeslagen. Daarnaast werden er vragen gesteld om te achterhalen of de gebruiker bekend was met FAIR en of de datasets (deels) FAIR zijn gemaakt. De enquête werd opgesteld in Survalyzer ([zie bijlage 1](#)) en door TNO gedeeld op LinkedIn en via netwerken van experts die actief zijn op het gebied van pandemisch onderzoek, inclusief ZonMw en Health-Holland. Een aantal van de respondenten van de enquête zijn vervolgens uitgenodigd voor een interview om extra diepgang te krijgen over hun ervaring met FAIR binnen het domein. Daarnaast is het congres 'Oversterfte in tijden van corona'⁷ bezocht om aanvullende bronnen en stakeholders te identificeren en extra informatie op te halen.

Van de geïdentificeerde databronnen en met name de portalen werd de zogenaamde staat van FAIRness, de mate waarin data voldoet aan de FAIR-data principes, geëvalueerd. Daarbij werd gekeken of en hoe de verschillende FAIR principes zijn meegenomen in het beschikbaar maken van de data.

Verder is er een overzicht gemaakt van beschikbare FAIR datamanagement tools, en of deze tools geschikt zijn voor verbeteren van FAIRness van data. Om beter inzicht te krijgen in één van deze tools, heeft het Utrechtse bedrijf The Hyve, een demo-sessie verzorgd van de tool FAIRspace. FAIRspace is een opensource research datamanagement tool dat zich richt op het opslaan en het FAIR organiseren van biomedische data. Daarnaast is een domein specifiek FAIR Implementation Profile voor pandemische paraatheid ontwikkeld met behulp van de FIP Wizard⁸. Al deze activiteiten hebben uiteindelijk geleid tot een kennissynthese waaruit verschillende adviezen voor het verbeteren van herbruikbaarheid van data voor pandemische paraatheid en infectieziekten zijn gedestilleerd. De kennissynthese wordt beschreven in het volgende hoofdstuk.

⁷ <https://www.zonmw.nl/nl/artikel/verslag-congres-oversterfte>

⁸ <https://fip-wizard.ds-wizard.org/wizard/>

3 Kennissynthese data-herbruikbaarheid voor Pandemische Paraatheid

De kennissynthese voor de herbruikbaarheid voor pandemische Paraatheid is opgebouwd uit de volgende componenten:

- Een overzicht van de stakeholders die verantwoordelijk zijn voor de databronnen of het belang dat deze partij heeft bij data beschikbaarheid in geval van een pandemie (in onderzoek, beleid en de zorg) (zie ook [bijlage 2](#));
- Een overzicht van de databronnen en de waarde die deze bronnen toevoegen aan vraagstukken over pandemische paraatheid en daarmee deel zouden moeten zijn van een FAIR data-infrastructuur voor onderzoek en beleid naar pandemische paraatheid (zie ook [bijlage 3](#));
- Een overzicht van FAIR tools en services;
- Aanbevelingen de beschikbaarheid van FAIR-data voor pandemische paraatheid verbeteren;

3.1 Stakeholder en databronnen analyse

Om te kunnen analyseren hoe pandemisch paraat Nederland is wat betreft de data beschikbaarheid zijn de volgende zes aspecten van belang:

- Wat zijn de belangen van de verschillende stakeholders bij data beschikbaarheid en herbruikbaarheid voor pandemische paraatheid?
- Welke data, relevant voor pandemische paraatheid, hebben stakeholders nodig of worden door stakeholders geproduceerd?
- Welke diensten leveren partijen voor data hergebruik?
- Hoe veel kennis en capaciteit hebben de stakeholders om data herbruikbaarheid te stimuleren?
- Ervaren de stakeholders ethische en juridische obstakels voor data hergebruik?
- Welke data zijn beschikbaar voor pandemische paraatheid en hoe herbruikbaar zijn ze?

Om deze vragen te beantwoorden is een online vragenlijst uitgezet en zijn er interviews afgenomen. Hieronder volgt een samenvatting van deze werkzaamheden en daarna wordt een overzicht gegeven van de relevante stakeholders en hun belang.

3.2 Vragenlijst en interviews: Identificeren van relevante stakeholders en bronnen

De online enquête werd gedeeld via platforms zoals de LinkedIn groepen van TNO, Health Holland en ZonMw. De enquête was oorspronkelijk beschikbaar van 9 mei 2024 t/m 20 mei 2024 maar is verlengd tot 24 juli 2024. In totaal zijn er 58 deelnemers gestart met de vragenlijst, waarvan 10 deze volledig hebben ingevuld. De response op de enquête nam aan het einde van de periode af en daarom was verlengen niet opportuun. Helaas was het niet mogelijk om te analyseren waarom de vragenlijst niet volledig werd ingevuld aangezien de data pas werden vastgelegd op het moment dat alle vragen waren beantwoord. De meerderheid van de respondenten was betrokken bij onderzoek naar pandemische paraatheid en/of infectieziekten. Van deze 10 respondenten zijn er 5 uitgenodigd voor een interview van een half uur om extra diepgang te verkrijgen over hun ervaring en kennis van FAIR binnen hun domein. Deze respondenten zijn geselecteerd op basis van ervaring met hergebruik van data en de relevantie van organisatie die ze representeren. Hieronder een samenvatting van deze gesprekken.

Interview 1: GGD Haaglanden

De geïnterviewde persoon is werkzaam op afdeling infectieziektenbestrijding als epidemioloog. Ze benadrukte de moeilijkheden bij het verkrijgen van betrouwbare wijkgerichte data, waarbij de bekendheid en toepassing van FAIR-principes binnen de GGD minimaal zijn. De geïnterviewde gaf aan dat FAIR kan helpen bij het verkrijgen van waardevolle inzichten. Zo kan er door het combineren van data worden geanalyseerd wat de effecten zijn van specifieke preventie maatregelen. Dit is nu lastig door gebrek aan detail in de databronnen. Binnen de GGD is data delen moeilijk, mede veroorzaakt door splitsing in regionale en landelijke onderdelen en een gebrek aan overzicht van beschikbare data. Het huidige informatievoorziening (IV) systeem schiet te kort op privacy niveau en maakt delen van data ingewikkeld, mede daarom wordt het uitgefaseerd. Het nieuwe IV systeem wordt op zijn vroegst in 2025 in gebruik genomen. De geïnterviewde ervaart privacywetgeving en de data beschikbaarheid als grote obstakels en heeft behoefte aan een betere coördinatie en centralisatie van data om pandemische paraatheid te verbeteren.

Interview 2: Programma Samenwerking Infectieziekten

Voor dit interview zijn twee respondenten uitgenodigd, één is werkzaam als programmadirecteur bij het programma Samenwerking Infectieziekten (een initiatief van Nederlandse Vereniging Internist Infectiologen en Nederlandse Vereniging Medische Microbiologen). De tweede respondent is parasitoloog bij het Erasmus MC en daarnaast betrokken geweest bij het opzetten van het programma Samenwerking Infectieziekten. Doel van de samenwerking is het onderzoek naar infectieziekten een boost geven, door o.a. gecombineerd gebruik te maken van patiënten-informatie en laboratoriumdata. Actuele problemen die de geïnterviewden ondervinden bij herbruikbaarheid van data, zijn regelgeving en de wetgeving die daarbij komt kijken. Een oplossingsrichting voor dit probleem is om een systeem te ontwikkelen waarbij gecombineerd, vraagafhankelijk en voor onderzoek specifieke parameters uit verschillende databronnen bij elkaar gebracht kunnen worden. Uitdaging is niet *per se* het vindbaar maken van data, maar de interoperabiliteit ervan: data is dikwijls niet gestructureerd of machine-leesbaar waardoor ze niet (her)bruikbaar zijn. Daarnaast is er behoefte aan het ontsluiten van EPD-data via regionale datahubs en het koppelen van lab-data aan bestaande systemen zoals die van het RIVM. Het ontwikkelen van gemeenschappelijke datamodellen voor infectieziekten onderzoek is cruciaal. De geïnterviewden benadrukken het belang van verdere samenwerking en

afstemming binnen het netwerk en met externe partijen, zoals ZonMw, om de voorgestelde methoden en systemen te implementeren.

Interview 3: Amsterdam UMC

De geïnterviewde van het Amsterdam UMC is senior onderzoeker ouderengeneeskunde en epidemioloog. FAIR is wel bekend in haar werkveld, maar de praktische en technische implicaties zijn vaak onduidelijk voor veel mensen. Er wordt veel data uit EPD's gebruikt, de verantwoordelijkheid voor de datavoorbereiding ligt dan bij de softwareleverancier. Deze data missen regelmatig labels of codering en moet dan door de leverancier gelabeld worden. Er wordt vanuit de leverancier wel gewerkt aan een codeboek om data beter herbruikbaar te maken. Dit is een stap richting FAIR data als dit codeboek breder wordt gebruikt. Koppelen van data uit verschillende EPD's is op dit moment complex, elk nieuw project vereist dat er opnieuw toestemming gevraagd wordt aan de data eigenaar (auteurs: beheerder). Veel verpleegtehuizen zijn 'eigenaar' (auteurs: waarschijnlijk bedoelt de geïnterviewde beheerder van de data) van data en maken gebruik van een commercieel software pakket voor databeheer. De ontwikkelaar geeft richting de onderzoekers aan dat het juridisch vaak ook niet mogelijk is om data te hergebruiken. Voor onderzoek is het noodzakelijk dat data delen (auteurs: herbruikbaar maken) makkelijker wordt.

Interview 4: Nivel

Tijdens dit interview zijn twee respondenten uitgenodigd, beide zijn werkzaam als senior onderzoeker bij Nivel. Ze zijn betrokken bij een ZonMw gesubsidieerd project 'Verbeteren van FAIRness van Nivel Zorgregistraties Eerste Lijn'. Het Nivel heeft al veel ervaring met data en met zorgregistraties maar dit is het eerste project dat zich puur richt op het verbeteren van FAIRness binnen de zorgregistraties. Nivel heeft een unieke positie, de peilstations van Nivel zijn in staat om vanuit de eerstelijnszorg de verspreiding van infectieziektes relatief snel op te pikken. Dit maakt hun werk extra relevant voor Pandemische Paraatheid. Het huidige ZonMw-project dient als een beginpunt en blauwdruk om alle data van Nivel zorgregistraties FAIR te maken (en mogelijk ook andere databronnen). Nivel streeft naar het creëren van een machine- en mens-leesbaar FAIR datapunt.

Interview 5: GGD Amsterdam

De respondent is onderzoeker/epidemioloog bij GGD Amsterdam. Omdat er geen specifieke rol als data steward bestaat bij organisatie, vervuld de respondent ook deels deze rol. Data zijn essentieel om naar pandemische paraatheid toe te werken. FAIR management is echter lastig doordat patiëntgegevens niet zomaar openbaar gemaakt mogen worden (opmerking auteurs: veel van de meta-data is niet privacy-sensitief en zou wel beschikbaar kunnen worden gemaakt). Het eerste project waarbij data werd hergebruikt, is uitgevoerd met een samenwerkingsovereenkomst en een Data Protection Impact Assessment (DPIA). Het verkrijgen van overeenstemming tussen data-leverende partijen over de specifieke invulling van de DPIA bleek moeilijk, ondanks de uniforme AVG-richtlijnen. Daarnaast ontbreekt het aan adequate documentatie en tools om data effectief te beheren en te delen (auteurs: hergebruiken). Door gebrekkige documentatie van data is het voor externen moeilijk om aan de slag gaan met de data. Op dit moment maken ze bij de GGD gebruik van een zelfgemaakt codeboek voor data labeling. Daarnaast worden timestamps wel geregistreerd, maar helaas nog niet goed geëxporteerd. Een ander obstakel dat wordt ervaren is als data gedeeld wordt met het RIVM, de GGD de nieuw ontstane dataset zelf niet meer mag inzien. De geïnterviewde benoemt nog dat een nieuw GGD GHOR project het huidige systeem moet vervangen in de komende twee jaar (auteurs: zie ook de opmerking van GGD Haaglanden over het nieuwe informatievoorziening (IV) systeem).

3.3 Relevante stakeholders voor pandemische paraatheid

Het deskresearch heeft zich gericht op publieke en private Nederlandse partijen, die actief zijn op het gebied van pandemische paraatheid, of daarvoor van belang (kunnen) zijn. Het gaat om Nederlandse partijen die actief zijn in de wetenschap, gezondheidszorg, beleid en relevante onderzoeksnetwerken die werken aan pandemische paraatheid in Nederland. Daarnaast zijn internationale initiatieven opgenomen in het overzicht voor zover deze samenwerken met de Nederlandse stakeholders en netwerken en directe relevantie hebben (zoals WHO en GloPID-R). De partijen zijn gegroepeerd in de categorieën wetenschap, gezondheidszorg, beleid en regelgeving. Sommige partijen die in meer dan een groep ondergebracht kunnen worden, zijn opgenomen in de categorie waar ze het meest aan pandemische paraatheid kunnen bijdragen (zie Tabel 3.1 en zie ook [bijlage 2](#)).

Om beleid te kunnen voeren rondom een pandemie is onderzoek met relevante data nodig. In Tabel 3.1 zijn partijen die actief in onderzoek, beleid en zorg en die van belang zijn wat betreft data voor pandemische paraatheid geïdentificeerd. Onderzoek met deze data kan inzicht geven in de progressie van een pandemie, het ziekteverloop, de behoeftes voor interventies, en de gevolgen van beleid monitoren en evalueren. Beleidspartijen hebben zelf regelmatig direct (via overheidssystemen) toegang tot data of via gerelateerde instanties (zoals het RIVM en CBS). Ze hebben echter ook belang bij aanvullende data vanuit de zorg en onderzoek om vragen te kunnen beantwoorden. Voor beleid is een relatie met beheerders van internationale databronnen in geval van een pandemie belangrijk. Deze databronnen kunnen inzicht geven in de ontwikkeling van een pandemie in landen met veel directe relaties (in- en uitreizende personen) en kan leiden tot inzicht in de effecten van beleidskeuzes van andere landen.

Voor de gezondheidszorg is toegang hebben tot data voor onderzoek van grote waarde voor het identificeren van nieuwe diagnostische methoden en behandelingen. Hiervoor is de ontsluiting van gestandaardiseerde Real World Data uit patiënten dossiers belangrijk. Real World Data (RWD) zijn gegevens die worden verzameld buiten de gecontroleerde omgeving van traditionele klinische studies, bijvoorbeeld uit elektronische patiënten dossiers. Met behulp van Real World Data kunnen pandemische signalen sneller worden opgepikt en kan op een snelle manier inzicht worden verkregen in de effectiviteit van behandelingen. Belangrijk hierbij is wel dat data vanuit verschillende patiënten dossiers onderling efficiënt kunnen worden gekoppeld. Interactie met data uit het buitenland is hierbij van groot belang vanwege het grensoverschrijdende karakter van een pandemie en omdat signalen beter worden opgepikt als de groep patiënten groter is.

Tabel 3.1: De lijst van stakeholders die belang hebben bij data beschikbaarheid voor pandemische paraatheid (zie [bijlage 2](#) voor details over belang, capaciteit, obstakels etc.). Een groot belang is aangegeven met 5 en een laag belang met 1 (kwantificatie belang)

Stakeholder	Belang bij Pandemische paraatheid	Kwantificatie Belang
Beleid		
RIVM	Inzicht in effecten van een pandemie op de publieke gezondheid & Advisering richting overheid	4
Rjksoverheid (VWS)	Beleidskeuzes in geval van en pandemie	4
Gemeenten	Beleidskeuzes en zorg ondersteuning in geval van en pandemie	4
VNG	Beleidskeuzes en zorg ondersteuning in geval van en pandemie	1
CBS	Analyses leveren over de ontwikkelingen van een infectie uitbraak op Nederlands niveau	3
WHO	Analyses leveren over de ontwikkelingen van een infectie uitbraak op wereld niveau	3
Samenwerking Infectieziekten	De werkstroom Onderzoek stimuleert onderzoek naar infectieziekten. Dat gebeurt door het verbinden van alle data-initiatieven op het gebied van infectieziekten, zodat de data-infrastructuur duurzaam geschikt is voor het eenvoudig multicenter, gecombineerd beschikbaar maken van patiëntgezondheid-data en microbiologische laboratoriumdata voor onderzoek naar infectieziekten.	3
Gezondheidszorg		
GGD/GHOR	Bron- en contact onderzoek en diagnostiek en vaccinaties coördineren	4
ICT leveranciers: FAIR services	Ondersteunend in data beschikbaarheid voor onderzoek, additionele financiële middelen nodig ingeval van pandemie	1
ICT leveranciers: EPD	Pandemische data kunnen leveren aan partners	1
ICT leveranciers: Third trusted party	Ondersteunend in data beschikbaarheid voor onderzoek, additionele financiële middelen nodig ingeval van pandemie	1
ICT leveranciers: PGO	Services voor diagnostiek en interventies rondom de pandemie	1
Zorgverzekeraars	Betaalbaar houden zorg in een pandemie	2
GGD	Inzicht in verspreiding infectie ziekten (bron- en contact onderzoek), coördinatie van diagnostiek en vaccinatie	4

