

## Oproep en advies

# Betere overheidsdata voor sterk beleidsonderzoek en goed bestuur

Brussel, 13 september 2021

# Oproep

## Betere verzameling en deling van overheidsdata besparen tijd, geld en levens

De Sociaal-Economische Raad van Vlaanderen (SERV) en de Vlaamse Interuniversitaire Raad (VLIR) roepen de Vlaamse Regering op om steviger te investeren in de beschikbaarheid, kwaliteit, vindbaarheid en ontsluiting van overheidsdata.

Rik Van de Walle, voorzitter VLIR: *“Het probleem is al langer gekend en de coronacrisis toonde opnieuw dat datastromen nog veel te versnipperd zijn in ons land. Data die in het bezit zijn van de overheid moeten sneller worden vrijgegeven voor academisch onderzoek zodat de overheid vervolgens met kwaliteitsvolle data en wetenschappelijke inzichten beter gerichte en meer proportionele beslissingen kan nemen.”*

Ann Vermorgen, voorzitter SERV: *“Data en hergebruik van overheidsdata spelen ook een cruciale rol in de digitale transitie. We waren destijds met de Kruispuntbank van de Sociale Zekerheid voorloper bij het koppelen en ontsluiten van overheidsdata maar lopen nu steeds meer achter op andere landen. Structureel overleg met onderzoekers en het middenveld moet toelaten het terug beter te doen. De SERV en de VLIR zijn bereid om hier hun schouders onder te zetten.”*

### Vier dringende werven

Ondanks herhaalde ambities van de Vlaamse Regering om werk te maken van een datagedreven overheid – recent nog in het Relanceplan ‘Vlaamse Veerkracht’ – ligt er nog veel werk op de plank om ze in praktijk te brengen. Volgens de Vlaamse Universiteiten en de Vlaamse werkgevers- en werknemersorganisaties zijn er vier dringende werven:

1. De kwaliteit en beschikbaarheid van de data. Die zijn niet altijd toereikend voor onderzoekers of beleid. Er zit vaak te veel vertraging op de data, soms zelfs meerdere jaren en er zijn belangrijke datalacunes. Dit kan leiden tot minder gepaste beleidsbeslissingen en het noodgedwongen bijstellen van onderzoeksambities.
2. De vindbaarheid en ontsluiting van data. Het is te weinig duidelijk welke data beschikbaar zijn. Het overzicht ontbreekt. Relevante datareeksen zitten verspreid over verschillende entiteiten en bestuursniveaus. Dat zorgt voor heel wat puzzelwerk en tijdverlies.
3. De gebruiksvriendelijkheid van de privacyprocedures en de ethische omgang met data. Het kost veel tijd en moeite om procedures voor data-aanvragen met persoonsgegevens te doorlopen. De privacyregels worden bovendien vaak erg strikt toegepast. Ze lijken soms een alibi om data niet te delen.
4. De cultuur en organisatie. Overheidsentiteiten zijn nog te veel gericht op wat ze zelf met hun eigen data kunnen doen. Ze zien te weinig wat er allemaal mogelijk is met de data en onderschatten het maatschappelijke potentieel van brede datadeling. Het delen van data wordt dan gezien als een gunst, niet als een publiek goed dat het beleid en de samenleving ten goede komt.

## Concrete aanbevelingen

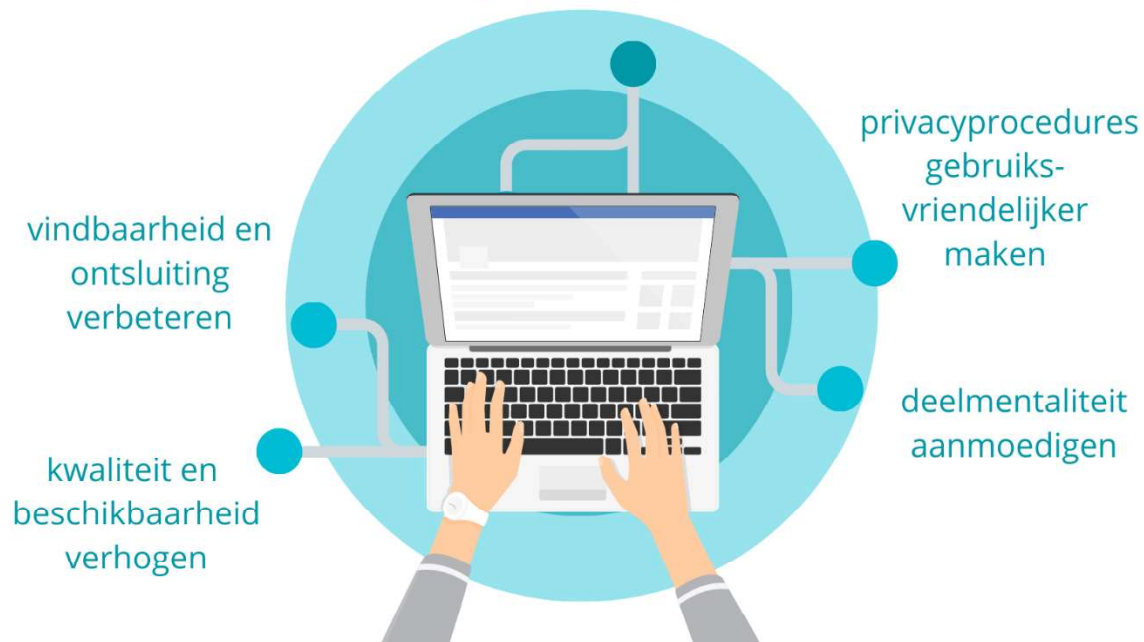
Buitenlandse voorbeelden tonen dat het beter kan. Enkele aanbevelingen van de SERV en de VLIR zijn:

