

**Microdata voor statistisch en wetenschappelijk onderzoek,
de mogelijkheden voor onderzoekers om individuele gegevens van Statbel te bekomen**

Erik Meersseman en Patrick Lusyne

INTRODUCTIE

Statbel, het Belgische statistiekbureau verzamelt, produceert en verspreidt betrouwbare en pertinente cijfers over de Belgische economie, de samenleving en het territorium. Om deze statistieken te kunnen opmaken, zamelt Statbel gegevens in. Hiervoor vraagt Statbel bestaande data op bij openbare of private instellingen en organiseert het enquêtes bij burgers of ondernemingen.

Het verspreiden van cijfers gebeurt op de eerste plaats via globale en anonieme statistieken. Hiermee worden resultaten bedoeld waar een cijfer steeds betrekking heeft op meerdere burgers of ondernemingen en waarbij er dus geen vertrouwelijke informatie wordt vrijgegeven. Voor sommige onderzoekers volstaan dergelijke globale en anonieme statistieken echter niet om hun onderzoek uit te kunnen uitvoeren. De wetgeving voorziet de mogelijkheid dat Statbel in bepaalde gevallen individuele gegevens (microgegevens) meedeelt, waarbij elke lijn in het bestand gegevens bevat van één persoon, gezin of onderneming.

Dit artikel beschrijft de mogelijkheden voor onderzoekers om in het kader van statistische en wetenschappelijke doeleinden individuele gegevens waarover Statbel beschikt, te bekomen.

WETTELIJK KADER

De verwerking van gegevens door Statbel en ook het delen van deze gegevens aan derden is wettelijk geregeld via de *Wet betreffende de openbare statistiek van 4 juli 1962* (hierna “Statistiekwet” genoemd). Het uitgangspunt van de Statistiekwet is dat alle gegevens die Statbel bekomt en verwerkt voor het opmaken van statistieken, beschermd zijn door het statistisch geheim.

Verder is, voor wat persoonsgegevens aangaat, ook de *Algemene verordening gegevensbescherming* (Verordening (EU) nr. 2016/679 van het Europees Parlement en de Raad van 27 april 2016 – Hierna “AVG”) en de *wet van 30 juli 2018 betreffende de bescherming van natuurlijke personen met betrekking tot de verwerking van persoonsgegevens* van toepassing (hierna “AVG-kaderwet” genoemd).

Elke mededeling van individuele gegevens zal gebeuren binnen wat toegelaten is volgens de hierboven vermelde wetgeving. Merk op dat de Statistiekwet ook gegevens beschermt die betrekking hebben op ondernemingen of personen die overleden zijn, wat buiten het toepassingsgebied valt van de AVG.

VOOR WELKE DOELEINDEN KUNNEN MICROGEGEVENS AANGEVRAAGD WORDEN?

Onderzoekers kunnen microgegevens aanvragen als de data noodzakelijk zijn voor hun statistisch of wetenschappelijk onderzoek.

Met *noodzakelijk* wordt bedoeld dat het onderzoek niet mogelijk is op basis van globale en anonieme gegevens die Statbel zonder beperkingen verspreidt via de website www.statbel.fgov.be:

- **Be.STAT¹**, de elektronische databank van Statbel, laat onderzoekers toe om tabellen aan te maken op maat;
- Via **Open data²** stelt Statbel geaggregeerde gegevens ter beschikking op een vrij gedetailleerd niveau.

Het doel waarvoor de gegevens opgevraagd worden, beperkt zich tot statistisch en wetenschappelijk onderzoek. Dit doel mag echter ruim geïnterpreteerd worden. Ook beleidsvoorbereidend onderzoek of onderzoek door private instellingen komen in aanmerking op voorwaarde dat de verwerking onafhankelijk, transparant en via wetenschappelijke methoden gebeurt. Ook dienen de resultaten van het onderzoek publiek gemaakt te worden.

