



Rijksinstituut voor Volksgezondheid
en Milieu
Ministerie van Volksgezondheid,
Welzijn en Sport

nota

MT I&V

A. van Leeuwenhoeklaan 9
3721 MA Bilthoven
Postbus 1
3720 BA Bilthoven
www.rivm.nl

KvK Utrecht 30276683

T (10)(2e)
info@rivm.nl

Nota nummer

Datum
14 mei 2020

Betreft:	Voorstel ondersteuning inrichten FAIR Data Point COVID-19 data
Auteur	(10)(2e) (10)(2e) (10)(2e) (10)(2e)
Afgestemd met	(10)(2e) (10)(2e) (10)(2e) (10)(2e) (10)(2e)
Besluit	Ja
IDocument	Nee
Medezeggenschap	n.v.t.
Gevolgen voor ROS	Ja
Bijlage(n)	1

Doel

Vaststellen of er draagvlak is bij MT I&V:

- voor het FAIR maken van data over Covid-19
- om dat in de vorm van een FAIR data point te doen
- of ondersteuning door de IV-organisatie zoals beschreven in het voorstel, gewenst is.

Kernpunten

- Vanuit de samenleving en de onderzoekswereld wordt gevraagd om transparantie en het delen van data over COVID-19. Er is een eerste stap gezet door het publiceren van open data.
- Voor wetenschappelijk onderzoek is echter toegang nodig tot meer gedetailleerde data. Nationaal en internationaal zijn er verschillende initiatieven gestart om data over COVID-19 beschikbaar en herbruikbaar te maken voor wetenschappelijk onderzoek. ZonMw vraagt de Nederlandse onderzoekers (en het RIVM) aan te sluiten bij de initiatieven van Health RI (nationaal) en VODAN (internationaal)
- We stellen voor dat CIO office en SIM uitzoeken wat de meest geëigende weg is om COVID-19 data te delen. Op basis hiervan komen CIO office en SIM met een voorstel hoe RIVM COVID-19 data het beste volgens FAIR principes ter beschikking kan stellen.
- Op basis van goedkeuring van dit voorstel, zullen CIO office, SIM en SSC Campus een FAIR data point inrichten bij het RIVM. Hierbij zullen de organisatorische en infrastructurele kant van het Fair data point door CIO office, SIM en SSC Campus geregeld worden. Onderzoekers en centra die hun data willen delen (bijvoorbeeld CIB en V&Z) zorgen voor een metadata beschrijving Het streven is de onderzoekers

zo veel mogelijk te ontlasten, met simpele en logische procedures, eventueel met ondersteuning van de datasteward.
 - Commitment vanuit de relevante centra (bij I&V en V&Z) om hun data als FAIR data te willen delen is een vereiste.

Toelichting

Zie bijlage

MEDEZEGGENSCHAP

RISICO (t.a.v. financiële, juridische of personele aspecten)
 Groot imago risico door niet bereid te zijn om onze data te delen met de onderzoekswereld. Van de andere kant moet geborgd worden dat we voldoen aan de AVG bij het delen van onze data. Inrichten van een FAIR data point zou dit moeten waarborgen.

TOELICHTING FINANCIËLE CONSEQUENTIES

Dit past binnen de begroting van CIO office.

TOELICHTING RIVM ORGANISATIE SYSTEEM CONSEQUENTIES

COMMUNICATIE naar de betrokken doelgroepen
 Volgt na besluitvorming MT I&V.

VERDERE AFHANDELING

Beslispunt en advies

MT I&V wordt aan te geven of er draagvlak is:

1. voor het FAIR maken van data over Covid-19
2. om dat in de vorm van een FAIR data point te doen
3. of ondersteuning door de IV-organisatie gewenst is zoals beschreven in het voorstel

BIJLAGE

Voorstel: RIVM richt een FAIR data point in voor COVID-19 data.

Data is valuable, so let's be mindful of how we're sharing it.
Clara Shih (en.wikipedia.org/wiki/Clara_Shih)

Het belang van transparantie en open science voor RIVM is evident ((10)(2e) , 7 mei 2020) . Tegelijkertijd is de data die wij bewerken vaak gevoelig en kan niet altijd als open data gedeeld worden. Meteen na de rol die RIVM in de COVID-19 pandemie kreeg , rees de vraag naar de gegevens die de adviezen onderbouwen.