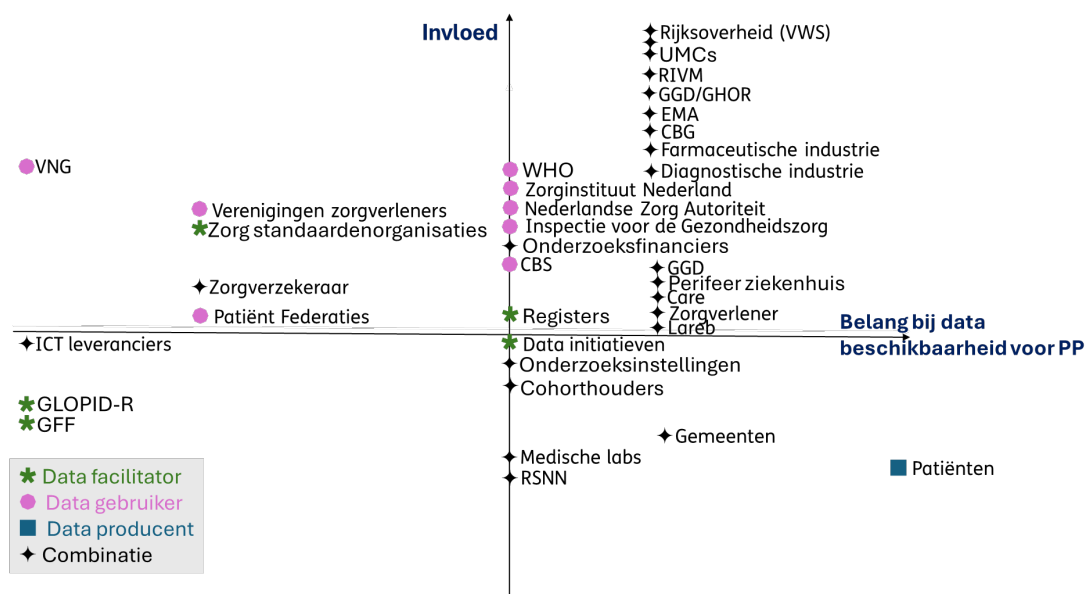
Huisarts(praktijk)	Directe verbetering zorg kunnen doorvoeren in geval van een pandemie, extra omzet door extra zorg , capaciteit op orde houden	4
Ziekenhuizen (met name perifeer)	Directe verbetering zorg kunnen doorvoeren in geval van een pandemie, extra omzet door extra zorg , capaciteit op orde houden	4
Zorgprofessional	Directe verbetering zorg kunnen doorvoeren in geval van een pandemie, extra omzet door extra zorg , capaciteit op orde houden	4
Thuiszorg	Directe verbetering zorg kunnen doorvoeren in geval van een pandemie, extra omzet door extra zorg , capaciteit op orde houden	4
Verpleeghuizen en zorginstellingen	Directe verbetering zorg kunnen doorvoeren in geval van een pandemie, extra omzet door extra zorg , capaciteit op orde houden	4
VZVZ	Nieuwe pandemie specifieke standaarden, diensten en initiatieven voor digitale zorg uitrollen	2
NICTIZ	Nieuwe pandemie specifieke standaarden ontwikkelen en implementeren voor digitale zorg uitrollen	2
Data initiatieven	Ondersteunend in data beschikbaarheid voor onderzoek, additionele financiële middelen nodig ingeval van pandemie	3
Medische labs	Nieuwe lab analyses voor een mogelijk nieuwe pathogeen & Meer lab analyses ten tijde van een pandemie	3
Verenigingen voor zorgverleners	Leden goed kunnen informeren over nieuwe diagnostiek en behandelings-opties	2
Patienten	Nieuwe diagnostiek en interventies voor de nieuwe infectie ziekte & Passende diagnostiek en behandelingsopties en de beste welzijnsuitkomst	5
Wetenschap		
Health-RI	Ondersteunend in data beschikbaarheid voor onderzoek, additionele financiële middelen nodig ingeval van pandemie	3
GFF	Ondersteunend in data beschikbaarheid voor onderzoek, additionele financiële middelen nodig ingeval van pandemie	1
GLOPID-r	Ondersteunend in data beschikbaarheid voor onderzoek, additionele financiële middelen nodig ingeval van pandemie	3
Universiteiten en andere onderzoeksinstellingen	Ondersteunend in data beschikbaarheid voor onderzoek en nieuwe onderzoeksvelden	3

Academische ziekenhuizen	Ondersteunend in data beschikbaarheid voor onderzoek en nieuwe onderzoeksvelden & Directe verbetering zorg kunnen doorvoeren in geval van een pandemie	4
Registers	Ondersteunend in data beschikbaarheid voor onderzoek, additionele financiële middelen nodig ingeval van pandemie	3
Cohort houders	Ondersteunend in data beschikbaarheid, additionele financiële middelen nodig ingeval van pandemie	3
ODISSEI	Ondersteunend in data beschikbaarheid, additionele financiële middelen nodig ingeval van pandemie	3
Beprepared	Ondersteunend in data beschikbaarheid, additionele financiële middelen nodig ingeval van pandemie	3
Onderzoeksfinciers	Versneld toegepast onderzoek naar infectieziekten	3
Netherlands Centre for One Health	Versneld toegepast onderzoek naar infectieziekten	3
ERRAZE@WUR	Versneld toegepast onderzoek naar infectieziekten	3
Pandemic and Disaster Preparedness Centre	Versneld toegepast onderzoek naar infectieziekten	3
Farmaceutische industrie	Nieuwe geneesmiddelen en vaccins kunnen verkopen	4
Diagnostische industrie (MedTech)	Nieuwe diagnostiek kunnen verkopen	4
Regelgeving		
RSNN	Ondersteunen regulatorisch onderzoek ten tijde van een pandemie, extra funding mogelijkheden	3
EMA	Snelle toegang van nieuwe geneesmiddelen en vaccins kunnen ondersteunen	4
CBG	Snelle toegang van nieuwe geneesmiddelen en vaccins kunnen ondersteunen	4
Lareb	Bijwerkingen nieuwe geneesmiddelen en vaccins snel kunnen rapporteren	4
Zorginstituut Nederland	De zorg ook ten tijde van een pandemie betaalbaar houden	3
Nederlandse Zorg Autoriteit	Versnelde analyse van de verzekerde zorg	3
Inspectie voor de gezondheidszorg	Versnelde analyse van de relatie tussen zorgvraag en aanbod	3

Overig		
Patient Federaties	Goede zorg voor hun achterban die snel kan worden aangepast op nieuwe bevindingen	2

Onderzoekers van verschillende organisaties genereren vaak zelf veel data door onderzoek naar pandemisch uitbraken naar aanleiding van een vraag van beleidsorganisaties of zorginstanties. Door het samenbrengen van deze databronnen kunnen nieuwe hypothesen worden gegenereerd die relevant zijn voor pandemische paraatheid. Het belang van deze organisaties in het beschikbaar en herbruikbaar maken van data, hangt erg af van welke type van onderzoek er wordt gedaan. Als een partij data verzamelen is het belang soms om er eerst zoveel mogelijk zelf over te publiceren en niet de prioriteit te geven aan de toegang tot de data aan anderen. Terwijl als er onderzoek wordt gedaan gericht op pandemische paraatheid op grond van data van andere partijen er een belang is in het goed vastleggen van de beschikbare data. In [tabel 3.1](#), [figuur 3.1](#) en [bijlage 2](#) wordt het belang van de stakeholder bij de data beschikbaarheid voor pandemische paraatheid beschreven.

De relatie tussen het belang van een stakeholder bij dataschikbaarheid t.o.v. de invloed is gevisualiseerd in [Figuur 3.1](#). Te zien is dat sommige partijen soms een groot belang bij het gebruik van bepaalde data hebben, maar weinig invloed op de beschikbaarheid van dergelijke data. Deze onevenwichtigheid behoeft aandacht: er moeten manieren worden gevonden waarop partijen die een groot belang hebben bij de beschikbaarheid van data voor pandemische paraatheid, maar weinig invloed op de beschikbaarheid hebben toch invloed uit kunnen oefenen. Een specifiek voorbeeld hiervan zijn de patiënten. Patiënten zouden mogelijk via de PGO meer invloed kunnen krijgen op de toegang tot data. Patiëntorganisaties zouden dit kunnen ondersteunen doormiddel van advisering.



Figuur 3.1: Stakeholder map belang bij vs. invloed op data beschikbaarheid voor pandemische paraatheid. Op de x-as het belang van de benoemde stakeholder bij data beschikbaarheid voor pandemische paraatheid en op de y-as de invloed erop. Het symbool geeft aan welke rol de specifieke stakeholder heeft (data producent, data facilitator, data gebruiker of een combinatie van rollen). De positie van de stakeholders in dit figuur is gebaseerd op Tabel 3.1.

3.4 Relevante databronnen voor pandemische paraatheid

Zoals hierboven beschreven zijn voor het onderzoek naar pandemische paraatheid niet alleen onderzoeksdata van belang, maar juist ook data vanuit beleidspartijen, zorgorganen en commerciële partijen. Om deze reden hebben we verschillende strategieën gevolgd om een overzicht te krijgen van relevante databronnen. Hierbij is ook input vanuit de eerder genoemde enquête meegenomen. Voor deze databronnen is gezocht met behulp van Pubmed (zoekterm ‘data AND COVID-19’), [FAIRsharing](#) en DuckDuckGo (zoekterm ‘data AND COVID-19’ en ‘database COVID-19’ (bij beide met/zonder taalgebied NL)). DuckDuckGo is gekozen als zoekmachine omdat deze zoekt zonder enige historische voorkennis waardoor de gevonden bronnen met elke zoekactie (onafhankelijk van de computer of degene die ingelogd is) hetzelfde vindt.

Het is van belang om onderscheid te maken in oorspronkelijke databronnen en data portals/beheerders (zoals CBS). Oorspronkelijke databronnen zijn bronnen waar de data vanuit de data verzameling direct wordt vastgelegd. Data portals zijn systemen die door organisaties zijn ontwikkeld met het doel om data en databronnen vindbaar te maken. Dergelijke systemen zijn vaak gericht op databronnen uit een bepaald domein en zijn vooral gericht op het vindbaar maken van deze databronnen voor mensen. Daarmee zijn ze dikwijls niet geschikt voor uitwisseling van data met andere partijen bovendien zijn ze meestal ook niet ingericht voor ‘machine-vindbaarheid en interpreteerbaarheid’. In [bijlage 3](#) hebben we omschreven wat voor soort databron de stakeholders beschikbaar hebben en de bronnen ingedeeld op grond van de type stakeholder (beleid, gezondheidszorg en wetenschap en regelgeving).

Databronnen uit de dagelijkse zorgpraktijk zijn een belangrijke bron voor beleidskeuzes, capaciteitsvraagstukken en nieuwe diagnostiek en behandelingen. Tijdens de COVID-19 pandemie bleken belangrijke databronnen als die van GGD, RIVM en CBS moeilijk te koppelen door technische en juridische uitdagingen. Het FAIR maken van deze bronnen zou de technische uitdagingen grotendeels oplossen. De juridische uitdagingen behoeven een organisatorische oplossing (zie aanbevelingen verder in dit document). Ook tussen zorgpartijen zijn de data regelmatig lastig herbruikbaar door gebrek aan standaardisatie. Deze uitdagingen zijn onderwerp van data infrastructuur initiatieven als zoals [Cumuluz](#), [Mprove](#) en [HIPS](#) en zullen ook aandacht krijgen binnen de Health Data Access Bodies (HDAB) de uitvoeringsorganisaties van de European Health Data Space (EHDS). Als de afstemming tussen deze infrastructuur initiatieven beter georganiseerd zouden zijn, bijvoorbeeld d.m.v. FAIR data, dan zouden we beter pandemisch paraat zijn.

Wat betreft beschikbaarheid van databronnen heeft het congres ‘[Oversterfte in tijden van corona](#)’ aanvullende inzichten gegeven. Zo bleek bijvoorbeeld dat er bijna geen bronnen voor langdurige zorg (care) beschikbaar waren waardoor de effecten van de pandemie op deze groepen niet goed onderzocht kon worden. Partijen die dergelijke data verzamelen zouden dus actief moeten worden benaderd om deze data FAIR beschikbaar te maken. Dit vergemakkelijkt pandemisch onderzoek met behulp van deze bronnen voor deze toch al kwetsbare groep mensen.

Tabel 3.2: De lijst van databronnen relevant voor pandemische paraatheid (zie [bijlage 3](#) voor details over locatie, FAIRness en kwaliteit).

Stakeholder	Databron beschrijving
Beleid	
RIVM	Ziekte data RIVM (infectieradar, roodvonk ed.), Luchtvervuiling (RIVM?), groenvoorzieningen, Vaccinatie register (RIVM/CIMS) etc.
	RIVM Covid-19 portal: verzameling van allerlei downloadbare data files relevant voor COVID-19
	RIVM-LFI: snelle opschaling en centrale aansturing van medisch-operationele* processen, zoals testen, vaccineren en bron- en contactonderzoek
	Acht studies van academische ziekenhuizen en een algemeen ziekenhuis de afweerrespons na Covid-19-vaccinaties
Rijksoverheid (VWS)	Diverse databronnen van de rijksoverheid, regionale overheid en overheidsinstanties
Gemeenten	Sterftedata aantallen, sterftedatums (leeftijd), geslacht, migratieachtergrond, huishoudgrootte, bouwjaar woning, bevolkingsdichtheid, Nabijheid zorgvoorzieningen (bv huisartsen), etc.
VNG	Alleen een data strategie voor coordinatie van gemeentelijke data
CBS	Doodsoorzakenregistratie primaire doodsoorzaak, sterftecijfers, WLZ-zorggebruik, geslacht, sociaal economische status, CBS microdata, genoomdata en fenotypische data van (resistente) gisten/schimmels afkomstig van humane, veterinaire en agri-oorsprong etc.
WHO	Internationale statistische data over gezondheid
World Mortality Dataset	Internationale statistische data over sterfte
GERDA	Regionale data rond brede gezondheid
Samenwerking Infectieziekten	Data rondom infectieziekten (initiatief van NVMM, NVII, Samenwerkende Gezondheid Fondsen (SGF) en Immunis Foundation)
Gezondheidszorg	
GGD/GHOR	Coördineren van GGD data
	Coördineren van GGD data
ICT leveranciers: EPD	Live klinische data: Data uit de ziekenhuis dossiers
ICT leveranciers: PGO	Actuele data van verschillende zorgverleners: Data uit zorg beschikbaar voor de betreffende patient in de omgeving van de specifieke PGO gedownload, gestandaardiseerd door het Medmij stelsel
Zorgverzekeraars	Bekostigingsdata zorg

GGD	Vaccinatiegraad, bron en contactonderzoek gegevens, met daarin informatie over symptomatologie, transmissieketens, clusters en persoonskenmerken, eventueel aangevuld met gerelateerde gedragsdata
Commerciële vaccinatie en test organisaties	Vaccinatie en diagnostiek data
Huisarts(praktijk)	Live klinische data: Data uit de huisarts dossiers
Ziekenhuizen (met name perifeer)	Live klinische data: DHD stroomlijnt samenwerking in de medisch-specialistische zorg op het gebied van data.
Zorgprofessional	Live klinische data: Data uit de overige zorgprofessional (ook paramedisch) dossiers
Thuiszorg	Live klinische data: Data uit de thuiszorg dossiers
Verpleeghuizen en zorginstellingen	Live klinische data: Data uit de dossiers van instellingen
Medische labs	Diagnostische data
Patienten	Verzameling van zorgdata van verschillende zorgverleners van 1 individu
European Center for Disease Control	Data rond infectie ziekten: Data van bijvoorbeeld microbiële of andere medische labs (voorbeeld labviewer)
Data initiatieven	Coördineren van klinische data: Cumuluz, Mprove, HIP, Castor FDP
Vektis	Klinische data van verschillende zorgverleners: Inzichten genereren voor goede zorg
COVIDPredict	ICU zorgdata van patienten met COVID-19
Stichting-NICE	Historische klinische data rond ICU bezetting
Wetenschap	
Health-RI	Data uit wetenschappelijk onderzoek en zorg
	Data uit wetenschappelijk COVID-19 onderzoek
	Biobank voor samples, images en data
ZonMw	COVID-19 onderzoeksprojecten
	Pandemische paraatheidonderzoeksprojecten
GLOPID-r	Internationaal overzicht van gefund onderzoek voor pandemische paraatheid
Universiteiten en andere onderzoeksinstituten	Data uit wetenschappelijk onderzoek
ELIXIR	Data in onderzoeks databases van bijvoorbeeld ELIXIR deposition databases
	ELIXIR core recourses
Onderzoeksprojecten (internationale funders)	Data uit onderzoeksprojcten van bijvoorbeeld Health Holland, Horizon Europe, NIH, Wellcome Trust etc., relevant zijn met name infectie ziekten
VODAN	Data rondom de verspreiding van COVID-19 in diverse landen

Academische ziekenhuizen	Actuele data uit Amsterdam UMC
	Actuele data uit UMCs
Registers	STIZON: Ziekteverloop en medicatie
	NIVEL: Gestructureerde huisartsen data
	NKR: De geaggregeerde gegevens van de kanker registratie: geaggregeerde data open toegankelijk. Meer gedetailleerde gegevens: gegevensaanvraag indienen.
Panel houders	Data uit panels zoals LISS panel
Cohort houders	Grote Nederlandse cohorten
ODISSEI	Gedragsdata van grote consortia
Beprepared	Gedragsdata zoals telefoonmast data voor bewegingen en supermarkt camera data voor onderzoek naar effecten van looppaden
COVID-RED	Wearable data om COVID-19 infectie te detecteren
Netherlands Centre for One Health	Data uit wetenschappelijk onderzoek
ERRAZE@WUR	Data uit wetenschappelijk onderzoek
Pandemic and Disaster Preparedness Centre	Data uit wetenschappelijk onderzoek
Farmaceutische industrie	Klinisch onderzoek data voor de ontwikkeling van nieuwe medicatie en vaccins
Diagnostische industrie (MedTech)	Klinisch onderzoek data voor de ontwikkeling van nieuwe diagnostiek
Literature and other sources	Diverse databronnen uit wetenschappelijk onderzoek
<i>Regelgeving</i>	
EMA	Database toegelaten geneesmiddelen Europa
CBG	Database toegelaten geneesmiddelen Nederland
Lareb	Database of bijwerkingen geneesmiddelen
	Database of bijwerkingen vaccins
Zorginstituut Nederland	Openbare data kwaliteitsgegevens
Nederlandse Zorg Autoriteit	Vergoedingsdata wordt bij NZA samengebracht

Uit de inventarisatie van de databronnen en van de gesprekken met de GGD kwam naar voren dat de COVID-19 crisis heeft geleid tot het inzicht dat er een omgeving nodig is waarin test- en vaccinatiedata van de lokale systemen snel ontsloten kunnen worden. Hiervoor wordt op dit moment door de GGD GHOR het IZB-systeem ontwikkeld dat gaat aansluiten bij de lokale data systemen van de verschillende regionale GGDs.

Cohorten en registers zijn andere relevante databronnen. Over het algemeen geldt voor deze partijen dat er een duidelijk doelbinding van het gebruik van de data is en dat herbruikbaarheid van data goed op orde is. In het geval van een pandemie en daarmee een ander doelbinding zijn er ethische en juridische obstakels. Zo kunnen commerciële partijen regelmatig niet gebruik maken van cohorten omdat dat binnen de AVG vaak niet wordt gezien als wetenschappelijk onderzoek. Echter toegang tot deze data kan soms de ontwikkeling van diagnostiek en/ of vaccins versnellen, wat van groot belang is in het geval van een pandemie. Voor registers geldt dat ze over het algemeen duidelijke doelbinding en grondslagen hebben voor data (her)gebruik voor hun taken. In geval van een pandemie zou voor gebruik van deze data voor andere doelen een tijdelijke aanpassing van wet- en regelgeving nodig kunnen zijn.