Het resultaat van de verwerking moet bestaan uit globale en anonieme statistieken en onderzoeksrapporten die geen individuele gevolgen mogen hebben voor de betrokken burgers of ondernemingen. Uiteraard kunnen deze gepubliceerde globale en anonieme resultaten nadien vrij gebruikt worden voor niet-statistische of niet-wetenschappelijke doeleinden.

¹ <https://bestat.statbel.fgov.be/bestat/>

² <https://statbel.fgov.be/nl/open-data>

WIE KAN GEGEVENS OPVRAGEN?

De Statistiekwet lijst de bestemmelingen op die in aanmerking komen om microgegevens van Statbel te ontvangen. Het gaat in de eerste plaats om federale, regionale, provinciale en lokale administraties en publieke instellingen. Hiertoe behoren ook de regionale statistische autoriteiten, het Federaal Planbureau en de statistische dienst van de Nationale Bank van België.

Ook natuurlijke en rechtspersonen die een wetenschappelijk doel nastreven, komen in aanmerking. Concreet gaat het om binnen- en buitenlandse universiteiten, studiediensten en internationale organisaties. Voor natuurlijke personen is het zo goed als onmogelijk om te voldoen aan de veiligheidsvereisten (zie verder) die Statbel oplegt om microgegevens te bekomen. Het gevolg is dat (doctoraats-)studenten enkel in aanmerking komen als de aanvraag uitgaat van de universiteit zelf en deze ook garant staat voor de bescherming van de gegevens.

WELKE GEGEVENS KOMEN IN AANMERKING?

Zowel bestaande (administratieve) gegevens die Statbel heeft opgevraagd als gegevens die Statbel zelf bij burgers of ondernemingen heeft ingezameld via enquêtes komen in aanmerking.

Wat de bestaande (administratieve) gegevens betreft, moet Statbel de voorwaarden nakomen die zijn vastgesteld door het Statistisch Toezichtscomité, het Informatieveiligheidscomité of het contract dat afgesloten werd met de leverancier van de data. Meestal is bepaald dat Statbel enkel gegevens ter beschikking mag stellen die het bewerkt heeft. Onderzoekers die op zoek zijn naar de ruwe gegevens dienen contact op te nemen met de leverancier van de data. Verder dient voor iedere variabele (of voor iedere cluster van variabelen) gemotiveerd te worden waarom de gegevens nodig zijn voor het onderzoek. Enquêtegegevens die Statbel zelf heeft ingezameld, worden aanzien als één samenhangende set van variabelen, waarbij de motivatie dient te gebeuren op het niveau van de enquête als geheel.

Statbel stelt aan onderzoekers enkel *gepseudonimiseerde gegevens* ter beschikking. Het gaat om gegevens op individueel niveau (persoon, gezin of onderneming) waarbij het voor de onderzoeker niet mogelijk is om de identiteit van de persoon, het gezin of de onderneming te achterhalen. Dit betekent in de eerste plaats dat de dataset geen informatie mag bevatten zoals een rijksregisternummer, ondernemingsnummer, naam, adres, telefoonnummer, ... die een *directe identificatie* mogelijk maken. De data moeten echter ook beschermd worden tegen *indirecte identificatie*. Dit betekent dat een combinatie van gegevens niet mag leiden tot één enkele persoon, gezin of onderneming. Dit wordt dataset per dataset geëvalueerd, maar over het algemeen komt het erop neer dat:

- de geboortedatum omgezet wordt naar een leeftijd of leeftijdsklasse;
- de postcode of gemeente vervangen wordt door het arrondissement, de provincie, het gewest of de urbanisatiegraad;
- bedragen afgerond, afgetopt of uitgedrukt worden in klassen of percentielen;
- NACE-codes van ondernemingen gehergroepeerd worden.