Deels kan deze data als open data gepubliceerd worden, zie bijvoorbeeld het overzicht van sterfte en ziekenhuisopnames op gemeenteniveau:
<http://indata.rivm.nl/geonetwork/srv/eng/catalog.search#/metadata/1c0fcd57-1102-4620-9cfa-441e93ea5604>. Spoedig zal meer open data over de epidemie op deze manier gedeeld gaan worden.

Meestal is voor wetenschappelijk onderzoek echter behoefte aan toegang tot meer gedetailleerde data, bijvoorbeeld op individueel niveau. Nationaal en internationaal zijn er verschillende initiatieven gestart om data over COVID-19 beschikbaar en herbruikbaar te maken voor wetenschappelijk onderzoek. Naast toegankelijk (accessible) moet de data ook vindbaar (findable) en (machine) bewerkbaar (interoperable) zijn. Deze eisen worden aangeduid met FAIR. FAIR is het acroniem voor Findable, Accessible, Interoperable en Reusable.

De financier van gezondheidsonderzoek in Nederland, ZonMw, ondersteunt verschillende initiatieven voor FAIR datamanagement. In reactie op COVID-19 is met steun van ZonMw het **Virus Outbreak Data Network** (VODAN) opgezet. Door ZonMw wordt dit initiatief gezien als het centrale punt voor het FAIR beschikbaar stellen van data over COVID-19 en het maken van wereldwijde afspraken over de datastandaarden.

Wat is VODAN?

Het doel van VODAN is dat alle data over een virusuitbraak vindbaar en bruikbaar voor verschillende partijen te maken, waardoor kwalitatief hoog onderzoek in kortere tijd uitgevoerd kan worden. VODAN is een gezamenlijk implementatie netwerk van 4 internationale organisaties: CODATA, RDA, WDS, en GO FAIR (<https://www.go-fair.org/implementation-networks/overview/vodan/>). De data en implementatie moet voldoen aan de FAIR principes. Bijvoorbeeld door het inrichten van een FAIR Data Point.

Wat is een FAIR Data Point (FDP)

Een FDP is een lokale repository waar de data is opgeslagen (de data blijven dus binnen de organisatie). Dit wordt ook wel het zogenaamde station genoemd.

Onderzoekspartijen met rechten – dit kan door de data-eigenaar bepaald worden – daarop kunnen een specifieke vraag aan dit station (een algoritme ‘langs’ sturen). Omgekeerd kan RIVM ook aan andere stations een vraag stellen en door middel van een zogenaamde data trein de antwoorden van verschillende stations bijeenbrengen. Omdat de data bij de bron blijven, niet zichtbaar worden voor de onderzoekers en alleen gebruikt worden door het leer-algoritme, is onderzoek mogelijk zonder AVG en privacy wetten te overtreden.

Een Fair Data Point is een mogelijkheid om data met ‘restricted access’ (d.w.z. niet open, wel toegankelijk onder voorwaarden) te delen. Omdat deze wens op het RIVM speelt (breder dan alleen voor COVID-19) willen we een project starten dat de infrastructuur voor een FAIR Data Point regelt.

Een groot aantal Nederlandse universiteiten en organisaties (bijvoorbeeld ZIN) participeren al in VODAN.

Naast VODAN zijn in korte tijd een groot aantal COVID werkgroepen , netwerken en initiatieven gestart, zie voor overzicht bij Health RI : <https://www.health-ri.nl/covid-working-groups-networks-initiatives>

Health-RI is een Nederlands publiek-privaat initiatief van organisaties gericht op public health- en zorg- onderzoek. Het wordt ondersteunt door meer dan 70 NL organisaties. Uit het Health RI overzicht zoeken we het best passende initiatief.

Waarom deelname RIVM?

Algemeen belang : Toegang tot meer en gevarieerdere data + onderzoek brengt sneller bevindingen die te vertalen zijn tot beleid.

Reputatie: De samenleving wenst transparantie. Transparantie en kennisdelen horen bij een kennisinstituut van de overheid. Door de centrale rol die het RIVM speelt in de coronacrisis is er grote druk vanuit de samenleving om de data over COVID-19 te delen.