- Organiseer structureel overleg en informatie-uitwisseling met onderzoekers en andere gebruikers om datanoden te signaleren, dataleemtes te identificeren en datafouten recht te zetten. Dat gebeurt bij voorkeur met alle bestuursniveaus samen.
- Maak werk van een versnelde uitbouw van een Vlaamse datavindplaats om de vindbaarheid van Vlaamse (overheids-) data te verhogen en ook interbestuurlijke data meer zichtbaar te maken. Ontwikkel een data-infrastructuur met gevalideerde data, gekoppelde databanken en gestandaardiseerde producten (vaste datasets) waarop gebruikers vlot beroep kunnen doen. Dat voorkomt dat voor de meest voorkomende vragen en behoeften iedere keer opnieuw data-aanvragen moeten worden ingediend en behandeld. De schaarse capaciteit kan dan worden ingezet voor complexere aanvragen.
- Zorg voor meer duidelijkheid en uniformiteit in de toepassing van de privacyregelgeving, vervang het amalgaam aan protocollen, machtigingen en beraadslagingen zoveel mogelijk door eenvormige afsprakenkaders en aanvraagprocedures en interbestuurlijk afgestemde en vereenvoudigde werkwijzen en formulieren. Dat verhoogt de efficiëntie, de doorlooptijd van de aanvraag en de klantgerichtheid.
- Versterk het hergebruik van data bij overheden en stimuleer een cultuur van data-uitwisseling, bijvoorbeeld door de mate waarin entiteiten het 'once only'-principe toepassen en data delen in te schrijven in de te bereiken maturiteit qua organisatiebeheersing. Werk een datastrategie uit in elk beleidsdomein naast een gedeelde Vlaamse datastrategie, met Chief Data Officers / Chief Policy Scientists die fungeren als aanspreekpunt voor beleidsmedewerkers en onderzoekers en zo zorgen voor continuïteit en langetermijnopbouw van data en expertise. Investeer in datageletterdheid om de kennis over data te bevorderen in alle lagen van de Vlaamse overheid (topmanagement, middenkader, operationeel en technisch personeel).

## Zorgen voor systematische verbetering

Om een dynamiek gericht op systematische verbetering op gang te brengen, is er structureel overleg nodig tussen overheid, onderzoekers en gebruikers. Een 'trefdag' zou jaarlijks in een bredere kring de resultaten van dat overleg kenbaar moeten maken, de vorderingen toelichten en de noden verder inventariseren. De SERV en de VLIR zijn bereid om hun schouders te zetten onder zo'n structurele dialoog.

structureel overleg tussen overheid, onderzoekers en  
andere gebruikers opzetten



Betere verzameling en deling van overheidsdata nodig

# Advies

## Investeer in betere overheidsdata voor sterk beleidsonderzoek en goed bestuur

### Belang van een datagedreven overheid

Beleidskeuzes onderbouwen met kwaliteitsvolle data en wetenschappelijke inzichten is een basiskenmerk van **goed bestuur**. Het versterkt de kwaliteit van de besluitvorming en de publieke dienstverlening en draagt bij aan publieke waarden zoals efficiëntie, effectiviteit, transparantie en verantwoording. Het leidt tot proportionele beleidsbeslissingen die geld, tijd en mensenlevens sparen.

Data spelen bovendien een cruciale rol in de **digitale transitie**. De Europese digitale agenda, waarin ook Vlaanderen zich heeft ingeschreven, beoogt onder meer een versterking van het grootschalig hergebruik van overheidsdata en met overheidsmiddelen gefinancierde onderzoeksgegevens. De data die de overheid verzamelt en produceert vormen immers een enorme, diverse en waardevolle hoeveelheid data die de samenleving ten goede kunnen komen. Ze vormen een belangrijke grondstof voor nieuwe toepassingen en digitale informatieproducten en -diensten. Dat moet bijdragen aan economische groei en de concurrentiepositie van de Europese Unie verbeteren.