BEVOORRECHTE ONDERZOEKERS

De statistiekdiensten van Vlaanderen (Vlaamse Statistische Autoriteit), Wallonië (L'Institut wallon de l'évaluation, de la prospective et de la statistique) en Brussel (Brussels Instituut voor Statistiek en Analyse), de statistiekafdelingen van de Nationale Bank van België, het Federaal Planbureau en het Prijzenobservatorium van de FOD Economie mogen wettelijk gezien in het kader van hun opdracht binnen het Instituut voor de Nationale Rekeningen (INR) ook *niet-gepseudonimiseerde* microgegevens ontvangen. Ze vormen daarmee een uitzondering. Hierbij geldt wel de voorwaarde dat de niet-gepseudonimiseerde microgegevens noodzakelijk zijn in het kader van hun wettelijke opdracht om statistieken op te maken. Merk op dat deze instellingen en hun personeelsleden in het kader van hun statistische opdracht net als Statbel onderworpen zijn aan het statistisch geheim (zie verder).

KOPPELEN VAN GEGEVENS

Het kan voor onderzoekers een meerwaarde betekenen dat gegevens van Statbel gekoppeld worden aan gegevens van derden waarvoor de onderzoeker een toelating heeft bekomen in het kader van het onderzoeksproject. Omdat Statbel wettelijk erkend is (KB van 13 juni 2014) als *vertrouwde derde*, kan Statbel instaan voor het koppelen van beide gegevens en het pseudonimiseren ervan. Statbel bewaart de sleutel van de pseudonimisering, zodat op een later tijdstip eventueel bijkomende gegevens toegevoegd kunnen worden.

Bij een koppeling is het belangrijk dat de onderzoeker zelf geen toegang heeft tot gegevens met directe identificatie die de onderzoeker in de mogelijkheid stelt om de identiteit van de persoon, het gezin of de onderneming te achterhalen.

STATISTISCH GEHEIM EN DATALEKKEN

Alle gegevens waarover Statbel beschikt, zijn beschermd door het statistisch geheim. Dit betekent dat de gegevens op geen enkele manier gebruikt mogen worden voor een andere finaliteit dan statistisch en wetenschappelijk onderzoek. Bij een mededeling van microgegevens legt Statbel daarom enkele veiligheidsmaatregelen op:

- de finaliteit wordt contractueel vastgelegd, evenals een verbod om de microgegevens verder te verspreiden;
- de aanvrager dient te beschikken over een functionaris voor de gegevensbescherming (Data Protection Officer of DPO) en een veiligheidsadviseur (CSI);
- de gegevens dienen geplaatst te worden op een beveiligde server, die zich binnen Europa bevindt en waarvan de toegang beperkt blijft tot de onderzoekers die betrokken zijn bij het project;
- contractueel wordt vastgelegd dat de output enkel mag bestaan uit globale en anonieme statistieken.

HOE MICROGEGEVENS AANVRAGEN?

Er kunnen enkel gegevens aangevraagd worden waarover Statbel beschikt en die een meerwaarde betekenen voor het onderzoek. Om na te gaan of deze voorwaarde is vervuld, dient de onderzoeker voorafgaandelijk te overleggen met een statisticus van Statbel die vertrouwd is met het onderzoeksthema en de data. Tijdens dit gesprek zal op basis van de behoefte van de onderzoeker bepaald worden welke microgegevens in aanmerking komen voor een aanvraag. Een niet onbelangrijk aspect daarbij zijn de kwaliteit van de gegevens, het universum, de referentieperioden en het moment dat de data beschikbaar zijn. Tijdens deze fase zal ook nagegaan worden of de aanvraag geen (verhoogd) risico inhoudt inzake dataveiligheid, want in dat geval dient eerst overlegd te worden met de DPO.

Indien Statbel door het ter beschikking stellen van microgegevens een meerwaarde kan bieden voor het onderzoek en er op het eerste gezicht geen sprake is van een (verhoogd) risico, kan de onderzoeker een formele aanvraag indienen via een gestandaardiseerd formulier. Het formulier bevindt zich op de website van Statbel (<https://statbel.fgov.be/nl/over-statbel/privacy/microdata-voor-onderzoek>). Het formulier bestaat uit twee delen. Het eerste deel is het *aanvraagformulier voor microgegevens* en behandelt de volgende aspecten:

- de identiteit van de verwerkingsverantwoordelijke (of de wettelijke vertegenwoordiger);
- de finaliteit waarvoor de gegevens worden aangevraagd (mag enkel statistisch en wetenschappelijk onderzoek zijn);
- de gegevens die aangevraagd worden (variabelen, referentieperiode, bron, ...);
- de gewenste bewaartermijn met motivatie;
- welke categorie van personen toegang zullen krijgen tot de microgegevens;
- waaruit de output zal bestaan (moet beperkt blijven tot globale en anonieme resultaten);
- de identiteit van de DPO.