Een ander inzicht uit de inventarisatie van de databronnen is dat verschillende portals een verschillende aandachtsgebieden representeren. Zo richt bijvoorbeeld Pandemic Pact⁹ zich op door de WHO geprioriteerde infectie ziekten en het Health-RI COVID-19 portal¹⁰ zich op projecten binnen het specifieke ZonMw programma. Portals zijn vaak gebouwd op een eigen zelf gebouwd systeem wat het vaker lastig maakt om met een opgelegd meta-data schema te werken zoals de FAIR principes voorstellen. Een meerwaarde van het FAIR maken van deze de portals is dat het gemakkelijker wordt om selecties van de data te presenteren. Voor een dergelijk FAIR data portal zou de data maar één keer vastgelegd te hoeven te worden ten behoeve van verschillende doeleinden (bijvoorbeeld COVID-19 data of infectieziekten data).

Op de portal van Pandemic Pact (het portal van GloPID-R) is een uitgebreide zoekfunctie beschikbaar om een overzicht te krijgen van onderwerpen waar in het kader van de pandemie onderzoek naar word gedaan. Deze informatie is vastgelegd op het niveau van project beschrijving. Als de onderliggende data uit deze projecten ook FAIR zouden zijn, was er gemakkelijker vast te stellen welke bevindingen er zijn gedaan en welk vervolg onderzoek daar bij past.

De data in verschillende portals gaan dikwijls uit van anonieme of geaggregeerde data. Voor beide handelingen, het anonimiseren en het aggregeren van data, moet ook grondslag zijn volgens de AVG. Het is belangrijk dat er ten tijde van een pandemie grondslagen bestaan voor deze handelingen zodat cruciale data overzichten kunnen worden gemaakt. Behalve technische en juridische obstakels spelen er ook vaak vertrouwensproblemen die samenwerking op data standaardisatie-gebied belemmeren. Een methode die regelmatig wordt gebruikt om zonder expliciete toestemming toch data te verzamelen, is het zogenaamde 'scraping'. Scraping is een methode waarbij websites geautomatiseerd worden bezocht en bepaalde gegevens van websites en portals worden gehaald. Het GloPID-R initiatief, een project om financiering van onderzoek naar infectie ziekten te coördineren, past deze methode toe om een overzicht te hebben van onderzoeksprojecten naar COVID-19.

Health-RI heeft binnen Nederland momenteel de grootste ambitie wat betreft de ontsluiting van FAIR databronnen. Health-RI beheert op dit moment twee portals die relevant zijn voor pandemische paraatheid. Eén portal met een overzicht van alle ZonMw COVID-19-gerelateerde databronnen en één portal met 184 algemene databronnen (datum check 29 juli

⁹ <https://www.pandemicpact.org>

¹⁰ <https://www.health-ri.nl/covid-19-related-data-initiatives>

2024). Beide portals zijn nog niet volledig FAIR, maar de meta-data van de COVID-19 projecten zijn dat wel (zie [bijlage 3](#)). Zo is er bijvoorbeeld in het algemene Health-RI portal wel gebruikt gemaakt van DCAT (een standaard voor het vindbaar maken van data), maar is er een zeer beperkte beschikbaarheid van meta-data. Inmiddels zijn alle COVID-19 projecten ook opgenomen in het algemene portal maar zijn de oorspronkelijke meta data beschrijvingen nog niet volledig overgenomen. De algemene Health-RI data portal zou de rijkere meta-data representatie zoals beschikbaar in de COVID-19 portal ook moeten includeren voor alle andere projecten.

Naast het meta-data overzicht in de Health-RI portal, heeft Health-RI ook als ambitie om gezondheidsdata te ontsluiten voor innovatie, zorg en onderzoek. Daarmee zouden zij een groot deel van de bronnen, die relevant zijn voor pandemische paraatheid, FAIR beschikbaar moeten maken. Health-RI is op dit moment druk bezig om de architectuur van de infrastructuur te ontwikkelen en dan te implementeren. De timing van de beschikbaarheid hiervan is op dit moment niet duidelijk en de vraag is of alle data die relevant zijn voor pandemische paraatheid ook binnen de huidige financiering mogelijk is, en belangrijk, of deze binnen het mandaat van Health-RI ligt. Zo is tijdens een pandemie bijvoorbeeld (anonieme) data rondom verplaatsing van personen relevant in geval van een avondklok en de naleving hiervan. Tijdens de COVID-19 pandemie zijn hiervoor telefoonmast meta-data voor gebruikt. Daarnaast werden tijdens de COVID-19 pandemie looprichtingen in winkels ingesteld. Het effect hiervan zou door informatie in camera beelden kunnen worden vastgesteld. Onderzoek naar dit soort gedragsaspecten van een pandemie is onderwerp van het Beprepared consortium. Sociale data worden deels ook beschikbaar gemaakt door het ODISSEI initiatief. Hiervoor zijn ook meta-data schema's ontwikkeld (bv door GLOPID-R). Echter, deze sociale data maakt voor zover bekend niet deel van de huidige doelstellingen van Health-RI en het FAIR beschikbaar maken. Beheer over deze data wat betreft financiering en breedte van de scope, is ook nog niet goed genoeg geborgd binnen de Beprepared en ODISSEI initiatieven. Een visie op welke partijen deze rol zouden moeten hebben en waar de financiering voor beheer van deze data vandaan zou moeten komen is essentieel om pandemisch paraat te zijn.

3.5 FAIRness van de databronnen

Vindbaarheid

Om data en databronnen zowel manueel en geautomatiseerd vindbaar te maken volgens de FAIR principes moeten data op z'n minst voldoen aan de volgende criteria: (meta)data hebben een unieke identifier, data zijn beschreven met 'rijke' meta-data en data zijn geregistreerd or geïndexeerd in een doorzoekbare bron. Deze richtlijnen zouden voor volledige FAIRness punt voor punt moeten worden nageleefd. In de praktijk blijkt dit nog uitdagend. Er zijn nog nauwelijks voorbeelden van databronnen die dit volledig hebben geïmplementeerd hooguit in demo's en synthetische datasets. Daarom wordt er in dit rapport op hoofdlijnen langs deze FAIR principes gegaan om te komen tot de belangrijkste aanbevelingen om de herbruikbaarheid van data te verbeteren.

Voor het beschrijven van databronnen worden vaak meta-data standaarden gebruikt. De belangrijkste meta-data standaard die veel in de FAIR community wordt gebruikt is DCAT (Data Catalog Vocabulary). DCAT is een gestandaardiseerde vocabulaire dat ontwikkeld is door de World Wide Web Consortium (W3C). De belangrijkste kenmerken van DCAT zijn:

- Beschrijven van datasets en data services
- Helpt bij het vinden van datasets over verschillende data catalogen heen
- Maakt eenvoudige publicatie van catalogo informatie mogelijk

Behalve DCAT wordt Dublin Core Metadata Terms (DMCT) gebruikt door data aanbieders. Dit zijn meta-datastandaarden die vanuit de FAIR community worden aanbevolen. Naast DCAT en Dublin Core bestaan er nog andere vocabulaires voor specifieke data catalogen. DCAT wordt al breed gebruikt binnen de overheid. Vanuit Health-RI wordt er gewerkt aan een nieuwe versie ([DCAT-AP-3.0](#)) en één die beter aansluit bij gezondheidsdata (HealthDCAT-AP¹⁷). Vooral op de dataportals van het RIVM en de overheid wordt meta-data op juiste wijze volgens DCAT en DCMT specificaties aangeboden (zie [bijlage 3](#) kolom G). Dit is niet gedaan voor het specifieke Covid-19-portal van het RIVM (dit in tegenstelling tot het hierboven genoemde algemene dataportal van het RIVM) als gevolg hiervan is de meta-data is niet machine-leesbaar verkrijgbaar. De vindbaarheid van de datasets op de algemene Health-RI portal is ook niet afdoende vooral door het gebrek aan rijke meta-data beschrijvingen. Deze zijn nodig voor data gebruikers om vast te stellen of de data relevant zijn voor hun onderzoek. Andere geïdentificeerde bronnen zijn vooral gericht op gebruik door mensen en daarmee niet geschikt voor automatisering. Voor de betrokken organisaties kunnen de [Metadata for Machines \(M4M\)](#) workshops behulpzaam zijn. Deze hackathon-achtige workshops brengen FAIR experts en domein experts samen om te werken aan metadata schema's en de data hierop te mappen.

Toegankelijkheid

Toegankelijkheid van data en databronnen beschrijft voor mens en machine voor wie, wanneer en waarvoor data toegankelijk zijn. Binnen de analyse zijn er geen voorbeelden van databronnen waar de toegankelijkheid van machine-leesbaar gemaakt zijn. Enkele bronnen zoals ODISSEI en RIVM hebben wel een mens-leesbare beschrijving toegevoegd en dan met name aangeven onder welke licentie data toegankelijk was (in dit geval ging het om bronnen die open access zijn) ([Bijlage 3](#) kolom H). Dergelijke informatie kan wel FAIR worden gemaakt. In [bijlage 3](#) geeft kolom L aan wat de privacy status (anoniem, pseudoniem of persoonlijke data) is. Deze informatie is relevant voor de toegankelijkheid van de databron.

Als het gaat over persoonlijke gezondheidsdata in onderzoek en uit de zorg is het nog een grote uitdaging om deze toegankelijkheid überhaupt te organiseren. De analyse van bronnen laat eigenlijk voor al dit soort bronnen zien dat alleen de meta-data van deze gezondheidsdata toegankelijk is en is het niet duidelijk (mens- en machine-leesbaar) onder welke voorwaarden de data wel (her)bruikbaar is. De toekomstige EHDS vereist dat dergelijke data wel (her)bruikbaar moet worden gemaakt, FAIR data kan hier dan bij helpen.

Interoperabiliteit & Herbruikbaarheid

Het interoperabele aspect van FAIR beschrijft hoe (meta) data moet worden beschreven zodat verschillende databronnen geïntegreerd kunnen worden. Bijvoorbeeld door het gebruik van 'controlled vocabulaires', ontologieën of URI's (Uniform Resource Identifier). Controlled vocabulaires, zijn gestandaardiseerde termen die worden gebruikt om informatie van een bepaald domein te beschrijven en te organiseren. Een voorbeeld voor gezondheid gerelateerde data zijn SNOMED-CT (een standaard voor medische gegevens) en LOINC (een standaard voor medische laboratorium bepalingen). Een ontologie is een gestructureerde vocabulaire waar de termen een relatie tot elkaar hebben, hetgeen extra informatie biedt voor mogelijke interpretatie. Een URI kan deel zijn van een ontologie, maar kan ook los van elkaar bestaan. URI's zijn essentieel om data machine interpreteerbaar te maken omdat een URI er voor zorgt dat de gebruikte term maar voor één uitleg vatbaar is. Zo kan de term BMI interpreteerbaar worden gemaakt door te verwijzen naar een URI die beschrijft dat het gaat om de Body Mass Index (kg/m²) en niet om het eiwit BMI. Deze URI kan dan deel zijn van een ontologie die de verschillende lichaamseigenschappen allemaal in URIs vastlegt en ook

¹⁷ <https://healthdcat-ap.github.io>

hun relatie beschrijft. Zo zal de ontologie dan ook lichaamslengte en gewicht bevatten en expliciet maken dat deze aan de basis liggen voor de BMI.

De analyse van de bronnen en interviews laten zien dat er in weinig databronnen gebruik maken van vocabulaires, ontologieën en URI's ([Bijlage 3](#) kolom I). Dit is niet perse nodig in de brondata, maar kan door middel van een mapper ook later worden toegevoegd tussen de oorspronkelijke bron en een portal. De snelle ontwikkeling van Large Language Modellen (LLMs) kunnen hier behulpzaam zijn. Voor de zorgsector kan de EHDS hier helpen, omdat deze gaat vereisen dat data herbruikbaar wordt inclusief de beschrijving doormiddel van specifieke vocabulaires. Er zou getoetst moeten worden of er (zorg-)data is die essentieel is voor pandemische paraatheid maar niet valt onder de EHDS.

Als laatste wordt de herbruikbaarheid van data beschreven, het uiteindelijke doel van FAIR. Hiervoor moet (meta)data zodanig worden beschreven dat data beschikbaar kan worden gemaakt in verschillende omgevingen en voor verschillende doeleinden. Hiertoe moeten aspecten zoals voor welk doel de data oorspronkelijk was verzameld, context variabelen zoals lab omstandigheden, welke software was gebruikt voor het generen van data, en de versie van de data goed worden gerepresenteerd. Belangrijk voor de herbruikbaarheid is de zogenaamde provenance (herkomst en kwaliteit van data) die specifiek moet worden beschreven, wie heeft de data gegenereerd, hoe is de data verwerkt, etc. De provenance zou idealiter in een te automatiseren workflow moeten worden beschreven. Een goede beschrijving van de provenance van data wordt sinds de opkomst van Generatieve AI modellen, steeds belangrijker. Bijvoorbeeld als het gaat om de kwaliteit van data voor het trainen van LLMs. Bovendien is het regelmatig onduidelijk hoe bedrijven aan hun data komen voor de ontwikkeling van Gen-AI. FAIR en in het bijzonder het beschrijven en machine-leesbaar maken van de provenance kan deze onduidelijkheid voorkomen. Naast de behoefte aan provenance voor hergebruik (secundair gebruik) is deze er ook bij gebruik van data (primaire gebruik). De zorg van een individu kan namelijk plaats vinden bij twee of meer partijen. Als deze partijen gebruik willen maken van elkaars diagnostiek, dan moet er helder zijn dat de meting bij de ene partij misschien net anders wordt uitgevoerd en wat dit betekent voor de interpretatie van de meting. Ook hiervoor is de provenance van de data verzameling essentieel. Om deze provenance machine-leesbaar te maken kunnen LLMs ook ondersteunend zijn op dezelfde manier als voor interoperabiliteit.

De analyse van de bronnen en de interviews hebben laten zien dat er nauwelijks aandacht is voor het voldoende toevoegen van rijke meta-data ([Bijlage 3](#) kolom J). Dit is echter wel essentieel voor herbruikbaarheid en zou daarom actief door financiers en systeemverantwoordelijken moeten worden gestimuleerd en getoetst. Data kwaliteit is een ander belangrijk aspect voor herbruikbaarheid van data, maar is lastig kwantificeerbaar omdat het afhangt van het doel waarvoor de data wordt hergebruikt. [Bijlage 3](#) bevat ook een kolom herbruikbaarheid die per bron een advies geeft wat de eerste stappen kunnen zijn om verdere herbruikbaarheid te stimuleren.

3.6 Evaluatie van FAIRness van de databronnen

De databronnen zijn conceptueel beoordeeld op elk van de vier FAIR principes (zie Tabel 3.2 en [bijlage 3](#)). Deze analyse laat zien dat naleving van de FAIR principes op deze punten grote verschillen vertoont. De beschikbaarheid en toegankelijkheid zijn bij de meeste databronnen in enige mate geregeld, maar vooral de interoperabiliteits- en herbruikbaarheidsaspecten laten veel te wensen over. Zo zijn dataportals vaak door organisaties ingericht voor een specifiek doel. De databronnen zijn voor mensen prima doorzoekbaar maar over het algemeen niet machine-leesbaar en daardoor niet geschikt om het zoekproces te

automatiseren. Daarnaast zijn dataportals meestal ook niet ingericht met als doel om data te hergebruiken, maar om inzicht te geven in beschikbaarheid van data of te voldoen aan vereisten van financiers.

Als het gaat om de gebruikte standaarden voor het vinden van dataset worden er enkel de meta-data vocabulaires gebruikt, zoals DCAT en Dublin Core. Vooral op de dataportals van het RIVM en de overheid wordt meta-data op juiste wijze volgens DCAT en DC-terms specificaties aangeboden, dit zou breder kunnen worden uitgerold naar andere databronnen en portals. De vindbaarheid van de datasets zoals in de Health-RI portal is ook niet afdoende door gebrek aan voldoende meta-data. Dit heeft waarschijnlijk te maken met de doorlopende ontwikkeling van dit portal. Health-RI is voornemens om DCAT meer toe te passen. Binnen de gezondheidszorg wordt DCAT nog nauwelijks gebruikt, het implementeren van deze standaard zou voor pandemische paraatheid een goede toevoeging zijn.

Naast de goede representatie van meta-data is het van belang voor data herbruikbaarheid dat ook de onderliggende data interoperabel zijn en daarvoor gebruikt maakt van identifiers die verwijzen naar vocabulaires en ontologieën. Op het niveau van de databronnen zien we na analyse van de databronnen en portals dat er nauwelijks tot geen gestandaardiseerde modellen worden gebruikt. Dat wil zeggen dat de datavelden enkel worden beschreven met een (lokale) naam en niet wordt verwezen naar een vocabulaire of ontologie voor deze termen. Hierdoor zijn de datapunten niet semantisch interoperabel en kunnen bepaalde metingen minder makkelijk worden gevonden (vindbaarheid). De keuze van het vocabulaire kan binnen een afsprakenstelsel in een FAIR implementation Profile (FIP) worden vastgelegd. FIPs hebben nog nauwelijks bekendheid bij de onderzochte partijen, toch kan het gebruik van FIPs juist binnen een domein zorgen voor de afstemmen van de gebruikte vocabulaires. In hoofdstuk 4 komen we terug op het gebruik van FIPs doormiddel van de FIP-Wizard, een tool om een FIP te creëren.

Data toegankelijkheid wordt vaak ook geremd door data-houdende partijen vanuit de angst om controle over data te verliezen. De wetgeving (AVG) wordt dan opgevoerd als belemmering voor data beschikbaarheid. Zo wordt het (achteraf) verkrijgen van consent opgevoerd als enige grondslag voor hergebruik van data terwijl er misschien ook andere grondslagen mogelijk zijn. Voor het verkrijgen van toestemming voor hergebruik van gezondheidsdata uit een patiëntendossier zou het PGO een oplossing kunnen brengen. Sinds enkele jaren kunnen Nederlandse burgers hun gezondheidsdata (huisarts, ziekenhuis en later ook apotheek en verloskundige) in hun Persoonlijke Gezondheidsomgeving (PGO) bij elkaar brengen, dat kan worden gezien als datakluis. Data vanuit het PGO staat onder controle van het individu die consent voor hergebruik kan geven. Dit principe is onderzocht in het door Health Holland gefinancierde project C4yourself¹². Data vanuit de 3 deelnemende PGOs werden gerepresenteerd in een FAIR datapoint en konden worden gebruikt voor data visualisatie vergelijkbaar met het rijksoverheid COVID-19 dashboard. Deze methode van ontsluiting van Real World Data (klinische en zelf verzamelde data) zou in geval van een nieuwe pandemie de FAIR data beschikbaarheid en de participatie van burgers enorm kunnen verhogen. In het project bleek dat individuen met Long COVID-19 zeer gemotiveerd waren om hun data beschikbaar te stellen voor onderzoek. Deze aanvulling op de andere manieren van het organiseren van data beschikbaarheid zou in de praktijk nog verder moeten worden onderzocht op toepasbaarheid in geval van een pandemie.