**COVID-19** heeft nogmaals het sociaal-economische en maatschappelijke belang van data en wetenschappelijke informatie onderstreept, niet alleen als fundament voor alle fasen van de beleidscyclus (voorbereiding, beleidskeuzes, uitvoering, evaluatie), maar zeker ook als hoeksteen voor een robuust systeem van risicomangement op lange termijn. De crisis legde bovendien al langer bestaande pijnpunten bloot in de databeschikbaarheid, de datatoegankelijkheid en het databeheer. Ondanks herhaalde ambities moet de performantie van het datasysteem in Vlaanderen beter.

De SERV en de VLIR waarderen dat de Vlaamse Regering daartoe **diverse initiatieven** heeft genomen in het kader van de uitvoering van het Witboek Bestuur, de totstandkoming van Digitaal Vlaanderen en de omzetting van de PSI-Richtlijn. Het Relanceplan 'Vlaamse Veerkracht' wil een bijkomende impuls geven. Het is bovendien positief dat er over alle beleidsdomeinen heen gewerkt wordt aan een datastrategie in reactie op eerdere vragen vanuit de SERV, de VLIR en de relancecomités. Toch ligt er nog veel werk op de plank om de ambities concreet in te vullen en geïntegreerd in de praktijk te brengen.

Met hun **advies** willen de Vlaamse Universiteiten en de Vlaamse werkgevers- en werknemersorganisaties de noodzaak van een **ambitieuw databeleid** onderstrepen en aanbevelingen formuleren voor het wegwerken van belangrijke knelpunten. Die knelpunten werden geïdentificeerd door analyse en gesprekken met experts en getoetst in twee relevante 'cases': gezondheidsbeleid en arbeidsmarktbeleid.

De **aanbevelingen** zijn gericht op het verbeteren van

- (1) de beschikbaarheid van kwaliteitsvolle data,
- (2) de vindbaarheid en ontsluiting van data,
- (3) de gebruiksvriendelijkheid van de privacyprocedures en ethische omgang met data
- (4) de benodigde cultuur en organisatie.

## Beschikbaarheid van kwaliteitsvolle data

### Analyse

De voor onderzoek en derde partijen beschikbare overheidsdata zijn lang niet altijd even *kwaliteitsvol*. Data voldoen niet altijd aan de FAIR-principes (Findable, Accessible, Interoperable, Reusable). Dat ondermijnt de mogelijkheid tot gerichte data-uitwisseling en geautomatiseerde verwerking. Belangrijke beleids- en onderzoeksvragen worden niet voldoende aangevat doordat goede koppeling tussen data niet altijd mogelijk is.

De resultaten van data-analyses worden niet altijd gepubliceerd waardoor externe kwaliteitscontrole uitgesloten is. Er wordt te weinig geïnvesteerd in longitudinale datareeksen. De adoptiegraad van bestaande *datastandaarden* ligt te laag.

De data bevatten nog te veel *fouten*. Er zijn weinig incentives voor dataleveranciers om voor kwaliteitsvolle data te zorgen en de data van metadata en documentatie te voorzien. De datakwaliteit lijdt ook onder een gebrek aan afstemming tussen beleidsmedewerkers (business) en technische medewerkers (IT).

Er zijn ook nog belangrijke *datalacunes* en op veel data zit (te veel) vertraging. De beschikbaarheid van real-time data uit beleids- en dienstverleningsprocessen is beperkt.

Data moeten niet alleen kwaliteitsvol zijn, maar ook *fit for purpose*, wat afhangt van de doeleinden waarvoor de data worden gebruikt.

De beschikbaarheid van data wordt soms beperkt door vereisten inzake *confidentialiteit*, waarbij de data wel beschikbaar zijn voor academisch onderzoek, maar minder voor beleidsondersteunend werk (bv. RSZ-data, NBB-data op microniveau). Toegang tot deze data gebeurt nu meestal case-by-case.

### Aanbevelingen

De SERV en de VLIR vragen structurele, bij voorkeur interbestuurlijke, *mechanismen* om datanoden te signaleren, dataleemtes te identificeren en datafouten te remediëren.

Duidelijke *verantwoordelijkheden* en werkzame *incentives* zijn nodig om data te documenteren (metadata) en kwaliteitsvol te houden (datamanagement).

De kwaliteit van data moet beter. Dit kan door het systematischer gebruik van *datastandaarden*, gevalideerde methodologieën en door *terugkoppeling* door gebruikers over kwaliteit van de aangeleverde data. Dit vraagt initiatief van dataleveranciers om met gebruikers in dialoog te gaan. Door deze informatie te delen met andere onderzoekers kan ook toekomstig onderzoek vlotter verlopen.

Een *strategie* rond datakwaliteitsmanagement is nodig. Het protocol van Statistiek Vlaanderen over datakwaliteit (van statistische data) biedt daarvoor een basis.

Het voorzien van een algemene *toegang tot confidentiële overheidsdata* voor expertcentra in beleidsondersteuning zou veel voordelen bieden en vermijden dat er op een case-by-case basis moet worden gewerkt.