RIVM bewaakt middels het maken van een FAIR Data Point het veilig beschikbaar stellen van data en uitkomsten.

Mogelijkheden Naast het delen van data aan anderen biedt dit initiatief ook de mogelijkheid om zelf gebruik te maken van de data van andere instituten. Zichtbaarheid bevordert nieuwe samenwerkingen.

Ervaring

De ervaring die met VODAN en Fair data points opgedaan wordt is voor al het data delen op het RIVM waardevol, niet uitsluitend met betrekking tot de huidige virus uitbraak.

Daarnaast zal het RIVM zelf ervaring op kunnen doen met deze nieuwe manier van onderzoek doen (bekend onder de termen personal health train, federated learning, privacy preserving analytics). De investering zal zich uitbetalen.

Steeds vaker een vereiste

Subsidieverstrekkers zoals ZonMw, NWO en Horizon2020 hebben inmiddels als eis voor het verstrekken van subsidie opgenomen dat de onderzoeksdata na afloop van het onderzoek als FAIR data voor hergebruik ter beschikking gesteld moeten worden.

Voorstel

CIO office en SIM bestuderen in het Health RI overzicht welk initiatief voor het RIVM de meest geëigende weg is om COVID-19 onderzoeksdata te delen. Belangrijke vragen daarbij zijn de procedure volgens welke bepaald wordt of iemand toegang krijgt tot de data in het FAIR data point en hoe het RIVM haar zeggenschap daarin kan borgen. Daarnaast hoe wederkerigheid is geborgd. Op basis hiervan komen CIO office en SIM met een voorstel hoe RIVM 'restricted access' COVID-19 data het beste volgens FAIR principes ter beschikking kan stellen.

Op basis van goedkeuring van dit voorstel, zullen CIO office, SIM en SSC Campus een FAIR data point inrichten bij het RIVM, dat voldoet aan onze standaarden. De data blijven op het RIVM, het RIVM stelt de data wel voor hergebruik ter beschikking met geregelde toestemming onder voorwaarden. En het RIVM kan zelf ervaring opdoen met analyses conform personal health train principe/federated learning.

Wie doet wat en kosten

(10)(2e), recent gestart als datasteward bij CIO office, zal met collega's bij CIO office en SIM met een voorstel komen voor infrastructuur van een Fair Data Point. Bij goedkeuring zal (10)(2e) de verdere uitwerking ter hand nemen. Hierbij zullen de organisatorische en infrastructurele kant van het FDP door CIO office, SIM en SSC Campus geregeld worden.

Onderzoekers en centra die 'restricted access' data willen delen (bijvoorbeeld CIB en V&Z), zorgen voor een metadata beschrijving en zijn betrokken bij het opstellen licenties en toegangsrechten. Metadata moet voldoen aan FAIR richtlijnen. Het streven is de onderzoekers zo veel mogelijk te ontlasten, met simpele en logische procedures, eventueel met ondersteuning van de datasteward.

De kosten worden betaald uit het budget dat CIO office voor 2020 heeft gereserveerd voor pilots/ondersteuning op het gebied van het FAIR maken van data.

Bijlage bij het voorstel2

(10)(2e) (10)(2e) bericht 7 mei 2020: "Bij burgers en collega wetenschappers is er een grote behoefte aan transparantie over ons werk en inzage in onze gegevens, modellen en onderzoeken. Dit zagen we rond de stikstofdiscussie vorig jaar en zien we nu opnieuw als het gaat om onze onderzoeksgegevens en adviezen rondom de coronacrisis. Velen van jullie hebben vast de [uitzending van Nieuwsuur](#) en/of de [open brief](#) hierover in NRC gezien. Wij onderschrijven vanzelfsprekend het belang van transparantie en open science. Maar de dynamiek van crisisbeheersing is wel anders dan die van de wetenschappelijke publicatie. Dat blijven we waar nodig consequent toelichten en uitleggen, steeds strevend naar zo groot mogelijke transparantie."

Afkortingen:

FAIR = Findable, Accessible, Interoperable and Reusable

CODATA= the Committee on Data of the International Science Council (codata.org)

RDA= Research Data Alliance (rd-alliance.org)

WDS=World Data System (world-datasytem.org)

Go FAIR = Global Open FAIR (go-fair.org)