De angst voor het verliezen van controle over data van data-houdende partijen zou ook kunnen worden bezworen met behulp van techniek. Een aantal initiatieven zetten daarom in op de ontwikkeling van gefedereerde data infrastructures waar data bij de bron blijven. In deze

¹² <https://www.health-holland.com/project/2021/2021/citizen-centred-and-controlled-covid-19-data-reuse>

infrastructuren worden data niet gekopieerd en gedeeld maar bezocht door een algoritme (gefedereerd leren) of wordt alleen cryptografisch versleuteld toegankelijk gemaakt. Deze oplossingen vereisen een goede machine-leesbaarheid van de verschillende databronnen, precies zoals de FAIR principes beschrijven. Voorbeelden van projecten die deze concepten hebben ontwikkeld en/of getest zijn Plugin¹³ een initiatief van de DHD, EZA en het IKNL om EPDs op een veilige manier te ontsluiten waarbij de privacy van de patiënten gewaarborgd blijft. Andere voorbeelden van de ontwikkeling van gefedereerde data infrastructuren zijn, [GERDA](#)¹⁴ een regionaal initiatief dat in de achterhoek data op een decentrale privacy vriendelijke manier data uit ziekenhuizen, ouderenzorg, huisartspraktijken, GGZ en de GGD ontsluit, en [Heracles](#)¹⁵ een project waarin een privacy-by-design data infrastructuur wordt ontwikkeld gericht op het gefedereerd ontsluiten van data uit kankerregistratie (IKNL), ziekenhuizen en (huis)artsendata (STIZON/Pharmo).

[Health-RI](#)¹⁶ [onderzoekt](#) of de gefedereerd leren methode breder ingezet kan worden, bijvoorbeeld binnen HDAB-NL. HDAB-NL is een programma van VWS dat een Health Data Access Body (HDAB) gaat ontwikkelen. De HDAB is een instantie die binnen de EHDS verantwoordelijk is voor het beheer en toegang tot gezondheidsdata en richt zich vooral op het secundair gebruik. De EHDS verplicht elk lidstaat een of meerdere HDABs op te richten. Een HDAB moet de beschikbaarheid van elektronische gezondheidsgegevens verbeteren, het proces van gegevensaanvragen en -afhandeling uniformeren en een beveiligde verwerkingsomgeving bieden. Binnen Nederland zijn VWS, CBS, ICTU, RIVM en Health-RI betrokken bij HDAB-NL.

Wat betreft de interoperabiliteit en de herbruikbaarheid van de onderzochte databronnen zijn deze minder ver ontwikkeld dan de vindbaarheid en de toegankelijkheid. Uit de evaluatie komt naar voren dat de verschillende databronnen en portals de vindbaarheid (F) en de toegankelijkheid (A) eerst hebben aangepakt. Dit is ook enigszins logisch want deze twee FAIR principes blijven op het niveau van de meta-data waar de meeste portals zich nu op richten. Het verbeteren van de interoperabiliteit (I) en herbruikbaarheid (R) vereist doorgaans het aanpassen van de databron zelf. Het *a priori* gebruik maken van gestandaardiseerde modellen of vocabulaires (woordenboeken) voor de data, in plaats van *ad hoc* de data structureren, wordt als te grote tijdsinvestering gezien en behoeft een directe incentive om wel aangepakt te worden. Dergelijke incentives zouden kunnen werken als het gemakkelijker wordt om onderzoek te publiceren of dat er wetenschappelijk waarde wordt gehecht aan het beschikbaar maken van data (via een H-index) of via druk en controle vanuit de financier. Zolang FAIRificatie weinig meer is dan gestandaardiseerde meta-data, zullen data en databronnen weliswaar beter en geautomatiseerd vindbaar en toegankelijk zijn maar worden data niet beter herbruikbaar.

In de zorg, met name in de patiënten dossiers, worden data velden wel steeds vaker maar nog te weinig beschreven, met vocabulaires zoals [SNOMED](#) en [LOINC](#). De verschillende interviews die zijn gevoerd bevestigen dit beeld. Meerdere geïnterviewden gaven aan dat de data die zij ontvangen vaak alleen met de lokale termen wordt beschreven. Dit geeft onduidelijkheid wat betreft namen van variabelen. In de praktijk betekent dit dat gedurende de samenwerking met de dataleverancier, de datagebruiker geregeld over-en-weer mailt met vragen over de betekenis van de data-elementen. Gedurende de samenwerking bouwt de datagebruiker zo een vocabulaire op. Voor goede interoperabiliteit levert de leverancier een vocabulaire aan in samenhang met de data. Dit gebruik van de juiste vocabulaires zou door zorg- en wetenschappelijke financiers (in

¹³ <https://plugin.healthcare/>

¹⁴ [GERDA is live! - populationhealthdata.nl](#)

¹⁵ [HERACLES legt basis voor veilige datadeling in zorg](#)

¹⁶ [Gefedereerde analyse | Health-RI](#)

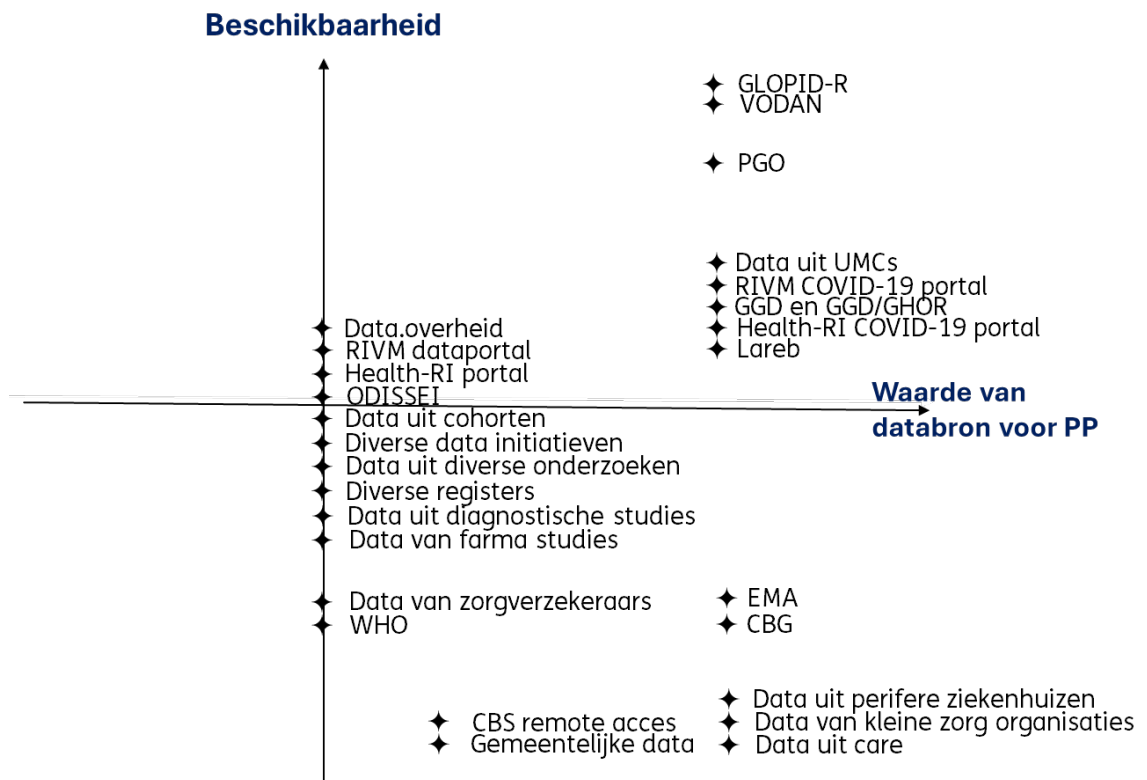
Nederland en Europa via de EHDS) kunnen worden gestimuleerd als eerste stap van interoperabiliteit. Deze vocabulaires zouden ook kunnen worden vastgelegd in een FIP voor het specifieke domein. Onlangs is er een kamerbrief¹⁷ gestuurd naar aanleiding van een rapport van Nictiz waarin wordt beschreven dat gebruik van de SNOMED-CT ontologie in de zorg op den duur verplicht gesteld zou moeten worden.

Optimale herbruikbaarheid van data zou additioneel vereisen dat de herkomst en manier van meten van de data op een gestandaardiseerde manier wordt weergegeven. Hiervoor zou [PROV-O](#) ontologie gebruikt kunnen worden. Op dit moment wordt deze ontologie nog niet gebruikt in databronnen voor pandemische paraatheid. Daarnaast vereist herbruikbaarheid genoeg meta-data om vast te leggen hoe en waarvoor de data zijn verzameld zodat ook duidelijk is of de data relevant is voor nieuw onderzoek. Deze verdieping van de meta-data is nog onderbelicht. Hierdoor kan onduidelijk ontstaan over welk meetplatform is gebruikt voor een dataverzameling of welke co-varianten een rol spelen in de waarden van gemeten parameters.

Wat betreft de herbruikbaarheid van data en vooral de provenance, is de grote datahonger van de bedrijven die (Gen)AI ontwikkelen, interessant. FAIR kan helpen om data beter te beschermen tegen het zonder toestemming gebruiken van de data maar biedt ook de mogelijkheid om data van hoge kwaliteit aan te bieden. Dit genereert een marktkans voor datahoudende partijen omdat zij zich mogelijk kunnen laten betalen door (Gen)AI bedrijven voor het gebruik van de data (indien dat ethisch en juridisch mogelijk is). Geautomatiseerde toegang en rijke provenance van de data is hiervoor cruciaal.

Figuur 3.2 laat zien dat er behoorlijk wat bronnen zijn die belangrijk zijn voor pandemische paraatheid maar die niet of nauwelijks beschikbaar zijn. Dit gaat om bronnen uit de gezondheidszorg en gedragsdata zoals van telefoonmasten. Het beschikbaar maken van deze data, het FAIR maken, behoeft aandacht en financiering om voorbereid te zijn op de volgende pandemie.

¹⁷ <https://open.overheid.nl/documenten/786237cd-bd33-459a-b500-315d7df4839b/file>



Figuur 3.2: Relevantie van databronnen. Op de x-as staat de waarde van de databron voor onderzoek, beleid en innovaties voor pandemische paraatheid. Op de y-as staat hoe beschikbaar de data (FAIR) op dit moment is. De positie in de map relateert aan de kolommen (F-I en M) in [bijlage 3](#). Omdat de bronnen zijn geselecteerd op een waarde die ze hebben voor Pandemische paraatheid is als minimale positie voor belang de nullijn gekozen. Als bronnen dus rechts van de y-as staat zijn ze meer van belang dan gemiddeld.

FAIRificatie van datasets vindt vaak plaats binnen organisaties en worden zichtbaar gemaakt in een specifiek (eigen) portal, dus bijvoorbeeld de RIVM-data beschikbaar in het RIVM-portal. Wanneer standaarden worden ontwikkeld of worden uitgebreid bij de ene partij tijdens FAIRificatie van de datasets en een ander heeft deels overlappende datasets beschikbaar, dan is het nuttig om deze standaarden af te stemmen. Een gezamenlijke FIP leent zich bij uitstek voor een dergelijke afstemming.

In een ideale situatie wordt een dataset met meta-data beschreven volgens afspraken uit de FIP, met unieke persistente (geen dode links) identifiers en correcte schema's door de partijen die data verzamelen (dus aan de bron). Hierna kunnen deze beschikbaar gemaakt kunnen worden in één of meerdere portal(s). Iedere portal of catalogus kan deze data FAIR aanbieden door middel van deze identifiers en meta-data en is het voor mens en machine meteen helder dat dit om dezelfde dataset gaat. Financiers kunnen in het data management plan toetsen of het project aan deze logica voldoen en zullen aan het einde van het project moeten toetsen of het data management plan (volledig) is uitgevoerd.

3.7 Data duplicatie

Vindbaarheid van data betekent dat het gemakkelijk wordt om relevante data in verschillende portals te vinden zonder dat er kopieën hoeven te worden gemaakt van deze (meta)-data. De analyse van databronnen heeft inzichtelijk gemaakt dat de data van de verschillende type stakeholders (beleid, onderzoek, zorg) binnen de groep regelmatig overlappen. Zo is er een overlap van data tussen RIVM en rijksoverheid data (beleid) en is er overlap is tussen de GLOPID-R, ZonMw en Healh-RI datasets (onderzoek). Aangezien Health-RI de ambitie heeft om data FAIR toegankelijk te maken voor beleid, innovaties en onderzoek zal deze overlap toenemen. Zo is Health-RI al in gesprek met RIVM om hun data ook vindbaar te maken in de Health-RI portal. Dit soort overlappen behoeven wel aandacht, aangezien deze kan leiden tot data duplicatie. Data duplicatie kost extra resources en heeft bepaalde risico's. Een serieus risico is dat data mogelijk onbewust twee keer kan worden meegenomen voor analyse als de duplicatie niet wordt herkend door mens en/of machine. Daarnaast is er het risico dat verschillende databronnen een andere versie van de data beheren. Het is dan van belang dat de versie voor de gebruiker (mens of machine) vast te stellen is. Door volledig FAIR te werken kan data op verschillende plekken vindbaar zijn zonder dat duplicatie nodig is en dat blijft inzichtelijk doordat FAIR ook voor de data unieke identifiers vereist. Dit voorkomt meer opslag dan nodig is en vermindert het risico op dubbeling van data in de analyse. Hiervoor is het van belang dat data bij de bron FAIR wordt gemaakt en beschikbaar komt in een FAIR data point. De ontwikkeling van een Nederlandse Health Data Access Body (HDAB-NL¹⁸) door Health-RI, RIVM, CBS en ICTU zal de beschikbaarheid en herbruikbaarheid van data verder moeten stimuleren.

¹⁸ <https://www.health-ri.nl/en/program-hdab-nl>

4 FAIR Research tools & services

Onderdeel van de kennissynthese is een overzicht en analyse van een aantal beschikbare FAIR datamanagement tools. Er is gekeken naar bestaande FAIR assessment tools die de FAIRness van databronnen helpen verbeteren en naar een methode om een FAIR Implementation Profile te genereren. Dit hoofdstuk start met meer achtergrond over hoe FAIR onderdeel uit kan maken van data management. Dit geeft de wat minder ingevoerde lezer de achtergrond informatie over hoe het FAIR maken van databronnen kan worden verbeterd.

4.1 FAIR maken van data

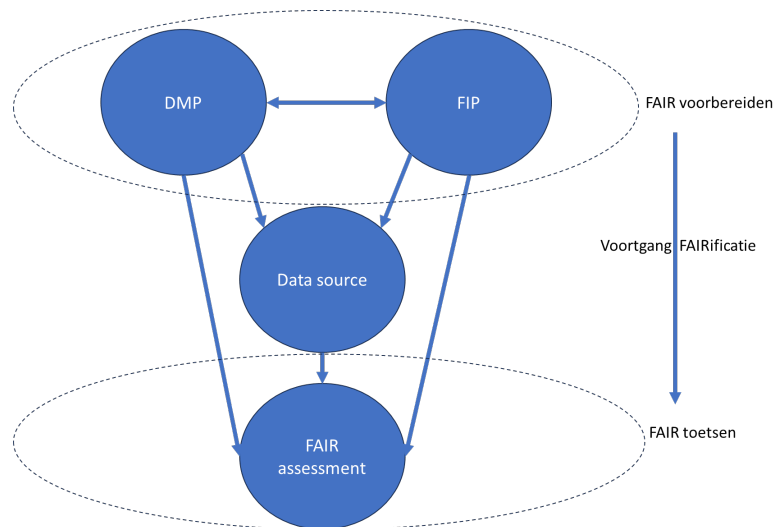
Om data binnen en na projecten goed te beheren wordt er steeds vaker een data management plan (DMP) opgesteld. Het DMP beschrijft hoe er tijdens het verloop en na afloop van een project omgegaan wordt met data en wat er voor het beheer van de data nodig is (aan tijd, hardware en software, etc.). Dit plan beschrijft bijvoorbeeld de bewaartermijn van data die soms door privacy wetgeving worden gelimiteerd. Soms spelen de hoge kosten van opslag van data ook een rol bij het verwijderen van data (bijvoorbeeld Peta-bytes aan ruwe data). Voor het maken van een DMP zijn verschillende generieke tools beschikbaar zoals de Data Steward Wizard¹⁹ een gemeenschappelijke inspanning van ELIXIR en GO FAIR) of DMP online²⁰ van het Digital Curation Centre en de University of California Curation centre).

De meeste projectfinanciers (bijvoorbeeld ZonMw) vereisen een DMP voor de start van een project en vragen het plan regelmatig te actualiseren. Als onderdeel van het DMP wordt data herbruikbaarheid volgens de FAIR principes steeds vaker geëist door project financiers. Daarmee is het DMP uitermate geschikt om herbruikbaarheid van data via de FAIR principes af te dwingen.

De praktijk van project financiers om data herbruikbaar te maken via de FAIR principes blijkt echter weerbarstiger dan gedacht. Er zijn verschillende uitdagingen waardoor nog niet alle databronnen volledig FAIR zijn. Voorbeelden van uitdagingen zijn: niet alle onderzoekers zijn bereid hun data voor hergebruik geschikt te maken (in de hoop zelf er meer over te kunnen publiceren); veel onderzoekers weten niet precies wat FAIR praktisch inhoudt en er zijn onvoldoende gebruikersvriendelijke methodes en technieken om te ondersteunen in het FAIR maken van data. Binnen een organisatie kan een data steward worden aangesteld als manager van de data. Een data steward beheert de kwaliteit en communiceert over data. Niet iedere organisatie heeft deze expertise en er zijn op dit moment ook niet genoeg data stewards. Daarnaast is het voor financiers de vraag wanneer een project de FAIR principes heeft nageleefd: wanneer is de data FAIR genoeg?

¹⁹ <https://ds-wizard.org>

²⁰ <https://dmponline.dcc.ac.uk/>



Figuur 4.1: FAIRificatie van data. Een DMP (data management plan), FAIR Implementation Profile (FIP), een data source (databron) en FAIR assessment tools zijn allen componenten die belangrijk zijn voor het FAIR maken van data en staan in relatie tot elkaar.