### Voorbeelden

#### Gezondheid:

- Risicomanagement in het algemeen en pandemiebeheer in het bijzonder vergen longitudinale reeksen van epidemiologische data. De structuren hiervoor moeten klaarstaan aan het begin van een pandemie maar zijn ook in niet-pandemische tijden essentieel.
- De overheid laat na, zeker in crisissituaties, om gebruik te maken van nochtans beschikbare wetenschappelijk gevalideerde methodologieën en instrumenten. Denk in de coronacrisis maar aan de veronachtzaamde basisprincipes uit het epidemiologisch onderzoek, zoals de epidemische boom voor fijnmazige analyses, het brononderzoek of het casemanagementprincipe. Hierdoor moet de overheid zich voor het ontwikkelen van beleid noodgedwongen beroepen op buitenlandse kennis en/of minder fijnmazige data. In de praktijk kan dit aanleiding geven tot disproportionele beleidsbeslissingen.
- De COVID-19-crisis heeft nogmaals het belang van welzijn en de sterke relatie tussen welzijn en gezondheid aangetoond. Welzijnsdata en ook armoededata zouden ook meer beschikbaar moeten zijn. Binnen het Steunpunt Welzijn Volksgezondheid en Gezin zijn er bijvoorbeeld studies die analyses bevatten van data die beschikbaar zouden moeten zijn, maar niet systematisch worden verzameld of niet zijn ontsloten. Betere dataverzameling en doorstroming van data zijn niet alleen belangrijk voor wetenschappelijk onderzoek maar ook voor betere zorg en ondersteuning en uiteraard voor een evidence-informed beleid.

#### Arbeidsmarkt:

- Er zijn nog belangrijke datalacunes, onder andere voor werk-naar-werk transitie, tijdelijke werkloosheid, personen met een migratieachtergrond, arbeidshandicap, tijdelijke werkloosheid, koppelingen tussen arbeidsmarktdata en onderwijsdata, enz. Administratieve data zijn niet altijd op beroepenniveau (ISCO) beschikbaar, vaak enkel data op sectorniveau (NACE). Dit bemoeilijkt arbeidsmarktprognosemodellen, het maken van een beroepenmonitor of discussies over zware beroepen.
- De Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM&SB) is één van de belangrijkste databronnen voor arbeidsmarktonderzoek. De data zijn echter niet actueel en slechts beschikbaar met vertraging (tot drie jaar). Deze termijn is te lang voor beleidsmakers maar ook voor beleidsgericht onderzoek dat actueel wil zijn. De DWH AM&SB speelt ook te traag in op nieuwe datanoden en volgt een model van eenmalig onderzoek. Hierdoor moeten instanties datastromen bij de dataleveranciers via de Kruispuntbank van de Sociale Zekerheid (KSZ) aanvragen die reeds opgezet zijn voor de DWH AM&SB.
- De kwaliteit van de data (juistheid, fijnmazigheid, actualiteit, enz.) is niet altijd toereikend voor onderzoekers of beleidstoepassingen en vaak niet ontsloten in een formaat dat voldoet aan de behoeften. Daardoor moeten onderzoekers maar soms evengoed administraties op surveydata terugvallen (bv. de koppeling van adviezen van arbeidsbemiddelaars aan klanten in (verschillende soorten) activeringstrajecten bij de Vlaamse Dienst voor Arbeidsbemiddeling (VDAB) om het hele parcours van een klant in kaart te brengen). Ook de survey-infrastructuur is te weinig uitgebouwd en gedeeld. De VLIR heeft het initiatief genomen voor het opzetten van een internetpanel, naar analogie met gelijkaardige paneldata in Duitsland, Nederland en het Verenigd Koninkrijk.
- Er bestaan regionale verschillen tussen datasets van Vlaanderen, Brussel en Wallonië. Dat belemmert onderzoekers om in te tekenen op Europese of internationale onderzoeksprojecten die vergelijkingen op land-niveau voorop stellen maar bovendien en vooral om analyses en evaluaties van het beleid te maken.

## Vindbaarheid en ontsluiting van data

### Analyse

Er zijn heel wat overheidsdata. De *transparantie* is echter beperkt en het overzicht ontbreekt.

Relevante datareeksen zijn *verspreid* over verschillende entiteiten en bestuursniveaus. Ze zitten ook bij niet-overheidsactoren zoals mutualiteiten, sociale secretariaten en externe diensten voor preventie en welzijn op het werk. Onderzoekers en andere gebruikers worden bovendien geconfronteerd met de silowerking van overheden. Dat zorgt voor

### Aanbevelingen

De VLIR en de SERV vragen dat er versneld werk wordt gemaakt van een Vlaamse *datavindplaats*. Die moet de vindbaarheid van Vlaamse (overheids)data verhogen en ook interbestuurlijke data meer zichtbaar maken. Belangrijk is dat dit dataportaal een hoge toegevoegde waarde heeft en de belangrijkste *high value datasets* (eerst) bevat.