Het tweede luik bevat een *verklaring in verband met technische en organisatorische maatregelen* die een schending van het statistisch geheim of datalekken moet voorkomen. Concreet bevat dit tweede luik:

- de identificatie van de veiligheidsadviseur;

- de interne procedure in verband met datalekken;
- de opsomming van specifieke maatregelen om de schending van het statistisch geheim en datalekken tegen te gaan;
- de veiligheidspolitiek van de organisatie;

Het ingevulde formulier vormt de basis voor de beraadslaging van de aanvraag door Statbel en het eerste deel van het document dient daarom ondertekend te worden door de verwerkingsverantwoordelijke (of de wettelijke vertegenwoordiger). Het tweede deel dient ondertekend te worden door de veiligheidsadviseur, de verwerkingsverantwoordelijke of de wettelijke vertegenwoordiger van de organisatie. Met *verwerkingsverantwoordelijke* wordt niet de persoon bedoeld die aan het hoofd staat van het concrete onderzoek waarvoor de gegevens gebruikt zullen worden, maar wel de rechtspersoon of natuurlijke persoon zoals omschreven in artikel 24 van de AVG. Concreet gaat het om de persoon die de doelen bepaalt, de middelen ter beschikking stelt, het vertrouwelijkheidscontract ondertekent en aansprakelijk is voor als er iets zou mislopen. Het gaat bijvoorbeeld om de rector of de beheerder van een universiteit, de voorzitter van een federale administratie, de secretaris-generaal van een regionale administratie, enz.

BERAADSLAGING EN MACHTIGING DOOR STATBEL

Elke aanvraag zal door een multidisciplinair comité binnen Statbel beoordeeld worden. Daarbij wordt niet alleen gekeken of er voldaan is aan de juridische voorwaarden, maar ook of de gevraagde data proportioneel zijn ten opzichte van het onderzoeksdoel en of de technische en organisatorische maatregelen in verhouding staan tot de risico's. Indien nodig worden er bijkomende maatregelen opgelegd om het risico op indirecte identificatie, een hergebruik voor niet-statistische of niet-wetenschappelijke doelen of datalekken tegen te gaan.

Op basis van het advies zal de DPO van Statbel een formeel DPO-advies opmaken. Uiteindelijk is het de directeur-generaal die als wettelijke vertegenwoordiger van Statbel de goedkeuring geeft door het advies van de DPO te ondertekenen. Tot slot wordt er een vertrouwelijkheidscontract opgesteld dat ter ondertekening aan de verwerkingsverantwoordelijke van de onderzoeker zal worden voorgelegd. De beraadslaging door Statbel en het opstellen van het vertrouwelijkheidscontract nemen 2 à 3 weken in beslag, te rekenen vanaf de ontvangst van het correct ingevulde en ondertekende aanvraagformulier. De DPO-adviezen en ondertekende vertrouwelijkheidscontracten worden gepubliceerd op de website van Statbel: <https://statbel.fgov.be/nl/over-statbel/privacy/privacy-gdpr>.

VEREENVOUDIGDE PROCEDURES

De formele procedure die Statbel heeft uitgewerkt om microgegevens mee te delen, beperkt zich tot de informatie die nodig is om de aanvraag correct te kunnen beoordelen. Als de onderzoeker een bestaande machtiging wenst aan te passen, voorziet Statbel een *vereenvoudigd aanvraagformulier*. Het kan daarbij enkel gaan om een verlenging van de bewaarduur, het toevoegen van extra variabelen, het toevoegen van bijkomende referentieperioden en het uitbreiden van de onderzoeksfinaliteit. De vereenvoudiging ten opzichte van de normale procedure situeert zich in het aanvraagformulier dat minder uitgebreid is.