Waar DMPs op dataproject niveau bijdragen aan het herbruikbaar maken van data d.m.v. de FAIR principes, doen FAIR Implementatie Profielen (FIPs) dit op het niveau van organisaties of voor specifieke domeinen. FIPs zijn in het leven geroepen om het probleem van inconsistentie en gebrek aan transparantie van organisaties voor de implementatie van FAIR principes aan te pakken. FIPs lenen zich daarom bij uitstek voor data standaardisatie en harmonisatie van projecten rondom pandemische paraatheid. Voor een FIP wordt door een representatieve groep mensen uit een bepaald domein per FAIR principe, consensus gezocht voor de invulling. Zo beschrijft een FIP onder andere welke technische en semantische datastandaarden worden gebruikt. Een FIP voor het pandemische paraatheid domein zou beschrijven welke keuzes zijn gemaakt voor bepaalde standaarden. In geval van een vogelgriep pandemie kan dan worden teruggevallen op een specifieke FIP waarin termen zijn vastgelegd voor data die nodig zijn om de verspreiding van de pandemie te volgen. Daarnaast wordt in een FIP vastgelegd hoe de verschillende databronnen kunnen worden gevonden. In verschillende landen worden er nu databronnen ontwikkeld om voorbereid te zijn voor een volgende pandemie. Door het uitwerken van een FIP kunnen deze systemen klaar staan voor wereldwijde samenwerkingen. Het maken van een FIP wordt ondersteund door een tool zoals de FIP [wizard](#). Een FIP helpt fragmentatie van initiatieven te voorkomen en bevordert harmonisatie en convergentie voor de FAIR data-infrastructuur, in dit geval voor pandemische paraatheid.

Zoals aangegeven kunnen DMPs de herbruikbaarheid van data volgens de FAIR principes bevorderen. Een FIP kan het maken van een DMP vergemakkelijken doordat de FIP de vereiste domein keuzes voor de FAIR principes al heeft gemaakt. Deze kunnen dan direct worden opgenomen in een DMP. Op haar beurt kunnen eerdere beschikbare DMP's van vergelijkbare projecten gezamenlijk input zijn om een nieuwe FIP te creëren. Het is wenselijk dat zowel een FIP als een DMP worden opgesteld in de voorbereiding van het (FAIR) maken van een dataset. Dit zou nog verder gestimuleerd kunnen worden door financiers van onderzoek door het gebruik van relevante FIPs als eis te stellen als onderdeel van een DMP.

In **Figuur 4.2:** Data integratie schema voor organisaties (uit het Playbook for Health Data Sharing Ecosystems) wordt de opname van data in een bestaande Fair Data Point portal beschreven. Dit proces start altijd met het authentifieren en autoriseren van een partij die

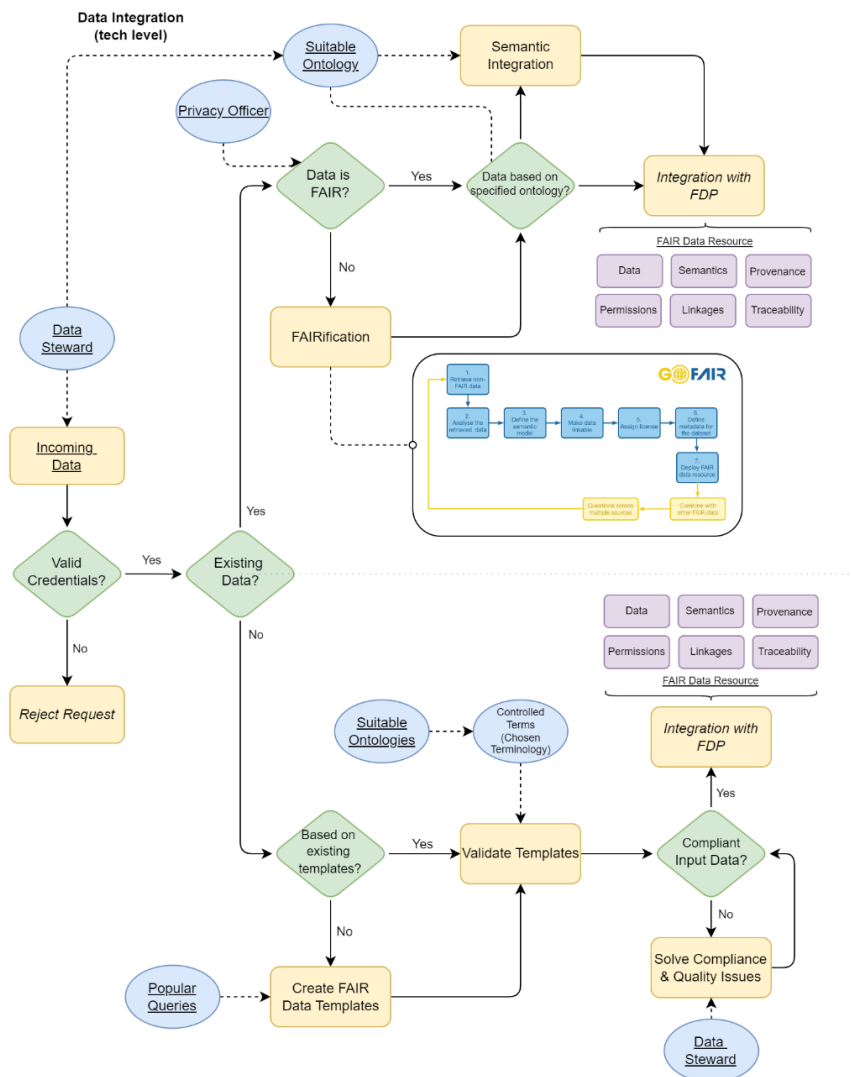
data inbrengt. Hierna wordt bepaald of het gaat om een al in de portal voorkomende databron of een nieuw databron. Daarna wordt de workflow van het proces gesplitst in twee takken:

Bestaande databronnen:

- Er wordt bepaald of de data al conform zijn volgens de FAIR principes
- Als dit niet het geval is dan gaan de data door het FAIRificatie proces.
- Check of de data gebruik maakt van een bekende ontologie
- Indien niet, het mappen van de termen in de data naar een bestaande ontologie
- Opname van de data in de FDP als FAIR Data Resource (FDR)

Nieuwe databronnen:

- Check of het nieuwe databron aansluit bij een bestaande FAIR meta-data template of data er een template is mee geleverd.
- Indien niet, dan moet er een nieuw template worden gespecificeerd anders is het niet mogelijk over te gaan tot data integratie.
- Check of de databron conform het geldende interne en externe beleid is.
- Indien niet, moeten deze beleidsissues worden opgelost voor opname in de FDP
- Opname van de data in de FDP als FAIR Data Resource (FDR)



Figuur 4.2: Data integratie schema voor organisaties (uit het [Playbook for Health Data Sharing Ecosystems](#))

Na deze stap kan een databron worden getoetst op de mate van FAIRness. Dit kan met behulp van FAIR assessment tools. Verder in dit document komen we uitgebreid op deze tools terug. Uit een assessment kan blijken dat verschillende databronnen voor verschillende type data een andere mate van FAIRness hebben. Dit kan komen door bijvoorbeeld een gebrek aan goede standaarden en dit inzicht kan daarmee ook helpen in het steeds beter toegankelijk maken van de data.

4.2 FAIR implementatie bronnen

Om een databron geschikt te maken voor hergebruik volgens de FAIR principes is specifieke kennis nodig. Deze informatie is beschikbaar d.m.v. verschillende online bronnen of tools. De belangrijkste hiervan zijn de [RDMkit](#), [FAIR cookbook](#), [FAIR toolkit](#) en het [FAIRplaybook voor Health Data Ecosystems](#). Voor het FAIR maken van databronnen zelf, ook wel FAIRification genoemd, zijn verschillende tools en services ontwikkeld. Om vast te stellen of FAIR

implementatie succesvol is uitgevoerd kunnen assessment tools worden gebruikt, deze worden hieronder beschreven en beoordeeld.

4.3 FAIR assessment tools

Binnen de opdracht zijn verschillende FAIR assessment tools geïdentificeerd en getest. Deze tools kunnen datagebruikers helpen databronnen op mate van compliance aan de FAIR principes te toetsen. Een belangrijk inzicht van de analyse van de tools is dat ze de FAIR principes op een andere manier valideren. Dit komt met name door de FAIR principes zelf, die globaal zijn gedefinieerd om breed toepasbaar te zijn. Dit maakt het lastig om een goede assessment tool te maken, want deze moet breed inzetbaar zijn en zelf geen ontwerpkeuzes afdwingen. Neem bijvoorbeeld principe R1 (volgens GO FAIR): “(Meta)data are richly described with a plurality of accurate and relevant attributes.” Wanneer is rijk, rijk genoeg en welke attributen zijn nodig voor herbruikbaarheid? De invulling zal dus verschillen per specifieke context. Ondanks deze limitatie kan een algemene FAIR assessment tool nog steeds van nut zijn om als gebruiker snel inzicht te krijgen in de FAIRness van een databron. Hieronder wordt een korte toelichting per tool gegeven zonder in te gaan op gedetailleerde verschillen tussen de tools:

- **FAIR Ontology Pitfall Scanner (FOOPS): FOOPS! (linkeddata.es) versie July 2021**

Deze tool is ontwikkeld door de ontologie-onderzoeksgroep aan de universiteit van Madrid (UPM). Het werk is nog in een bèta-staat en werkt voornamelijk met ontologieën in plaats van met datasets. Daarom is deze tool voor het Pandemisch Paraatheid werk van ZonMw niet direct interessant.

- **FAIR checker:** (French Institute for Bioinformatics: <https://fair-checker.france-bioinformatique.fr/>) **versie 1.2.7**

Deze tool is heel interessant voor ZonMw omdat gemakkelijk kan worden geëvalueerd in hoeverre de meta-data van een dataset FAIR toegankelijk is. Via enkele diagrammen wordt de mate van FAIRness gevisualiseerd.

- **F-UJI tool: versie 3.2**

Deze tool is ontwikkeld door het Europese Horizon2020 FAIRsFAIR project. Het is hier onduidelijk op welke manier de score voor een databron voor een bepaald FAIR principe kan worden verbeterd. Als dit zou worden toegevoegd kan de tool zeer bruikbaar zijn voor Pandemische paraatheid.

- **FAIR Evaluation Services: versie 2018**

Deze FAIR-tool is ontwikkeld door UPM, Universiteit Maastricht, Universiteit Oxford en de GO FAIR Foundation (GFF). Het is mogelijk een dataset te valideren aan de hand van verschillende criteria die samengevoegd kunnen worden in enkele tests. Het voordeel is dat de implementatie van de criteria gemakkelijk toegevoegd en uitgekozen worden door gebruikers, waar deze bij de meeste andere assessment tools vaststaat.

- **FAIR Data Maturity model** (RDA FAIR Data Maturity Model Working Group, https://www.rd-alliance.org/group_output/fair-data-maturity-model-specification-and-guidelines/): versie 1.0

Deze tool bestaat enkel uit een handmatige vragenlijst die gebruikers worden verwacht in te vullen. Het bevat geen (toekomstige) mogelijkheid om automatisch te worden toegepast op een dataset. Daarom zien we dit als een assessment service die uiteindelijk uitgefaseerd gaat worden.

Voor meer informatie geeft de blogpost [The Road to FAIRness: An Evaluation of FAIR Data Assessment Tools | The Hyve en het rapport van de European Open Science Cloud \(Microsoft Word - Report on the FAIR Evaluation events_final_sub.docx \(eosc.eu\)\)](#), een gedetailleerde

uiteenzetting van een aantal FAIR assessment tools die hierboven genoemd zijn. Overigens ontwikkelen deze tools zich snel en verwachten we dat de bovenstaande tools steeds gebruikersvriendelijker zullen worden.

4.4 FAIR Implementatie Profiel

Voor organisaties of onderzoeksgroepen is het vaak lastig om volgens de FAIR principes te werken vooral omdat FAIR geen standaardisatie- en implementatie keuzes afdwingt. Hiervoor is het concept van FAIR implementation Profiles (FIP) bedacht. Een FIP helpt een praktijkgemeenschap (community) op basis van gezamenlijk afspraken implementatiekeuzes te maken voor elk van de FAIR principes. Deze keuzes zijn te vergelijken met het maken van een afsprakenstelsel. Een FIP maakt het FAIRificatie-proces van een databron voor gemeenschappen eenduidig en verbetert de herbruikbaarheid van data binnen een gemeenschap of voor een bepaald onderzoeksdomein. Hoewel het realiseren van een FIP tijdrovend en complex werk is geeft het wel het meest diepgaande resultaat dat is toegespitst op de context van de databron. Specifieke FIPs voor pandemische paraatheid hebben daarmee zeker een toegevoegde waarde.

Als oefening is er in het kader van dit project een eerste versie voor een FIP voor data voor pandemische paraatheid opgesteld (zie [bijlage 4](#)). Aangezien er door ZonMw geprogrammeerd wordt op pandemische paraatheid is er gepoogd een generieke FIP voor alle type databronnen te maken. Deze FIP beoogt samenhang tussen de keuzes voor verschillende datatypes. Daarbij is gebruik gemaakt van de FIP-wizard waarin elk van de FAIR principes wordt behandeld in één of meerdere richtlijnen. Deze zijn soms toegespitst op een specifieke dataset of meta-data, omdat het niet altijd mogelijk is voor alle datasets eenduidig de bepalingen te beantwoorden. Dit leidt dus tot een soort netwerk van FIPs waarbij sommigen meer specifiek zijn en andere meer generiek. Deze gelden dan als voorbeeld om een specifieke keuze te maken voor andere data types. Deze Pandemische paraatheid-FIP (PP-FIP) is zelf ook beschreven als FAIR dataset en machine-leesbaar, geschikt voor verdere aanpassing en hergebruik. De PP-FIP is besproken met domeindeskundigen van ZonMw en FAIR-deskundigen van de GFF. Het document is te vinden via de DS Wizard (zie [bijlage 4](#)). Waar relevant zijn in de FIP *implementation considerations* toegevoegd. Dit zijn adviezen en opmerkingen over hoe elementen uit de FIP gebruikt kunnen worden in de praktijk en bij het gebruik van de FIP in de FAIRificatie van een dataset. Deze adviezen of opmerkingen kunnen inzicht geven bij het verdere beheer en werk aan een FIP voor pandemische paraatheid door financiers en systeemverantwoordelijken. Daarnaast zijn er bugs en ideeën over mogelijke verbeteringen van de FIP wizard gedeeld met GFF.

4.5 De FIP en FAIR assessment

Aangezien de FIP vastlegt welke implementatie keuzes zijn gemaakt voor een bepaalde gemeenschap is de FIP ook relevant voor de toetsing van FAIRness. Op dit moment is er geen assessment tool die de FIP automatisch mee kan nemen in het toetsingsproces. Hierdoor vereist het daarom behoorlijke kennis van de FAIR principes bij de uitvoerder en ook een zekere discipline om de assessment nauwkeurig uit te voeren. Dit zou kunnen worden verbeterd en vergemakkelijkt met een automatische assessment tool die een FIP als uitgangspunt kan nemen en kan inlezen. Vervolgens kan het assessment worden uitgevoerd door te controleren in hoeverre de keuzes per bepaling zijn nageleefd voor de databron. Een geautomatiseerde methode voor FAIR assessment die een dataset onderzoekt aan de hand van de bepalingen uit de FIP voor pandemische paraatheid, is op dit moment nog niet beschikbaar.

In de toekomst zou dit wel de meest uitvoerige methode van FAIR assessment kunnen opleveren. De GO FAIR Foundation(GFF) werkt aan een assessment tool voor automatische validatie volgens een FIP.

We noemen hier nog enkele andere overwegingen rond de FIP en FAIR assessment. Deze kunnen onder andere relevant zijn in de keuzes voor de ene of de andere assessment tool. Ten eerste is het belangrijk om het moment in het proces van FAIRificatie in overweging te nemen. Aan het begin van een project om een dataset te ontwikkelen is er nog geen noodzaak voor FAIR assessment. Kennis en bewustzijn van FAIR zijn wel belangrijk zodat de juiste keuzes voor standaarden en ontologieën kunnen worden gemaakt. De inhoud van de relevante FIP kennen en een data management plan hebben, zal in deze fase daarbij een goede basis vormen. Ook helpt het bij een goede start om (de opbouw van) andere (gevalideerde) FAIR data points (FDPs) te kennen. Dit zijn plaatsen of services waar data en/of meta-data volgens de FAIR principes beschikbaar zijn gemaakt.

Tijdens een project is een assessment van de FAIRness van de dataset nuttig om daarna zo nodig te kunnen bijsturen en verbeteringen aan te brengen. Een assessment aan het einde van een project kan gebruikt worden als bekrachtiging om te kunnen verklaren dat de geproduceerde data voldoet aan de FAIR principes. Een dergelijk assessment kan daarmee nut hebben voor organisaties als ZonMw die FAIR data binnen gesubsidieerde projecten eisen. Een assessment is belangrijk wanneer hergebruik van de dataset wordt overwogen of om te onderzoeken of de data geschikt is voor hergebruik en of dat er eventuele verbeteringen nodig zijn. Organisaties die databronnen beschikbaar maken en/of beheren zouden zelf deze FAIR assessment moeten uitvoeren en als meta-data toevoegen aan de data. Dit is typisch werk dat de data steward van een organisatie zou moeten uitvoeren.

Ook de rol van de gebruiker van een FIP of FAIR assessment tool is van belang. Mensen in verschillende rollen en op verschillende momenten in het FAIRificatie proces zullen met de data in aanraking komen. Afhankelijk van de rol van gebruiker en de kennis over FAIR zal er meer diepgang worden vereist van een assessment. Een data-steward die een bron beheert met veel kennis van FAIR en de FIP uit het domein, zal meer baat hebben bij een assessment tool die de FIP inleest en gebruikt, terwijl de eigenaar van data een algemener overzicht van de FAIRness kan vragen om te bepalen of er moet worden ingezet op verbetering. Een onderzoeker die wil bepalen of een gevonden dataset herbruikbaar is, kan een oplossing tussen deze twee uitersten in zoeken.

5 Conclusies

Herbruikbaarheid van de juiste data is van belang om goed voorbereid te zijn in de pandemische en de koude fase van een pandemie. Op basis van de bevindingen in dit rapport wordt de conclusie getrokken dat data beschikbaarheid en vooral de herbruikbaarheid van data bij de onderzochte databronnen niet op orde is en daarmee is Nederland qua data organisatie niet pandemisch paraat. De toepassing van FAIR zou een belangrijke en efficiënte manier zijn om relevante data pandemische paraatheid te maken. Daarbij zou het FAIR vindbaar maken van voor pandemische paraatheid relevante data al een belangrijke eerste stap zijn.