Het verdient aanbeveling om een dergelijke datavindplaats niet als een repository te concipiëren maar als een metadatabank met

heel wat puzzelwerk om data te zoeken, op te vragen en samen te leggen.

*Interbestuurlijke uitwisseling* van data stopt niet aan de regio- of landsgrenzen. De internationale dimensie komt te weinig aan bod bij het uittekenen en verfijnen van toekomstig overheidsbeleid, zeker bij het beheren van internationale crisissituaties.

Het aantal *open data sets* neemt weliswaar toe maar er is een grote verscheidenheid op het vlak van thema's, kwaliteit, entiteiten en toegankelijkheid. De kwaliteit en fijnmazigheid van open data is vandaag vaak te laag om ze ook doeltreffend systematisch in te zetten. Het huidige aanbod is nog beperkt en kan doelgerichter.

verwijzingen naar beschikbare datasets, mits deze voldoen aan de afgesproken kwaliteitsstandaarden. Een dergelijke aanpak zorgt er ook voor dat andere actoren, zoals de universiteiten, als *data custodians* kunnen optreden.

Daarnaast zijn *thematische dataportalen* nodig, zoals het COVID-19 Data Portal van de Europese Commissie, om in het licht van doortastend crisisbeheer data zo snel mogelijk ter beschikking te stellen van de wetenschap. Er is ook nood aan een gezamenlijke infrastructuur waar onderzoekers hun data kunnen archiveren met het oog op FAIR gebruik.

De SERV en de VLIR vragen daarnaast dat verder wordt gegaan dan ruwe data ter beschikking stellen. Naar buitenlands voorbeeld (INSEE in Frankrijk, Statistics Sweden, cf. ook ESFRI - European Strategy Forum on Research Infrastructures) mag worden verwacht dat de overheid ook zelf een *data-infrastructuur* uitbouwt met gevalideerde data, datakoppelingen en gestandaardiseerde producten (vaste datasets) waarop gebruikers vlot beroep kunnen doen. Dit zorgt voor efficiëntiewinsten: door via (gekoppelde) databanken en gestandaardiseerde (continue) producten te voldoen aan de meest voorkomende vragen en behoeften moeten niet iedere keer opnieuw (eenvoudige) data-aanvragen worden ingediend en behandeld. De schaarse capaciteit kan zo worden ingezet voor complexere aanvragen met een specifiek doel.

Tot slot moet de rol van technologie in het versterken van het gebruik van data worden beklemtoond (AI, linked data, enz.). In dat verband is interoperabiliteit erg belangrijk (cf. het European Interoperability Framework).



### Voorbeelden

#### Gezondheid

- Het internationaal delen van gezondheidsdata tijdens de coronacrisis zorgde voor heel wat problemen omdat moest worden samengewerkt tussen Europa en andere landen waar de Algemene Verordening Gegevensbescherming (AVG) niet van toepassing is. De aanbevelingen van het rapport International Sharing of Personal Health Data for Research (easac.eu) bieden handvaten voor remediëring.

#### Arbeidsmarkt

- Ervaringen van onderzoekers en beleidsmedewerkers geven aan dat dataleveranciers geen incentives hebben om data goed te documenteren en te communiceren. Zelfs kleine wijzigingen in methodologie of de betekenis van variabelen maakt longitudinaal onderzoek problematisch.
- Data zijn nog te versnipperd. Een voorbeeld is de doorstroom van jongeren uit beroepsonderwijs (of duaal leren) naar de arbeidsmarkt. Hierbij moeten onderzoekers zelf data van VDAB, het Departement Werk en Sociale Economie (DWSE) en het Agentschap voor Onderwijsdiensten (AGODI) met elkaar kruisen. Onderzoekers moeten deze oefening voor ieder onderzoek herhalen (of wachten tot de dataset door overheidsinstanties is opgebouwd en daarna gepseudonimiseerd aan onderzoekers wordt meegegeed). Dit vergt veel tijdsinvestering in een context waarin onderzoeksprojecten beperkt in de tijd zijn.
- Samen met de duurtijd van een machtigingsaanvraag kan het tot een jaar duren om een datakoppeling goed op te zetten. Een voorbeeld is de koppeling van Not in Education, Employment or Training (NEET)-data aan gegevens van de Rijksdienst voor Sociale Zekerheid (RIZIV) en data van de uitbetalers van het Groeipakket (voormalige kinderbijslagfondsen).

## Gebruiksvriendelijkheid van privacyprocedures en de ethische omgang met data

### Analyse

Regels voor *goed databeheer* vinden niet voldoende ingang. Goed databeheer vereist duidelijke spelregels over het eigenaarschap, de levenscyclus, de bescherming en de kwaliteit van data. Bovendien is transparantie noodzakelijk om steeds helder te kunnen duiden waarom bepaalde data worden verzameld.