Voor sommige onderzoeken volstaan *geaggregeerde gegevens met een klein risico op indirecte identificatie* om het onderzoek uit te voeren. Omwille van dit kleine risico kunnen de gegevens niet als globaal en anoniem beschouwd worden en dient er een beraadslaging door Statbel te gebeuren. Ook hier volstaat een sterk vereenvoudigd aanvraagformulier.

BERAADSLAGINGEN EN MEDEDELINGEN VAN MICROGEGEVENS SINDE MEI 2018

De hierboven beschreven procedure voor het bekomen van microgegevens is van toepassing sinds de invoering van de AVG op 25 mei 2018 en het afschaffen van het Statistisch Toezichtcomité van de voormalige Commissie voor de bescherming van de persoonlijke levenssfeer (Privacycommissie). Tijdens de jaren 2018, 2019 en 2020 keurde Statbel 209 aanvragen van onderzoekers voor het bekomen van microgegevens goed, waarvan 50 via een vereenvoudigde procedure om een bestaande machtiging uit te breiden. Door het voorafgaandelijk contact tussen de onderzoeker en de statisticus van Statbel komt het amper voor dat aanvragen geweigerd worden. In sommige gevallen was er wel een bijsturing van de aanvraag nodig.

Type aanvraag	2018	2019	2020	Totaal
<i>Gewone aanvragen</i>	24	39	36	99
<i>Aanvragen door een andere statistische autoriteit die lid is van het IIS</i>	5	26	16	47
<i>Aanvragen in het kader van het INR</i>	4	3	6	13
<i>Uitbreiden van een bestaand contract</i>	12	15	23	50
<i>Totaal</i>	45	83	81	209

DE PRAKTIJK – USER CASES

De data-technische evolutie van het laatste decennium heeft geleid tot een doorgedreven hervorming van de reguliere statistiekproductie binnen Statbel, met onder andere een veel grotere nadruk op secundaire bronnen, zoals administratieve gegevens. Dit heeft grote efficiëntiewinsten opgeleverd, die als hefboom ingezet zijn om steeds geavanceerdere dataproducten te ontwikkelen waarmee aan de groeiende noden van onze stakeholders tegemoet wordt gekomen (onder andere beleidsdepartementen, de academische wereld, ...).

De dataproducten van Statbel omvatten heel veel inhoudelijke onderzoeks- en beleidsdomeinen (economische, sociale en demografische data) en de innovaties nemen verschillende vormen aan:

- sommige nieuwe dataproducten zijn directe uitbreidingen van de reguliere data- en statistiekproductie. Vaak gebeurt dit op vraag van belangrijke institutionele spelers. Ad-hoc-modules bij enquêtes vormen hiervan een voorbeeld.
- tot op zekere hoogte zijn administratieve data nog steeds ondergevaloriseerd. Statbel heeft verschillende dataproducten ontwikkeld die de bronnen waarover het beschikt veel beter ontsluiten en tot rijkere valorisaties leiden, waardoor economische en maatschappelijke fenomenen die tot voor kort moeilijk in kaart te brengen waren, nu belicht kunnen worden. De recente ontwikkeling van herkomstgegevens en ook longitudinale datareeksen op basis van het rijksregister zijn hier voorbeelden van.
- via koppeling tussen administratieve data kunnen inzichten bekomen worden die volstrekt obscuur blijven wanneer men onderzoek of beleidsvoorbereidend werk baseert op één enkele bron. De administratieve census is een voorbeeld van een standaardproduct waarbij dit principe expliciet in de praktijk gebracht wordt. Het erg rijke data-ecosysteem van Statbel staat echter toe op een veel vrijere manier informatie uit verschillende databases te integreren. Bovendien gaat het niet alleen over data ingezameld door Statbel, maar wordt er ook samengewerkt met andere datahouders, zoals de Kruispuntbank Sociale Zekerheid (KSZ) en het InterMutualistisch Agentschap (IMA). Dit alles gebeurt strikt binnen de krijtlijnen van de wet.