5.1 Data beschikbaarheid

Uit dit onderzoek komt een beeld naar voren dat er voor pandemische paraatheid veel relevante databronnen zijn bij zeer uiteenlopende partijen: in de wetenschap, zorg en vanuit beleid. Deze data zijn essentieel om inzicht te krijgen in de sociaal-maatschappelijke en financiële effecten van de pandemie en om de gezondheidszorg optimaal te laten functioneren. De meeste van deze bronnen zijn op zichzelf bruikbaar maar zijn niet geschikt voor een geautomatiseerd proces van hergebruik van data. Voor machine-leesbaarheid, een van de centrale pijlers van FAIR om hergebruik van data te automatiseren, zijn nog aanvullende acties nodig bij alle drie de type stakeholders (onderzoek, beleid, en gezondheidszorg). Juist deze machine-leesbaarheid maakt dat databronnen snel en efficiënt kunnen worden gevonden en dat het nut voor hergebruik kan worden bepaald.

De databronnen die zijn gevonden in dit onderzoek voor pandemische paraatheid geven niet het volledige overzicht er missen ongetwijfeld nog relevante databronnen. Dat heeft enerzijds te maken met de reikwijdte van de opdracht en de moeite die het handmatig zoeken naar databronnen kost. Maar ook met de ‘known unknowns’: voor een nieuwe pandemie kunnen nog niet geïdentificeerde databronnen van belang zijn.

Wat betreft de aard van de data; er zijn vooral databronnen gevonden en onderzocht die in retrospectief van belang zijn voor pandemische paraatheid. Er zijn maar weinig databronnen gevonden met geschikte actuele data, laat staan dat deze databronnen op een FAIR manier beschikbaar zijn. Eén van de redenen dat juist actuele data moeilijk beschikbaar zijn is angst voor privacy wetgeving. Wat betreft de beschikbaarheid van actuele data zal de EHDS in eerste instantie niet veel bijdragen omdat daar nu wordt ingezet op een jaarlijkse update van de meta-data in de HDAB catalogus. Voor een pandemie waar beleidskeuzes dagelijks moeten worden genomen en snelle ontwikkeling van zorginnovaties levens kunnen redden is deze tijdslijn voor pandemische paraatheid dus een groot risico. Het is daarmee van groot belang dat er ook actuele ontsluiting van data mogelijk wordt.

Vanuit de interviews komt naar voren dat de manier van het stimuleren van data beschikbaarheid afhankelijk is van de data leverende partijen. Het is belangrijk dat er centrale regie wordt genomen om data herbruikbaarheid te vergroten (zie ook de aanbevelingen). De invoering van de EHDS geeft hiervoor handvaten. Voor pandemische paraatheid zullen echter ook andere databronnen herbruikbaar moeten worden gemaakt die niet onder de EHDS vallen, zoals de eerder genoemde ontsluiting van actuele en sociale data.

5.2 Organisatie van FAIR data voor Pandemische Paraatheid

Het proces van het FAIR maken van data maakt deze in principe beter beschikbaar en herbruikbaar. Uit de vragenlijst, de interviews en het deskresearch blijkt dat de FAIR principes nu bij veel partijen die een rol hebben in pandemisch paraatheid zijn doorgedrongen. Uit het vervolg onderzoek komt naar voren dat meer verantwoordelijken voor databronnen meer aandacht besteden aan de FAIR principes. Toch is er nog veel werk te verzetten, vooral in de zorg en bij organisaties verantwoordelijk voor publieke gezondheid blijft het begrip van FAIR data nog achter.

Een ander beeld is dat behalve bekendheid met FAIR er ook veel onbegrip is over wat FAIR in de praktijk nu echt betekent en wat er concreet voor nodig is om data FAIR te maken.

De huidige tools om FAIR data te creëren, vast te leggen of de bronnen te toetsen op FAIRness zijn nog van een te laag technologie readiness level (TRL). Ook de gebruikersvriendelijkheid van deze tools is nog onvoldoende. De FAIR assessment tools die zijn bekeken, zijn op dit moment nog te beperkt om breed gebruikt te kunnen worden. Verdere ontwikkeling moet leiden tot een tool die makkelijk in gebruik is, duidelijke scores geeft voor de realisatie per FAIR principe, en aanwijzingen geven voor verbeteringen. Bovendien moet het eenvoudiger worden om datgene wat in de ene tool is vastgelegd, bijvoorbeeld de keuzes die gemaakt zijn door een gemeenschap en vastgelegd in de FIP, ook in een andere tool is te gebruiken. Verder ontbreekt het bij partijen aan middelen om bijvoorbeeld een data steward aan te stellen om data FAIR te maken. Veel partijen werken op projectbasis, terwijl data management niet stopt na de uitvoering van een project. Hier zou budget voor moeten worden gealloceerd en moeten er mensen voor worden opgeleid.

Een probleem bij FAIR data kan zijn dat voor afzonderlijke databronnen specifieke keuzes voor standaardisatie worden gemaakt, die niet afgestemd zijn met vergelijkbare databronnen. Dit kan de uitwisselbaarheid van data, ondanks dat er gebruik wordt gemaakt van standaarden, juist tegengaan. Een FAIR Implementation Profile (FIP) gaat dit probleem tegen door verschillende stakeholders met databronnen gezamenlijk een standaardisatie strategie te laten ontwikkelen. In het kader van dit project is er een start gemaakt met de ontwikkeling van een FIP voor het domein pandemische paraatheid. De FAIR assessment tools die zijn bekeken, zijn op dit moment nog te beperkt om breed gebruikt te kunnen worden. Verdere ontwikkeling moet leiden tot een tool die makkelijk in gebruik is, duidelijke scores geeft voor de realisatie per FAIR principe, en aanwijzingen geven voor verbeteringen. De beste oplossing voor assessment is een tool die de keuzes uit de FIP voor pandemische paraatheid gebruikt en automatisch een databron kan toetsen in hoeverre deze voldoet aan deze FIP.

5.3 FAIR data infrastructuur voor Pandemische Paraatheid

Aansluitend op de conclusies voor data beschikbaarheid, ontbreekt het in Nederland aan een centraal portaal waar alle relevante data voor Pandemische Paraatheid vindbaar zijn, bij voorkeur via een standaard als DCAT-Health. In België worden hiervoor via de Health Data Agency, het equivalent van de Nederlandse HDAB-NL serieuze stappen gezet. Als een aantal centrale portals binnen Nederland zoals die van Health-RI, de HDAB(s) en ODISSEI etc. FAIR goed hebben doorgevoerd, goed op elkaar aansluiten en alle voor pandemische paraatheid relevante data representeren, dan is Nederland al veel beter op de volgende pandemie voorbereid dan nu. Als vervolgens rijke meta-data en de onderliggende data bij verschillende

partijen echt FAIR worden gemaakt en dus herbruikbaar worden dan zal dat onderzoek stimuleren waarmee het beleid en zorg voor infectieziekten kan worden verbeterd.

Dit onderzoek laat ook zien dat bovenstaande initiatieven vooral zijn gefocust op centrale ontsluiting van historische data: data worden centraal verzameld waarna analyse plaatsvindt. Aan de ontsluiting van actuele data die vooral interessant zijn voor actuele beleidsbeslissingen, wordt binnen een aantal andere projecten gewerkt zoals GERDA, Heracles en Plugin. Om actuele data beschikbaar te maken is het van belang ze decentraal bij de databron te laten. Dit vergt wel het nodige van data infrastructuur om met respect voor privacy toch met deze data te kunnen werken. [GERDA](#), het [Heracles](#) project en [Plugin](#) zijn voorbeelden van een dergelijke decentrale data infrastructuur. FAIR databronnen zijn een voorwaarde om met gedecentraliseerde databronnen te werken.

Vanuit deze conclusies zijn er een groot aantal aanbevelingen opgesteld, sommige meer technisch, andere meer organisatorisch. Een volledige lijst van aanbevelingen is te vinden in [bijlage 5](#). In het volgende hoofdstuk worden de meest belangrijke aanbevelingen beschreven.

6 Adviezen en aanbevelingen

Vanuit dit rapport komt het beeld naar voren dat Nederland op dit moment niet pandemisch paraat is wat betreft data beschikbaarheid en herbruikbaarheid. Regie hierop is essentieel om Nederland wel goed voor te bereiden. Er zijn een aantal spelers die een belangrijke rol vervullen bij het verkrijgen van een betere beschikbaarheid en herbruikbaarheid van data relevant voor pandemische paraatheid. De belangrijkste zijn de systeemverantwoordelijken zoals VWS en de gemeentes. VWS pakt de regie over de beschikbaarheid en herbruikbaarheid van gezondheidsdata steeds beter op maar er is nog duidelijk ruimte en behoefte aan meer aansturing met name op de niet-curatieve zorg, gedragsdata en de actuele data. De belangrijkste organisaties, die aangestuurd door VWS, infrastructuur voor betere data beschikbaarheid organiseren, zijn Health-RI en HDAB-NL, maar ook instanties zoals RIVM, GGD-GHOR en het CBS. ZonMw is de belangrijkste financier van onderzoek rondom pandemische paraatheid in Nederland. Ten slotte is ODISSEI de belangrijkste speler als het gaat om de ontsluiting van data voor sociale wetenschappen, welke essentieel zijn om de sociaal-maatschappelijke effecten van een pandemie te kunnen inschatten.

In deze sectie volgen een aantal aanbevelingen om pandemische paraatheid wat betreft data beschikbaarheid en herbruikbaarheid te verbeteren. De aanbevelingen zijn steeds opgebouwd uit een constatering, de aanbeveling zelf en hoe en door wie deze aanbeveling kan worden opgepakt. In [bijlage 5](#) worden nog een groot aantal specifieke organisatorische en technische aanbevelingen opgevoerd. Hoewel deze aanbevelingen worden gedaan in het kader van pandemische paraatheid zijn ze ook van belang voor de ontwikkeling van een lerend zorgsysteem. Een lerend zorgsysteem draagt bij aan het creëren van een efficiënt zorgsysteem, een van de kernpunten van het ministerie van VWS.

6.1 Algemene aanbevelingen pandemisch paraatheid

Constatering: Zoals aangegeven is Nederland wat betreft data beschikbaarheid en herbruikbaarheid van data niet voorbereid op een eventuele nieuwe pandemie. Het is te verwachten dat een nieuwe pandemie zijn eigen data uitdagingen zal kennen. Databronnen die nu als niet relevant worden beschouwd kunnen mogelijk essentieel zijn voor de bestrijding van een nieuwe pandemie. Meest belangrijk, er ontbreekt centrale aansturing om het probleem van data beschikbaarheid en herbruikbaarheid van data voor pandemische paraatheid aan te pakken.

Aanbeveling: Om de urgentie van dit probleem aan te pakken moet er een ‘crisis’-team worden opgetuigd die zich specifiek richt op data herbruikbaarheid voor Pandemische Paraatheid door middel van FAIR organisatie van data. De term ‘crisis’ geeft aan hoe urgent de beschikbaarheid en herbruikbaarheid van data tijdens alle fases van een pandemie is. Dit team is dus niet alleen actief ten tijde van een pandemie, maar ook in de koude fase (pre- en post-pandemisch) van de pandemie.

Uitwerking: Dit team, aangestuurd door VWS, bestaat uit verschillende expertises, zoals juridisch expert(s), data steward(s) en domein experts. De belangrijkste taak van het crisisteam is het ontsluiten van relevante data voor pandemische paraatheid door middel van het creëren en onderhouden van een afspraakstelsel via één of meerdere FIPs (zie specifieke aanbevelingen voor FIPs in [bijlage 5](#)). Verder zou het crisisteam actief op zoek moeten gaan naar nieuwe relevante databronnen en deze (laten) ontsluiten. Daarnaast zal het regelmatig voorkomen dat er voor een bepaald data hergebruik geen grondslag is. Het team zou ook voor cruciaal data (her)gebruik grondslagen moeten overwegen, waar mogelijk samen met het obstakel verwijder traject (Health-RI), zodat acute (beleids)vragen kunnen worden beantwoord.

Probleem: geen centrale aansturing voor de aanpak van data beschikbaarheid en herbruikbaarheid voor pandemische paraatheid.

Wie: VWS, Health-RI.

Hoe: Opzetten van een team met juridische experts, data stewards en domein experts aangestuurd door VWS.

6.2 Aanbevelingen voor data beschikbaarheid

Constatering: Uit dit rapport komt naar voren dat Nederland een complex (zorg)data landschap heeft. Dit maakt het lastig om een goed overzicht te krijgen van welke partijen welke data produceren, door wie data worden verwerkt en wie de data gebruikers zijn. Deze bevindingen komen ook naar voren in het OECD rapport²¹ 'Toward an Integrated Health Information System in the Netherlands' uit 2022. Een dergelijk compleet overzicht is essentieel om pandemisch voorbereid te zijn maar ook om tot een lerend zorgsysteem te komen. Een belangrijk aspect van een dergelijk overzicht zouden de actuele data moeten zijn. De prioriteit voor het maken van een dergelijk totaal overzicht voor pandemische paraatheid zeker wat betreft actuele data, lijkt momenteel niet bij organisaties als Health-RI en HDAB-NL te liggen, met name doordat pandemische paraatheid additionele databronnen vereist dan in scope ligt van deze organisaties (zoals gedragsdata, data uit de langdurige zorg en publieke gezondheid).

Aanbeveling: Het ministerie van VWS zet een onderzoeksopdracht uit via ZonMw naar een externe onderzoekspartij om het versnipperde Nederlandse data landschap voor pandemische paraatheid met voldoende detail in kaart te brengen en een strategie op te stellen om deze data FAIR te ontsluiten.

Aanbeveling: Het ontwikkelen van een afsprakenstelsel voor Pandemische Paraatheid op basis van de beschikbare databronnen. Dit zou bij voorkeur worden uitgevoerd door een partij die sterk is verbonden met het FAIR gedachtengoed (zoals de publiek-private

²¹ [OECD - Health Information System NL - 17feb2022.pdf](#)

samenwerking LIFES). Ook organisaties zoals Nictiz en Nivel zouden aangesloten moeten zijn.

Uitwerking: Een voorbeeld van hoe een complex data landschap in kaart kan worden gebracht is te vinden in het data insights project²² van het NHS. Voor het FAIR ontsluiten van dergelijke data kunnen tools zoals een aangepaste versie van FAIRspace worden gebruikt of de tools zoals in ontwikkeling bij Health-RI. De eerste stap zou zich richten, op het in eerste instantie, FAIR vindbaar maken van deze data. Voor het maken van een afsprakenstelsel voor het hergebruik van data voor pandemische paraatheid kan het uitwerken van een netwerk van FIPs de basis leggen. De FIP die voor dit project is opgeleverd kan op hoog niveau afspraken vastleggen voor Pandemische Paraatheid en voor specifieke soorten van data kunnen dan FIPs worden gemaakt die voor die data aanvullende afspraken opnemen.

Probleem: er is geen duidelijk overzicht van het complexe (zorg)data landschap als het gaat om retrospectieve maar vooral als het gaat om actuele data.

Wie: VWS en ZonMw

Hoe: opdracht van ZonMw waarbij een externe onderzoeksorganisatie het versnipperde Nederlandse data landschap in kaart brengt. Op basis van deze kaart wordt een afsprakenstelsel ontwikkeld voor de herbruikbaarheid van data voor pandemische paraatheid.

6.3 Aanbevelingen rondom de organisatie van FAIR voor pandemische paraatheid

Constatering: Hoewel uit de bevindingen van dit rapport blijkt dat de FAIR principes bij veel partijen die een rol hebben in pandemische paraatheid zijn doorgedrongen, wordt tegelijkertijd geconstateerd dat is nog veel onbegrip is over wat FAIR nu in de praktijk echt betekent. Ook de verantwoordelijkheden en rollen van datahouders ten opzichte van de herbruikbaarheid van data voor bijvoorbeeld pandemische paraatheid is vaak onduidelijk. Voor verschillende gezondheid-gerelateerde vragen zijn ook sociaal-maatschappelijke en financiële data relevant. Deze data vallen vaak onder de verantwoordelijkheid van andere ministeries en andere wetgeving

Aanbeveling: Er moet door VWS een visie worden ontwikkeld op data herbruikbaarheid voor pandemische paraatheid. Deze visie, die voortborduurde op de visie die in dit rapport beschreven staat, moet duidelijkheid verschaffen over de rollen en verantwoordelijkheden van de partijen die databronnen beheren die geschikt zijn voor de pandemisch en de koude fase. De ontwikkelde visie moet ook duidelijkheid geven waar financiering vandaan moet komen, hoe de wet- en regelgeving voor data beschikbaarheid moet worden aangepast/aangevuld en welke data essentieel zijn om pandemisch paraat te zijn met dit rapport als belangrijk uitgangspunt. De opdracht voor zo'n uitgebreide visie zou moeten worden gegeven door VWS. Deze visie zal moeten worden afgestemd met andere ministeries zoals SZW, BZK en LNV.

²² [About – Data Insights](#)

Uitwerking: Aan de hand van de hier beschreven stakeholders en de relevante databronnen voor pandemische paraatheid wordt een visie uitgewerkt voor data beschikbaarheid voor pandemische paraatheid voor onderzoek, beleid en innovatie. Deze visie beschrijft stap voor stap: 1) Welke stakeholders welke rol en welke verantwoordelijkheden hebben. Daarin moeten de cruciale rollen van bijvoorbeeld de GGD, gemeenten, zorginstanties en burgers worden meegenomen. 2) Op welke manier worden data op een eenduidige manier FAIR gemaakt. 3) Welke databronnen extra aandacht (en financiering) en prioritering behoeven om pandemische paraat te zijn. 4) Wat het proces en de timing is van (componenten van) dit werkplan. Ook wordt beschreven of onderdelen pas in werking treden in het geval van een pandemie en wat weer op slot gaat indien de pandemie over is. 5) De rol van een 'crisis'-team in het organiseren van data beschikbaarheid en herbruikbaarheid.

Probleem: onbegrip en onduidelijkheid over wat FAIR in de praktijk betekent. Onduidelijkheid over rollen van datahouders.

Wie: VWS, SZW, BZK en LNV.

Hoe: Aan de hand van de hier beschreven stakeholders en de relevante databronnen voor pandemische paraatheid wordt een visie uitgewerkt voor data beschikbaarheid voor pandemische paraatheid voor onderzoek, beleid en innovatie.