Het kost vandaag vaak veel inspanningen en tijd (meerdere maanden tot wel een jaar) om *procedures voor data-aanvragen met persoonsgegevens* te doorlopen. De privacyprocedures komen niet altijd tegemoet aan de datanoden, timing en flexibiliteit die het beleid of een onderzoeksdesign vereisen. Dat geldt in het bijzonder voor longitudinale studies, kortlopend onderzoek of onderzoek waar lichte wijzigingen in de data nodig blijken. Problemen die onderzoekers ondervinden bij data-aanvragen of datakoppelingen worden doorgaans geval per geval bekeken, vaak zonder dat veel overleg mogelijk is of aan structurele oplossingen kan worden gewerkt. Buitenlandse voorbeelden zoals het Centraal Bureau voor de Statistiek (CBS) in Nederland

### Aanbevelingen

Overheidsdata zijn een globaal *publiek goed*. Dit zou de hoeksteen van de ethische omgang met data moeten zijn.

De SERV en de VLIR menen dat een adequate privacyregelgeving en bovenal een cultuur van informatieveiligheid en gegevensbescherming nodig is. Maar ze mogen *geen onnodige belemmering* zijn voor het koppelen en gebruiken van data voor onderzoeksdoeleinden. De evaluatie van de Belgische privacywetgeving op federaal niveau en de ervaring met de implementatie van de AVG elders bieden een opportuniteit om het systeem van de controle op de uitwisseling en verwerking van persoonsgegevens kritisch tegen het licht te houden.

Meer *duidelijkheid en uniformiteit* in de toepassing ervan kan onderzoekers en ondersteunende juridische diensten ontlasten, capaciteit vrijmaken en snellere procedures mogelijk maken. Dat vergt ook interbestuurlijke afspraken over een gefedereerd systeem voor het indienen van

en L'Institut National de la Statistique et des Études Économiques (INSEE) in Frankrijk tonen dat het beter kan.

De coronacrisis heeft bovendien al te vaak een zeer *enge visie op de privacyregels* geëtaleerd. Die lijken soms een alibi voor overheidsorganisaties om terughoudend te zijn in het delen van hun data, wetgeving restrictief te interpreteren of datavragen op de lange baan te schuiven. De transparantie is beperkt: het is niet altijd duidelijk wat de redenen zijn voor het verwerpen of aanvaarden van zeer gelijkaardige aanvragen. De verplichting om data te vernietigen staat haaks op de repliceerbaarheid van onderzoek en strookt niet met de Algemene Verordening Gegevensbescherming (AVG).

machtigingsaanvragen, meer frequente beraadslagingen, evenals op elkaar afgestemde en vereenvoudigde procedures en formulieren.

De individuele machtigingen, beraadslagingen en huidige protocollen worden best zoveel mogelijk vervangen door eenvormige *afsprakenkaders* die het mogelijk maken om sneller toegang te krijgen tot data en die verzekeren dat de doelstellingen waarvoor de data mogen worden gebruikt niet al te zeer worden beperkt. Een eenvoudige toegang, uiteraard met respect voor het juridische en ethische kader, versterkt de gebruiksvriendelijkheid.

Meer algemeen zou de *hele keten* van het bekomen van data voor onderzoek geëvalueerd moeten worden (voorbereiding, privacyprocedures, dataverzameling, opstellen en cleanen van data, ontsluiting richting onderzoekers).

Inspraak van en samenspraak met onderzoekers is een must voor een *gebruikersgerichte en vertrouwensvolle dienstverlening*. Kennisinstellingen stellen best een gedragscode 'responsible data management' op, dat als zelfregulerend instrument kan worden ingezet voor sensibilisering en het stroomlijnen van de onderzoekspraktijk.

Naar de toekomst toe moet resoluut de kaart worden getrokken van '*open science*', waarbij de nodige randvoorwaarden (o.a. voor anonimisering, pseudonimisering, informatieveiligheid, duidelijke rolverdeling) worden ingebouwd om misbruik te vermijden.

Een bijzonder aandachtspunt is de ethische omgang met data en in het bijzonder zorgen voor *e-inclusie* (cf. 'Berlin Declaration on Digital Society and Value-based Digital Government').

### Voorbeelden

#### Gezondheid:

- De privacyregels werden tijdens de coronacrisis ten onrechte gebruikt als argument om overheidsdata lange tijd niet vrij te geven (bv. cijfers over de besmettingen op de werkvloer).
- Tijdens de coronacrisis werden herhaaldelijk niet-versleutelde persoonsgegevens via onbeveiligde IT-systemen getransfereerd. Voor het gebruik en de valorisatie van de data die door Sciensano worden aangeleverd aan onderzoeksinstituten ontbreekt momenteel een afsprakenkader en een gedragscode. De recentelijk afgesloten samenwerkingsovereenkomst tussen Sciensano en de Belgische universiteiten (VLIR, CRef) is een te volgen model.