Zeker wat het laatste betreft, maar in feite globaal, biedt Statbel maatwerk aan, hoewel nu grote dataconsolidaties aan het ontstaan zijn doordat verschillende op maat gemaakte databases een verspreiding kennen bij uiteenlopende gebruikers. Dit heeft dan weer een harmoniserend effect en leidt er zelfs toe dat onderzoeksconsortia zich enigszins rond onze data organiseren.

De onderstaande niet-exhaustieve lijst omvat een korte bloemlezing van verschillende, veeleer recente onderzoeksprojecten die gebruik maken van door Statbel terbeschikkinggestelde microdata:

- **Immilab** is een door het Federaal Wetenschapsbeleid (BELSPO) gesponsord project dat de positie van personen met een migratieachtergrond op de Belgische arbeidsmarkt onderzoekt. Onderzoeksteams van de UA, de ULB en UMons participeren in dit project. Statbel draagt bij via een innovatieve koppeling tussen de enquête naar de structuur en de verdeling van de lonen (SES) en demografische gegevens;

- **Re-invest** betreft een project dat focust op de link tussen sociale bescherming, wonen en gezondheid. Dit project gebruikt de longitudinale data 2004-2018 van de Statistics on Income en Living Conditions (SILC) als ruggengraad, verrijkt met Census-data (2011) en met demografische informatie (onder andere over verhuisbewegingen). Partners in dit project zijn de KU Leuven – HIVA, het Steunpunt ter bestrijding van armoede en de UCL.
- **SUSPENS** is een door BELSPO gefinancierd project, uitgevoerd door onderzoekers van de Universiteit Antwerpen, de ULB en het Federaal Planbureau. Als bronnen worden de Statbel-enquêtes SILC en Huishoudbudgetonderzoek gebruikt, zodat het verband kan worden gelegd tussen het inkomen, het consumentengedrag en de uitstoot van broeikasgassen op gezinsniveau.
- De **VMM** (Vlaamse Milieumaatschappij) berekent een indicator om de betaalbaarheidstoets van de integrale waterfactuur per jaar uit te voeren. Uitgaven voor de integrale waterfactuur worden vergeleken met het gezinsinkomen, zodat beleidsadvies kan worden gegeven over betaalbaarheid van drinkwater. Statbel koppelde hiertoe informatie van SILC aan gegevens van de waterfactuur van de VMM. Om de data te beschermen tegen identificatie door de VMM werden bijkomende technische en organisatorische maatregelen opgelegd.
- **Causineq** en de **mortaliteitsdatabanken** tijdens de **COVID-19**-epidemie: Causineq is een groots opgezet onderzoek naar differentiële mortaliteit volgens sociale gradiënt waarin de UCL en de VUB participeren. Vanaf het begin van de COVID-19-crisis heeft Statbel gelijkaardige wekelijkse mortaliteitsdata geproduceerd die door verschillende instanties en universiteiten gebruikt werden om oversterfte in kaart te brengen (Sciensano, UCL, VUB, UHASSELT).

Misschien nog markanter is dat, parallel aan de technisch-methodologische vooruitgang die de laatste jaren geboekt werd, ook binnen Statbel een cultuurshift plaatsvond. De verschuiving van een louter “statistiek-producerende” organisatie naar een belangrijke dataproductent is hand in hand gegaan met het aannemen van een meer open manier van werken en vooral ook van een belangrijke toename in de dienstverlening naar de stakeholders.

CONCLUSIE

Statbel beschikt over hoogstaande en kwaliteitsvolle data en over het wettelijke mandaat om deze data ter beschikking te stellen aan onderzoekers in het kader van statistisch en wetenschappelijk onderzoek. Daarnaast heeft Statbel de nodige expertise in huis om technisch hoogstaande dataproducten op maat te maken. Hierdoor evolueerde Statbel de voorbije jaren naar een organisatie die onderzoekers actief helpt door data ter beschikking te stellen die inspelen op hun concrete noden.