6.4 Aanbevelingen rondom FAIR infrastructuur

Constatering: Aansluitend op de conclusies voor data beschikbaarheid, ontbreekt het in Nederland aan een centraal portaal waar alle relevante data voor Pandemische Paraatheid beschikbaar zijn. Dit portaal zal databronnen relevant voor de gezondheid, sociale en economische impact van een pandemie, vindbaar moeten maken. Een dergelijk portaal gaat daarmee dus verder dan het ontsluiten van gezondheidsdata wat momenteel binnen Health-RI en HDAB-NL de focus heeft. Gezien de aanvullende data behoeften voor pandemische paraatheid en de snelle dynamiek van ontsluiting die nodig is voor pandemische paraatheid (actuele data is cruciaal), zijn er aanvullende architecturale eisen voor dit domein. Gelukkig zijn er verschillende projecten die al werken aan relevante oplossingen.

Aanbeveling: VWS zou ook verdere regie moeten nemen op de ontwikkeling van FAIR data infrastructuur gericht op de ontsluiting van data voor pandemische paraatheid voor zowel (historische) data voor onderzoeksdoeleinden als voor de ontsluiting van voor beleid geschikte actuele data.

Uitwerking: Voor de data infrastructuur specifiek voor pandemische paraatheid, kan worden aangesloten en uitgebreid op data infrastructuren zoals die van Health-RI en de HDAB-NL (EHDS). Voor actuele data zou gebruik kunnen worden gemaakt van bestaande projecten zoals de gefedereerde architecturen van Heracles en GERDA. Daarbij moeten de herbruikbaarheid van data en data koppelingen actief worden gestimuleerd. Belangrijk is om op deze infrastructuren ook het GGD IV systeem aan te sluiten. Extra onderzoek rondom

data beschikbaarheid van actuele data kan doormiddel van gerichte ZonMw programma's worden gestimuleerd.

Probleem: het ontbreekt in Nederland aan een centraal portaal waar alle relevante data voor pandemische paraatheid beschikbaar zijn..

Wie: VWS, Health-RI, HDAB-NL, GERDA, HERACLES en ZonMw

Hoe: aansluiten op en uitbreiden van bestaande data infrastructuren. Extra onderzoek rondom data beschikbaarheid van actuele data stimuleren d.m.v. een gericht ZonMw programma.

7 Ondertekening

TNO › Healthy Living & Work › Leiden, 24 oktober 2024

Daan Kloet
Research Manager

Jildau Bouwman
Auteur

8 Bijlages

8.1 Bijlage 1 Vragenlijst

ZonMW FAIR data vragenlijst

Title / Hook

Is uw organisatie betrokken bij het verzamelen of analyseren van gegevens die relevant zijn voor infectieziekten met een endemisch of pandemisch risico? We hebben uw kennis nodig!

We zijn momenteel bezig met het identificeren van belangrijke stakeholders in het onderzoek naar infectieziekten met een endemisch of pandemisch risico, met speciale aandacht voor het bevorderen van FAIR data.

Als u werkzaam bent in dit veld of als uw organisatie relevante gegevens verzamelt of analyseert met betrekking tot infectieziekten met een endemisch of pandemisch risico, nodigen wij u van harte uit om deel te nemen aan onze enquête.

Uw deelname draagt direct bij aan het opbouwen van een veerkrachtigere respons op toekomstige pandemieën.

Meer informatie: <https://www.zonmw.nl/en/pandemic-preparedness>

#PandemicPreparedness #Infectieziekten #covid19 #FAIRdata

Intro

Wij zijn op zoek naar stakeholders die betrokken zijn bij onderzoek naar infectieziekten en/of pandemische paraatheid. Als u werkzaam bent in dit veld, of als uw organisatie relevante gegevens verzamelt of gebruikt met betrekking tot infectieziekten en/of pandemische paraatheid, dan nodigen wij u van harte uit om deel te nemen aan deze enquête.

Uw expertise en input zijn waardevol voor ons streven naar een effectievere benadering van pandemische paraatheid. Door deel te nemen aan onze vragenlijst draagt u bij aan de ontwikkeling van een FAIR data-infrastructuur die cruciaal is voor het onderzoek naar en de respons op pandemieën.

In opdracht van het ministerie van VWS voert ZonMw het Kennisprogramma Pandemische Paraatheid uit. Onderdeel van dit programma is het mogelijk maken van het werken met FAIR data in onderzoek naar pandemische paraatheid. ZonMW heeft TNO benaderd voor het opstellen van een kennissynthese met betrekking tot stakeholders en databronnen die van belang zijn voor een FAIR data-infrastructuur voor onderzoek naar pandemische paraatheid.

Het doel van de kennissynthese is ten eerste dat ZonMw daarmee een basis heeft voor het programmeren van onderzoek naar pandemische paraatheid. Ten tweede moet de kennissynthese ook voor andere partijen geschikt zijn om databronnen verder te ontwikkelen en pandemisch paraat te maken.

Vragen

Toestemmingsverklaring voor gegevensverwerking

Door deze enquête in te vullen, ga je akkoord met het volgende:

- Je gegevens worden opgeslagen en verwerkt voor analytische en onderzoeksdoeleinden.
- Indien je je e-mailadres onderaan deze vragenlijst achterlaat, stem je ermee in dat we contact met je kunnen opnemen voor eventuele vervolgvragen met betrekking tot deze enquête.

Je hebt het recht om op elk moment je toestemming in te trekken door contact met ons op te nemen via **<e-mailadres contact>**. Het intrekken van je toestemming heeft geen invloed op de rechtmatigheid van de verwerking op basis van de toestemming vóór de intrekking.

- Ik ga akkoord met de bovenstaande voorwaarden.

Checkbox (verplichte vraag)

- Is uw organisatie betrokken bij het verzamelen of analyseren van gegevens die relevant zijn voor infectieziekten of pandemische paraatheid?

Radiobuttons: Ja / Nee. (verplichte vraag)

- Voor wat voor een organisatie bent u werkzaam?

Radiobuttons: Universiteit / Kennisorganisatie / Bedrijf / Overheid / Anders, namelijk... (verplichte vraag)

- Wat is de naam van uw organisatie?

Open vraag (niet verplichte vraag)

- Welke specifieke data worden verzameld of geanalyseerd die relevant zijn voor infectieziekten of pandemische paraatheid? Voeg URL's nodig waar mogelijk.

Open vraag (verplichte vraag)

- Voldoen deze data aan de FAIR (Findable, Accessible, Interoperable, Reusable) principes?

Radiobuttons: Ja / Nee / Weet ik niet /Ik weet niet wat 'FAIR' betekent (verplichte vraag)

- Welke tools gebruiken jullie om data FAIR te maken?

Checkboxes met voorbeelden (verplichte vraag)

- FAIR Checker
- F-UJI
- FAIR Evaluation Services
- FAIR Data Endpoint
- FIP Wizard

- Anders, namelijk

- Staan de datasets die jullie gebruiken in catalogi, zoals die van Health-RI of Europese data portalen?

Checkboxes met voorbeelden (verplichte vraag)

- Data.overheid.nl
- Data.rivm.nl
- Health-RI
- BioPortal
- Pubmed
- CBS
- Nivel
- Anders, namelijk

- Welke problemen verwacht je dat het FAIR maken van de data (deels) kan oplossen?

Open vraag (geen verplichte vraag)

- Op basis van uw antwoorden vinden wij het wellicht interessant om hierover met u te spreken. Als u hiervoor openstaat, kunt u hieronder uw e-mailadres achterlaten. Wij zullen dan contact met u opnemen.

Invulveld e-mailadres (geen verplichte vraag)

8.2 Bijlage 2 Stakeholders

Zie volgende pagina alleen digitaal leesbaar door in te zoomen

8.3 Bijlage 3 Lijst van Databronnen voor pandemische paraatheid

Zie volgende pagina alleen digitaal leesbaar door in te zoomen

8.4 Bijlage 4 PP-FIP Pandemische paraatheid FAIR implementatie profiel.

Meta-Data FIP						
Publication Date	2024-10-22T08:22:54.941642604Z					
Community	GloPID-R Global Research Collabora					
Data Steward	0000-0002-5448-0543					
Start Date	2024-05-06					
End Date	2091-03-17					
Data FIP						
FAIR Enable Resource (FER)			Replacement FER			
Question	Name	URI	Usage	Name	URI	Considerations
F1 MD	URI Uniform Resource Identifier	http://purl.org/	currently in use			We prefer the use of DOI for the reason of strong persistency. However, acquiring a DOI for any dataset can be difficult, and therefore a simpler alternative is necessary currently. Here we choose the URI because: a) easy to create for many partners, easier than DOI. (Several relevant examples, such as data.overheid don't have DOI), b) subsumes other formats such as PURL. c) dereferenceable which UUID is not.
F1 MD	DOI Digital Object Identifier	http://purl.org/	planned use			We prefer the use of DOI for the reason of strong persistency. It is currently not used widely for this application, but we believe this should be the direction for the future.
F1 D	URI Uniform Resource Identifier	http://purl.org/	currently in use			We prefer the use of DOI for the reason of strong persistency. However, acquiring a DOI for any dataset can be difficult, and therefore a simpler alternative is necessary currently. Here we choose the URI because: - easy to create for many partners, easier than DOI. (Several relevant examples, such as data.overheid don't have DOI) - subsumes other formats such as PURL. - dereferenceable which UUID is not.
F1 D	DOI Digital Object Identifier	http://purl.org/	planned use			We prefer the use of DOI for the reason of strong persistency. It is currently not used widely for this application, but we believe this should be the direction for the future.
F1 D	UUID Universally Unique Identifier	http://purl.org/	to be replaced	URI Uniform Resource Identifier	http://purl.org/	The UUID is not dereferenceable, so one cannot send a SPARQL query to a UUID. This is a required future, so the UUID should be replaced in the future, easiest to be done with the URI.
F2	DCAT Data Catalog Vocabulary Version 3	http://purl.org/	currently in use			
F2	DC Dublin Core	http://purl.org/	currently in use			
F2	FDP FAIR Data Point	http://purl.org/	currently in use			This schema for metadata is currently in use, but its use can and should be expanded. It is more elaborate than DCAT and DC, and uses elements of both, so the FDP is the optimal choice when setting up the most expansive metadata.
F3	DOI Digital Object Identifier	http://purl.org/	planned use			In the future, we would require the datasets to have a DOI as GUPRI. This answer, in combination with the use of metadata schemas listed for F2, declare a schema of linking metadata and data. The DOI must be present in the metadata, adhering to a schema per F2. When using an URI as GUPRI for datasets, the same applies.
F4 MD	COVID-19 metadata and data portal	http://purl.org/	currently in use			We see this as a mandatory location for publishing metadata records.
F4 MD	BioStudies	https://doi.org/	planned use			Next to the COVID-19 portal, this location is a nice-to-have for publishing metadata records
F4 D	COVID-19 metadata and data portal	http://purl.org/	currently in use			We see this as a mandatory location for publishing datasets.
A1.1 MD	HTTPS Hypertext Transfer Protocol Secure	http://purl.org/	to be replaced	SPARQL (open) endpoint	http://purl.org/	We view the SPARQL endpoint as the preferable option and encourage replacing HTTPS protocols in many situations. However, when the HTTPS method provides the metadata in RDF/Turtle nicely, that can suffice as well.
A1.1 MD	SPARQL (open) endpoint	http://purl.org/	currently in use			
A1.1 D	HTTPS Hypertext Transfer Protocol Secure	http://purl.org/	to be replaced	REST Representational state transfer	http://purl.org/	We favor REST as an addition to just HTTPS.
A1.1 D	REST Representational state transfer	http://purl.org/	currently in use			
A1.1 D	SPARQL (open) endpoint	http://purl.org/	planned use			
A1.1 D	FTP File Transfer Protocol	http://purl.org/	currently in use			
A1.2 MD	OpenID Connect	http://purl.org/	currently in use			The OpenID Connect can be suitable as it is underlying technology for the SURF and Eduroam technologies. However, other options might be considered here as well and can be added to this FIP.
A1.2 D	OpenID Connect	http://purl.org/	currently in use			The OpenID Connect can be suitable as it is underlying technology for the SURF and Eduroam technologies. However, other options might be considered here as well and can be added to this FIP.
I1 MD	RDF Resource Description Framework	http://purl.org/	currently in use			This is suitable in combination with the selections made for F2.
I1 MD	Turtle Format	http://purl.org/	currently in use			We add this option since it is both machine and human-readable, relative to just RDF which is only machine-readable.
I1 D	JSON JavaScript Object Notation	http://purl.org/	currently in use			We choose here for Linked Data options like RDF and JSON-LD, which are our general preference. However, conversion of datasets to this might not be desirable in every situation, so a JSON notation can suffice in those cases.
I1 D	CSV File Format	http://purl.org/	to be replaced	JSON-LD JavaScript Object Notation for Linking Data	http://purl.org/	The CSV (or TSV) format is commonly used but lacks all the useful features of the RDF/Linked Data options listed here, and should be replaced in the future.
I1 D	RDF/XML XML syntax for RDF	http://purl.org/	currently in use			
I1 D	JSON-LD JavaScript Object Notation for Linking Data	http://purl.org/	currently in use			
I1 D	Turtle Format	http://purl.org/	currently in use			
I2 MD	DCAT Data Catalog Vocabulary Version 2	http://purl.org/	currently in use			This intentionally coincides with the decision for F2
I2 D	Schema.org	https://w3c.org/	currently in use			Many options can be correct in specific contexts, but the schema.org can be a good starting point.
I3 MD	SNOMED CT ontology SNOMED Clinical Terms Ontology	http://purl.org/	planned use			
I3 D	SNOMED CT ontology SNOMED Clinical Terms Ontology	http://purl.org/	planned use			
R1.1 MD	CC BY-ND 4.0 Attribution-NoDerivatives 4.0 International	http://purl.org/	currently in use			This choice promotes the metadata to be freely available.
R1.2 MD	PROV-O W3C PROV Ontology	http://purl.org/	planned use			We choose the PROV Ontology that works well in combination with the other choices for linked data implementations.
R1.2 D	PROV-O W3C PROV Ontology	http://purl.org/	planned use			We choose the PROV Ontology that works well in combination with the other choices for linked data implementations.

8.5 Bijlage 5 lijst van aanvullende aanbevelingen

Algemene aanbevelingen voor data beschikbaarheid voor pandemische paraatheid.

- Data beschikbaarheid heeft binnen de looptijd van projecten vaak aandacht van de stakeholder, echter na afloop van een project is er geen capaciteit en financiering meer om deze beschikbaarheid te onderhouden. Daarom zou de systeemverantwoordelijke (VWS en gemeentes) voor niet projectmatig werk, eisen moeten stellen aan data(bron)beheerders en ook de naleving moeten toetsen. De EHDS en als uitvoeringsinstantie de HDAB-NL zou voor zorgdata deze rol kunnen spelen. Financiers van onderzoek (specifiek voor pandemisch gerelateerd onderzoek) kunnen in het datamanagementplan toetsen of het project aan deze logica voldoet en zullen aan het einde van het project moeten toetsen of het plan (volledig) is uitgevoerd. Hiervoor zal wel extra budget voor gereserveerd moeten worden.
- Als partijen data aanleveren aan databronnen (FAIR data points) van partijen met wettelijke taken zoals CBS en RIVM, moeten deze data ook goed beschikbaar blijven voor de toeleverende partijen. Systeemverantwoordelijken moeten hierop toezien.

Incentives

- Geef waar mogelijk als financiers en systeemverantwoordelijken duidelijk aan welke partijen wanneer toegang zouden moeten krijgen tot de data en onder welke voorwaarden. Geef gefinancierde projecten voorbeelden mee van licenties waar zij mee kunnen/moeten werken.
- Stakeholders die een groot belang hebben bij pandemische paraatheid zoals patiënten en zorgprofessionals hebben vaak weinig invloed op data beschikbaarheid. Een deel van de oplossing om data beschikbaar te krijgen kan liggen in het verschuiven van de invloed naar groepen die een direct belang hebben. Patiëntorganisaties kunnen hierbij een belangrijke rol spelen. Zo kunnen patiënten consent geven voor het hergebruik van de data in hun PGO en kunnen patiëntorganisaties ondersteunen in de toetsing van de consent aanvragen.
- Financiers en systeemverantwoordelijken zouden een mate van FAIRness als voorwaarde moeten stellen als dit voor het algemeen belang zoals pandemische paraatheid belangrijk is. Hier moet dan ook een specifieke financiële ruimte voor zijn. Deze mate van FAIRness moet in overeenstemming zijn met het doel, domein en omvang van het betreffende project, zodat een redelijke inspanning van de datahouder wordt gevraagd. Deze mate van FAIRness zou kunnen worden geoperationaliseerd via een FIP (zie verder onder aanbevelingen FIP).
- Het a priori gebruik maken van gestandaardiseerde modellen of vocabulaires voor de data, in plaats van ad hoc de data structureren, wordt als te grote tijdsinvestering gezien en behoeft een directe incentive om wel aangepakt te worden. Incentives

zouden kunnen worden gevonden als het makkelijker wordt om onderzoek te publiceren, dat er wetenschappelijke waarde wordt gehecht aan het beschikbaar maken van data (via een H-index) of via druk en controle vanuit de financier.

Wet en regelgeving

- Om pandemisch paraat te zijn moet ook de wet- en regelgeving goed voorbereid zijn. Er moet bijvoorbeeld worden onderzocht hoe data beschikbaarheid tussen het gezondheidsdomein en het sociale domein kan worden georganiseerd en hoe wettelijke taken van partijen veranderen in geval van een pandemie (en mogelijk weer terug veranderen na een pandemie). De ministeries van VWS en SZW zouden hier opdracht voor moeten geven.
- Er zou getoetst moeten worden of data die essentieel zijn voor pandemische paraatheid allemaal onder de EHDS beschikbaar komen of dat er nog aanvullende acties voor pandemische paraatheid nodig zijn. VWS zou hiervoor een opdracht moeten uitzetten.