#### Arbeidsmarkt:

- Aanvragen moeten zeer gedetailleerd zijn (incl. op niveau van de variabelen). Privacy-autoriteiten hanteren een zeer strikte interpretatie van de AVG-regels. Als uit de analyse blijkt dat beperkte aanpassingen toch nog nodig zijn, dan is men genoodzaakt om de procedures opnieuw te doorlopen. Een andere vertragende factor is de beperkte capaciteit om aanvragen te beoordelen. Verder moeten bij iedere administratie inhoudelijk verschillende procedures worden doorlopen en formulieren ingevuld. Een voorbeeld zijn aanvragen ten aanzien van STATBEL of de DWH AM&SB. Bij de DWH AM&SB controleert een persoon van de KSZ bijvoorbeeld via een projectie het scherm waarop een gemachtigde data verwerkt. Remote-access, zoals in het buitenland, zou hier mogelijk moeten worden.
- Privacy-autoriteiten benaderen aanvragers eerder wantrouwig dan vertrouwensvol. Het is bovendien niet altijd duidelijk wat de redenen voor het verwerpen of aanvaarden van zeer gelijkaardige aanvragen zijn. Overheidsinstanties hebben niet altijd een gezamenlijke visie en motivering voor het ontsluiten van persoonsgegevens. Een voorbeeld bij VDAB is het conform stellen van de gegevens van mensen met een migratieachtergrond met de Vlaamse definitie. Deze houdt niet alleen rekening met de nationaliteit van het individu, maar ook met die van de ouders. De privacy-instantie was tegenstander, waarop VDAB voorstelde om enkel via de confrontatie van bestanden deze informatie te bekomen (dus ja of nee voor ieder betrokken individu). Maar de betrokken dataleveranciers vonden het geen prioriteit of zagen er geen behoefte in, waardoor VDAB deze gegevens niet kon verkrijgen.

## Cultuur en organisatie

### Analyse

Er is nog veel groeimarge qua *cultuur en motivatie* bij alle Vlaamse instanties om data met elkaar uit te wisselen. Entiteiten zijn vaak nog te veel naar binnen gericht, op de eigen data en doelstellingen. Ze hebben geen holistische opvatting over de overheid en het datapotentieel. Dat zorgt voor problemen bij onderzoekers die data van verschillende instanties willen koppelen. Er moet dan ook veel energie geïnvesteerd worden in het overtuigen om datastromen op te zetten, data te documenteren, toegang te verlenen of fouten in de data recht te zetten.

Dataleveranciers hebben vaak te weinig *inzicht in wat er met hun data kan gebeuren*. Het delen van data wordt dan gezien als een gunst die verleend wordt in plaats van iets dat het beleid en de samenleving ten goede komt.

### Aanbevelingen

De VLIR en de SERV dringen aan op een doorgedreven *mentaliteitswijziging*. De heersende (impliciete of expliciete) attitude 'data is power' moet wijzigen naar 'data sharing is even more power'. Een goede datadeling bespaart tijd, geld en levens.

Een cultuur van data-uitwisseling begint reeds bij de verzameling van data, vanuit het principe dat data slechts éénmaal wordt verzameld om daarna te worden hergebruikt. Dit vergt goede incentives. Voor administraties en andere, niet-academische instellingen, zou de praktijk en bereidheid om het once only principe toe te passen en data te delen met derden kunnen worden ingeschreven in de te bereiken maturiteit qua organisatiebeheersing en interne controle. Voor een wetenschappelijke instelling zou dit gestimuleerd kunnen worden door repositories te gebruiken voor

Er is eveneens een gebrek aan *vertrouwen en samenwerking* tussen de verschillende actoren om de data samen te brengen en waar mogelijk te koppelen in een geest van maximale complementariteit.

De huidige *versnipperde aanpak* leidt tot meerdere problemen zoals verschillende (juridische) regelingen, weinig transparantie over beschikbare data, lange procedures voor aansluitingen en integraties, te weinig kennis, capaciteit en schaalbaarheid, beperkte semantische en technische interoperabiliteit, een financiering die de uitwisseling van data hindert en een omgeving die datagedreven innovaties in de weg staat.

geanonimiseerde data. De DOI (digital object identifier) die daarbij wordt toegekend kan dan als een aparte categorie worden opgenomen in de academische bibliografie.

Tegelijk gaat een datagedreven overheid verder dan het louter delen van data. Het gaat over de hele *levenscyclus* van data en het inbouwen van een continue feedbackloop over wat men met de data kan doen, welke verdere stappen men kan zetten. Het gaat eveneens over duidelijke spelregels rond eigenaarschap, gebruiksrechten van onderzoeksdata en hun implementatie in de samenwerkingsovereenkomsten tussen overheid en onderzoekswereld.

De keuze voor continuïteit en langetermijnopbouw van data en expertise vergt een *aangepaste governance*. Gekoppeld aan de gedeelde datastrategie zou elk beleidsdomein een eigen datastrategie moeten ontwikkelen, samen met de belangrijkste stakeholders uit de maatschappelijk vijfhoek. Dit vergt een gestructureerde dialoog die zorgt voor een vraaggedreven aansturing van de data-opbouw en ontsluiting.

*Chief Data Officers / Chief Policy Scientists* in elk beleidsdomein kunnen de opmaak en uitvoering van zo'n datastrategie ondersteunen en fungeren als Single Point Of Contact (SPOC) voor beleidsmedewerkers en onderzoekers. Zij kunnen als netwerk gezamenlijke uitdagingen opnemen, goede praktijken delen en nieuwe initiatieven stimuleren. De rol en opdracht van het netwerk Statistiek Vlaanderen moet daarbij worden herijkt zodat het nog meer ondersteunend, oplossingsgericht en netwerkversterkend kan zijn.

Een performant databeleid vergt daarnaast investeringen in *datageletterdheid*. Beleidsmatige en technische kennis van data moet bij medewerkers doorheen alle geledingen van de organisatie gedeeld worden. Naar het voorbeeld van de Vlaamse

AI-Academie en de Nederlandse Rijksacademie voor Digitalisering en Informatisering Overheid zijn multidisciplinaire opleidingen op verschillende niveaus (topmanagement, middenkader, operationeel, technisch, enz.) een belangrijk instrument om de kennis en competenties die noodzakelijk zijn om een digitaal en datagedreven beleid te vergroten en aan te scherpen. Het gaat dan niet alleen over datakennis maar ook over soft skills zoals kritisch denken, het omgaan met evidence en communicatievaardigheden.

#### Voorbeelden

##### Gezondheid

- De fragmentatie van gezondheidsdata over de verschillende bestuursniveaus en de onwil tot tijdig delen van data heeft tijdens de coronacrisis geleid tot grote vertragingen bij de monitoring van de pandemie (in casu bij het in kaart brengen van de mortaliteitscijfers in woonzorgcentra en ziekenhuizen). Een bijkomend gevolg is dat ons land internationaal grote imagoschade heeft opgelopen, zowel politiek als wetenschappelijk. Een situatie die, alle communicatie-initiatieven ten spijt, niet meer kan worden gecounterd.
- De gebrekkige opleiding van de call agents in de eerste fase van het contactonderzoek tijdens de coronacrisis heeft tot aanzienlijke vertragingen geleid in de aanpak van de crisis.

##### Arbeidsmarkt

- Het arbeidsmarktbeleid heeft nog geen echte evidence-informed cultuur. Centrale elementen zijn daarin onder meer experimenteerruimtes en Randomized Controlled Trial (RCT) voor innovatief beleid, en een arbeidsmarktmonitoring die zowel binnen de beleidscyclus structureel ingezet kan worden als in het kader van wetenschappelijk onderzoek. Een evidence-informed cultuur zorgt ervoor dat vooraf bij elke maatregel wordt nagedacht welke data nodig zijn, hoe monitoring kan opgezet worden, welk beleidsgericht onderzoek nodig is en hoe er geëvalueerd moet worden.
- Het Steunpunt Werk is een belangrijke troef en heeft in haar opdracht een kruispuntbankfunctie toegewezen gekregen. Het huidige beheerskader zorgt zowel financieel als organisatorisch voor beperkingen. Een regelgevend kader dat het Steunpunt duurzaam verankert en financiert lijkt meer aangewezen.
- Betere gebruikersinspraak is nodig om samen met onderzoekers de mogelijkheden die data bieden verder te exploiteren. De bestaande gebruikersgroepen (op het niveau van bv. DWH AM&SB, de Vlaamse Statistische Autoriteit) moeten op een andere leest worden geschoeid. In het ontwerp van Strategische doelstellingen 2022-2026 van Statistiek Vlaanderen wordt dit bevestigd en worden verbeteringen aangekondigd. Statbel heeft nu ook een werkgroep rond telewerk en platformeconomie die zowel beleidsmedewerkers als onderzoekers bevat. Deze werkgroep kan als opstap dienen om ook daar stakeholders meer structureel te gaan betrekken rond data en evidence.

## Systematische verbetering door structureel overleg

De beschikbaarheid, kwaliteit, vindbaarheid, ontsluiting, procedures en governance van data voor zowel onderzoek als beleid moeten verbeteren. Om de bestaande knelpunten aan te pakken, formuleerden de VLIR en de SERV in deze oproep een aantal beleidssuggesties. Belangrijk is echter ook om een brede dynamiek op gang te brengen om te komen tot gerichte en systematische verbetering. Daarom is structureel overleg wenselijk tussen overheid, onderzoekers en gebruikers om datanoden te bespreken, knelpunten te signaleren en strategieën en prioriteiten te bepalen. Jaarlijks zou een 'trefdag' kunnen worden georganiseerd om in een bredere kring de resultaten van dat overleg kenbaar te maken, de vorderingen toe te lichten en de noden verder te inventariseren. De SERV en de VLIR zijn bereid om hun schouders te zetten onder zo'n structurele dialoog.