Technisch

- De (meta-)data voor pandemische paraatheid moet worden meegenomen in de ontwikkeling van het algemene Health-RI portal en de HDAB-NL portal. Deze portals moeten beschikbaar zijn voor pre-pandemische, pandemische en post-pandemische vragen.
- Als het gaat om pandemische paraatheid en om de toepassing van Real World Data dan zou de EHDS kunnen worden aangegrepen om interoperabiliteit af te dwingen. Als ziekenhuizen nieuwe EPDs gaan aanschaffen, kan datastandaardisatie volgens de EHDS meteen mee worden genomen (zie kamerbrief 21 maart 2024 over de EHDS verordening). VWS zou ervoor kunnen zorgen dat de ziekenhuizen en ICT-leveranciers zich hiervan bewust zijn. Zorg daarnaast dat de huidige aanbieders nu al met een plan komen hoe zij EHDS gaan toepassen en zorg dat helder is wat moet worden gestandaardiseerd (met de vraag of Nederland misschien verder wil gaan dan nu is afgesproken in Europa). Rond pandemische paraatheid zou in kaart moeten worden gebracht welke data uit de EPDs en daarbuiten relevant zijn.
- De analyse van de databronnen en de interviews heeft laten zien dat er nauwelijks aandacht is voor het voldoende toevoegen van rijke meta-data. Dit is echter wel essentieel voor herbruikbaarheid en zou daarom actief door financiers moeten worden gestimuleerd en getoetst. Hiervoor zouden project calls kunnen worden geformuleerd die voor specifieke domeinen deze vereisten zouden kunnen vastleggen.
- Door volledig FAIR te werken kan data op verschillende plekken zichtbaar zijn zonder dat duplicatie nodig is. Dit voorkomt meer opslag dan nodig is en vermindert het risico op verdubbeling van sommige data in de analyse. Hiervoor is het van belang dat data bij de bron FAIR wordt gemaakt en beschikbaar komt in een FAIR data point (zie ook onder architectuur aanbevelingen).

Aanbevelingen rondom de organisatie van FAIR voor pandemische paraatheid

- Nivel, Nictiz en andere koplopers in datastandaardisatie kunnen een relevante rol spelen in beschikbaarheid van primaire zorgdata en daarmee inspirerend zijn voor andere partijen. Financiers kunnen de FAIR datakoplopers in de schijnwerpers zetten tijdens specifieke bijeenkomsten, waaronder workshops om hun kennis over te dragen naar andere partijen.
- Gemeenten vallen vaak onder andere wet- en regelgeving en hebben andere zorgtaken. Data uit deze regionale rol moeten in geval van een pandemie soms samen kunnen worden geanalyseerd met nationale data. De [VNG Datastrategie](#) en het [GALA akkoord](#) beschrijven een visie voor het domein overstijgend hergebruiken van data, de vraag is of daarvoor genoeg ruimte is in de huidige budgetten. VWS en SZW zouden hierop een gezamenlijke visie moeten ontwikkelen. Er zijn bijna geen bronnen voor gehandicaptenzorg en verpleeghuiszorg herbruikbaar, waardoor de effecten van de pandemie op deze groepen niet goed onderzocht kan worden. Hier zouden middelen voor beschikbaar moeten worden gemaakt (door de centrale en regionale overheid afhankelijk van de verantwoordelijke van de bron). Dit vergemakkelijkt pandemisch onderzoek met behulp van deze bronnen.
- Er zou nog meer aandacht moeten komen voor het ontsluiten van data voor gedragsanalyses, zoals die binnen BePrepared worden gebruikt. ODISSEI ontsluit nu met name data voor grote onderzoekstrajecten, maar er is minder aandacht voor real-life gedragsdata. Hierop zullen de ministeries van VWS en SZW gezamenlijk een strategie moeten ontwikkelen om pandemisch paraat te zijn.
- Er moet aandacht komen voor het ontsluiten van data rond pandemische paraatheid signalen en ziekte symptomen bijvoorbeeld door middel van wearables (Bring Your Own Device) en het bijhouden van bijwerkingen met behulp van applicaties. Ook moet verder worden onderzocht of en hoe de Persoonlijke Gezondheids Omgeving (PGO) hier een rol bij kan spelen. Mogelijk kunnen de PGOs, in geval van een nieuwe pandemie, de FAIR data beschikbaarheid van Real World Data (klinische en zelf verzameld) en de participatie van burgers kunnen verhogen. Dit onderzoek zou door ZonMw of andere financiers kunnen worden gesubsidieerd.
- Omdat de rol van data steward cruciaal is voor data beschikbaarheid, moet deze rol beter gefinancierd worden door financiers en systeemverantwoordelijken en moet er in het onderwijs meer aandacht voor komen (afstemmen met universiteiten en OCW).
- Organisaties die databronnen beschikbaar maken en/of beheren zouden zelf deze FAIR assessment moeten uitvoeren en als meta-data toevoegen aan de data (zie FIP). Dit is werk dat de data steward van de organisatie zou moeten uitvoeren.
- Benoem enkele centrale data stewards die vanuit de systeemverantwoordelijken en financiers betrokken zijn bij de ontwikkeling van FIPs en data beschikbaarheid voor pandemische paraatheid. Deze data stewards zouden ook deel moeten worden van het data beschikbaarheids-crisisteam.

Aanbevelingen rondom FAIR infrastructuur

- De GGD ontwikkelt nu nieuwe systemen specifiek voor pandemische paraatheid. Dit zou in samenhang met ander data initiatieven moeten gebeuren en niet in isolatie. VWS zou hier regie moeten nemen.
- Stimuleer data-houdende partijen om rekening te houden met hergebruik van deze data vanuit hun systeem en hoe deze gekoppeld kunnen worden met andere bronnen voor pandemische paraatheid. Financiers en systeemverantwoordelijken zullen hierop moeten aansturen.

- Onderzoek wat er nog nodig is aan aanpassing van wetgeving met name op de interacties met andere domeinen (bijvoorbeeld zorg en sociale data).
- FAIR data maken het mogelijk om data op een gefedereerde en op een privacy vriendelijk manier te hergebruiken. Verschillende initiatieven (zoals [GERDA](#) en [Heracles](#)) hebben laten zien dat dit werkt. Voor volledige implementatie (hogere TRL-niveaus van de componenten) van deze concepten is verdere (financiële) ondersteuning van belang. Wetenschappelijke financiering past hier niet en de voorinvestering is nog te groot voor commerciële partijen. Omdat dit soort architectuur ook voor andere domeinen relevant is, is het nuttig financiering mee te nemen in de [Interbestuurlijke Datastrategie](#) van het ministerie Binnenlandse Zaken.

Aanbevelingen voor een FIP voor Pandemische paraatheid

- Maak FIPs voor Pandemische paraatheid. Voor kleinere partijen is FAIR data creëren uitdagend (over het algemeen geen data stewards), tijdrovend en kostbaar. Huidige initiatieven waarin dit wel gebeurt (bijvoorbeeld via ZonMw funding) leveren vaak toch niet het gewenste resultaat op (hergebruik van data), omdat er geen eenduidige standaarden etc. worden gebruikt. Een FIP zou hier kunnen helpen omdat deze ervoor zorgt dat de betreffende onderzoeksgemeenschap een expliciete keuze maakt voor een of slechts enkele standaarden waar andere onderzoekers in dat domein gevraagd worden/aangespoord worden (bijv door de funder) om die keuzes over te nemen. Dat zorgt voor beter uitwisselbaarheid van data. (zie ook aanbevelingen FIP).
- Advies aan financiers en systeemverantwoordelijken is om een consortium van relevante stakeholders (zie ook de tabel van pandemische paraatheid stakeholders) de opdracht te geven om voor pandemische paraatheid één of meerdere FIPs te creëren (voor verschillende typen data kan een specifieke FIP nodig zijn) en deze uit te dragen en bij te houden. In dit project moet rekening worden gehouden met pre-pandemische, pandemische en post-pandemische data beschikbaarheid. Bepaalde keuzes in de FIP zullen veel werk kosten om te realiseren, maar dat zal nodig zijn om de data echt FAIR aan te bieden. Echter, het is ook niet wenselijk als bepaalde keuzes echt onhaalbaar blijken, dat deze om die reden niet worden nageleefd. Hier moet dus een balans in worden gezocht. De FIP die is ontwikkeld in dit project is een startpunt voor de ontwikkeling van deze PP-FIP set.
- Daarnaast moet de FIP blijvend onderhouden worden door (een) vertegenwoordiger(s) uit het onderzoeksgebied (bijvoorbeeld RIVM, GGD GHOR, Health-RI en/of ODISSEI). We adviseren om deze aan te stellen en om deze mensen in contact te brengen met projecten die de PP-FIP gaan gebruiken. Voor deze taak moet structurele financiering worden geregeld. Het is belangrijk dat ZonMw het bestaan van de PP-FIP promoot en het gebruik ervan aanraadt en uitlegt voor nieuwe projecten, aangezien men niet altijd direct zal begrijpen waar een FIP toe dient. Dit advies zou ook door andere financiers moeten worden gegeven bij onderzoeksprojecten en er zou bij stakeholders in de zorg en beleid (met name systeemverantwoordelijken) ook meer kennis over moeten komen.
- Behalve het vindbaar maken van data is de interoperabiliteit ook een uitdaging: data is vaak niet gestructureerd of machine-leesbaar waardoor het niet (her)bruikbaar is. Het ontwikkelen van gemeenschappelijke datamodellen voor onderzoek naar infectieziekten is cruciaal. Een verdere samenwerking en afstemming binnen het netwerk en met externe partijen, zoals ZonMw, is van belang om de voorgestelde methoden en systemen te implementeren. Sommige partijen hebben te maken met afstemming van regionale en landelijke onderdelen, dat leidt tot een gebrek aan overzicht van beschikbare data. Echter er is een behoefte aan betere coördinatie en centralisatie van data om pandemische paraatheid te verbeteren. Deze organisatie moet uitgewerkt worden in een plan. Hiervoor moet VWS opdracht geven.

- Leid rondom een FIP ook personeel op, zoals data stewards, maar ook domeinexperts die de relevantie begrijpen.
- De FIP-wizard is een goede tool om de FIP verder te ontwikkelen, dus deze kan gebruikt blijven worden. Het verder ontwikkelen van de FIP wizard om deze gebruikersvriendelijk te maken, vereist middelen die op dit moment niet beschikbaar zijn. Dit soort middelen zijn lastig vrij te maken uit onderzoekfinanciering. De rijksoverheid en onderzoeksfinanciers moeten gezamenlijk een visie ontwikkelen hoe digitale middelen voor onderzoek kunnen worden onderhouden.
- Stimuleer het gebruik van een FIP en FAIR assessment tools en het vastleggen van de uitkomst van de assessment als metadata bij de data. Deze uitkomsten kunnen financiers en systeemverantwoordelijken gebruiken om naleving van FAIR te controleren, dit moet worden opgenomen in de processen van deze organisaties.
- De aard van de FAIR principes maakt het lastig om goede generieke FAIR assessment tools te maken. De heilige graal is hierbij een algemene tool die de FIP gebruikt als configuratie en daarmee onderzoek kan uitvoeren op bronnen. De mogelijkheden hiervoor moeten eerst verder worden onderzocht. Daarom adviseren wij dit onderzoek te bevorderen. In ons gesprek met de ontwikkelaars van de FIP-wizard is gebleken dat al vooronderzoek wordt gedaan naar het configureren van een FAIR assessment tool met een FIP.
- De FAIR gemeenschap heeft nog niet gezamenlijk besloten welke FAIR assessment tools gebruiksvriendelijk zijn en nuttige informatie geven. Er zijn wel verschillende tools, die nog in ontwikkeling zijn en we adviseren ZonMw, andere financiers en systeemverantwoordelijken om ten minste deze ontwikkelingen goed te blijven volgen en vervolgens de tools te promoten als deze ver genoeg ontwikkeld zijn.
- Het goed vastleggen van gemeenschapsstandaarden kan ook bedrijvigheid stimuleren, omdat services makkelijker ontwikkelbaar zijn als de data gestandaardiseerd worden ontsloten. Onderzocht moet worden welke kansen er nu blijven liggen, ook omdat dit deels de kosten van FIPs en FAIR maken van data kan dragen. Hier zouden financiers een subsidie ronde op kunnen openen.
- Als er voor andere velden rondom pandemische paraatheid ook een FIP wordt opgesteld, dan is het nuttig om deze FIP's te vergelijken door middel van een [FAIR Convergence Matrix](#). Daarmee kan worden uitgezocht in hoeverre de verschillende FIP's met elkaar overeenkomen of van elkaar verschillen. De gaten in de FIP's uit enkele gemeenschappen kunnen mogelijk worden opgevuld met keuzes uit die van de ander. FAIR principes waar alle gemeenschappen dezelfde keuzes voor maken, kunnen worden opgewaarderd naar keuzes voor een niveau hogere FIP, indien deze bestaat en gebruikt wordt. Voor keuzes die conflicteren moet of eensgezindheid worden gevonden of moeten er oplossingen worden ontwikkeld om tussen databronnen data te kunnen mappen.

Aanbevelingen rondom (Generatieve) AI en FAIR data

- Financiers moeten de ontwikkeling rondom LLM beter stimuleren. De snelle ontwikkeling van (generatieve) AI (Gen AI) stimuleert ook het FAIR data veld. Zo kunnen data sneller FAIR worden gemaakt met large language models (LLM) door bijvoorbeeld de LLM de termen te laten mappen op vocabulaires of ontologieën. Stimuleer als financiers deze ontwikkelen.
- Gen AI kan ook synthetische data genereren op grond van een databron. Dit kan herbruikbaarheid van data mogelijk maken indien er ethische en/of juridische issues zijn om deze data zelf niet beschikbaar te maken. Stimuleer als financiers deze ontwikkelen.
- Eén van de grote problemen in het gebruik van Gen AI is het hallucineren van de modellen. Als Gen AI-modellen worden getraind op kwalitatief goede data (FAIR) dan is

er genoeg provenance om inzicht te geven in de kwaliteit van de uitkomst van een Gen AI-model. Dit is een extra incentive voor FAIR data, communiceer hierover.

- De analyse van de bronnen en interviews laten zien dat er in weinig bronnen gebruik wordt gemaakt van vocabulaires, ontologieën en URI's. Dit hoeft niet altijd in de bron te gebeuren maar kan ook later door middel van een mapper worden toegevoegd. De snelle ontwikkeling van Large Language Modellen (LLMs) kunnen hierbij behulpzaam zijn. Voor de zorg kan de EHDS hier helpen omdat deze gaat vereisen dat data (her)bruikbaar wordt en vereist ook het gebruik van specifieke vocabulaires (zie ook hierboven).

-

Additionele aanbevelingen pandemische paraatheid.

- FAIR algoritmes zijn ook van essentieel belang, en moeten beter gestimuleerd worden.
De scope van FAIR heeft zich hier voor dit rapport exclusief gericht op FAIR data en FAIR databronnen. Waar geen aandacht voor is geweest is voor FAIR algoritmes. Het advies is om ook FAIR algoritmes te gaan stimuleren. Start met een onderzoek naar de wenselijkheid en haalbaarheid hiervan in het kader van pandemische paraatheid en check wat hieraan gebeurt in het ZonMw 'modeling voor pandemische paraatheid' programma.
- Ondersteun van infrastructurele projecten waarbij citizen science verder wordt uitgewerkt voor pandemische paraatheid. Dergelijke projecten zouden op innovatieve manieren gegevens rondom de vogelgriep of ander virussen die van dier naar de mens overspringen kunnen helpen verzamelen (data van boerderijen etc., tellen en fotograferen van dode vogels etc.).
- In het onderzoeksgebied rondom FAIR data zijn [FAIR Data Objects](#) (FDO) en [Nanopublicaties](#) recente ontwikkelingen waarin de FAIR ideeën worden voortgezet. In een FDO worden alle (meta) datagegevens gestructureerd en gerepresenteerd door een unieke identifier. In een Nanopublicatie wordt elke bewering, uitgedrukt in RDF triples (de kleinste eenheid om informatie gestructureerd op te kunnen slaan zoals 'COVID-19' 'is' 'een ziekte') en wordt ook de herkomst en publicatie-informatie toegevoegd. De FDOs en Nanopublicaties zijn allebei een uitbreiding van de functionaliteit van het meer simpele FAIR Data Endpoint (de machine-leesbare representatie van een databron). Daarom adviseren we hierover de ontwikkelingen goed in de gaten te houden en te ondersteunen met projectfinanciering. Uiteindelijk moet onderzocht worden of de toegevoegde expressiviteit van FDOs en Nanopublicaties inderdaad overlapt met de wensen van gebruikers en hoe.
- Van cruciaal belang is dat er uiteindelijk een éénduidige definitie komt van hoe je data machine-leesbaar maakt (vergelijkbaar met de TCP/IP waarmee het internet voor iedereen bruikbaar is geworden). Alleen dan is het mogelijk om alle FAIR gemaakte data te kunnen vinden en interpreteren. Het risico bestaat dat zonder deze definitie er FAIR data silo's kunnen blijven bestaan. Het blijft lastig om voor dit soort zeer fundamentele ontwikkelingen middelen (tijd en geld) te vinden. Zie ook hierboven de aanbeveling voor het uitwerken van een digitale strategie voor onderzoek.
- Privacy wetgeving wordt regelmatig gebruikt als argumentatie om data niet FAIR te maken. Ten eerste is van belang dat het concept FAIR daarvoor beter begrepen wordt. De A beschrijft namelijk onder welke condities de data toegankelijk is (ook voor machines leesbaar). Daarmee kan dus beschreven zijn dat er consent nodig is om data voor een bepaald onderzoek te kunnen hergebruiken. Voor het delen van metadata (bv. welke variabelen er in de database zitten, en hoeveel mensen er geïncludeerd zijn), zijn geen privacy bezwaren. En dit is vaak juist de informatie die nodig is om te weten of een database hergebruikt van worden. Daarnaast is er heel wat informatie beschikbaar hoe data anoniem (https://rdmkit.elixir-europe.org/sensitive_data) te maken en hoe bijvoorbeeld de AVG te interpreteren

(<https://www.coreon.org/gedragscode-gezondheidsonderzoek/>). De HDAB(-NL) krijgt hierin een belangrijke taak en deze communicatie moet binnen het huidige project worden uitgewerkt.

- Het proces van FAIRificatie kan verder worden ondersteund door middel van bijvoorbeeld video's, documentatie en tutorials, zodat het proces makkelijker wordt gemaakt voor de gebruiker. Een onderwerp waarover gecommuniceerd moet worden is het belang van FAIRificatie en de mogelijkheden binnen de geldende wetgeving. We merken namelijk dat data-eigenaren het makkelijk vinden te beantwoorden hoe ze data van andere partijen willen ontvangen, maar het moeilijk vinden van de eigen data te identificeren hoe het meer FAIR gemaakt kan worden. Een rollenspel of serious game kan helpen om bij de data-eigenaren duidelijk te maken hoe belangrijk hun taak van FAIRificatie is in de wetenschappelijke omgeving, binnen de zorg of bij beleidspartijen. Health-RI (en Cumuluz zodra echt gestart voor de zorg) zouden dit mogelijk kunnen oppakken binnen hun huidige taak.