



EDITORIAAL

Philippe Mauroy
Vicevoorzitter van de raad van bestuur

Belgische statistiekinstellingen onderworpen aan Europese audit

Dagelijks worden we overspoeld door een golf van informatie, feiten en statistieken. Openbare statistiekinstellingen spelen daarbij een belangrijke rol. Hun doel is om iedereen toegang te geven tot statistieken die kwaliteitsvol, betrouwbaar, vergelijkbaar en geloofwaardig zijn.

Langs de andere kant moeten gebruikers kunnen dubbelchecken of die missie ook waargemaakt wordt, en of de gegevens van hoge kwaliteit zijn. Alleen zo kunnen de statistiekinstellingen hun gebruikers overtuigen om vertrouwen te hebben.

Om dat vertrouwen verder te verhogen, was er in december 2021 een Europese audit in België. Die audit, of peer review, werd uitgevoerd bij verschillende instellingen die Europese statistieken produceren in België. Een groep internationale deskundigen onderzocht in hoeverre deze statistieken in overeenstemming zijn met de Praktijkcode voor Europese statistieken. Dat onderzoek zal leiden tot aanbevelingen die in het voorjaar van 2022 gepubliceerd zullen worden.

In deze nieuwsbrief vindt u een artikel waarin de principes van de peer review in het Europees statistisch systeem worden geduid.

In dit nummer komen verder volgende onderwerpen aan bod, die het werk van het Interfederaal Instituut voor de Statistiek voorstellen: de indicatoren voor de opvolging van de doelstellingen inzake duurzame ontwikkeling, de verbetering van de statistieken over woningen in het kader van de Census en de Europese Statistiekolympiade. Ten slotte zal in een artikel worden uiteengezet hoe de techniek van machine learning gebruikt kan worden voor de productie van openbare statistieken.

We wensen u een aangename leeservaring toe.

IIS NIEUWSBRIEF

INHOUD

- p. 1** [Editoriaal](#)
- p. 2-3** [De peer reviews in het Europees statistisch systeem](#)
- p. 4-5** [Indicatoren voor de opvolging van de duurzame-ontwikkelingsdoelstellingen \(SDG's\)](#)
- p. 6-7** [Census Woningen – Samenwerken voor meer en betere statistieken](#)
- p. 8-9** [Europese Statistiekolympiade daagt jongeren uit](#)
- p. 10-11** [Machine Learning voor openbare statistieken](#)



EUROPEAN STATISTICAL SYSTEM

De peer reviews in het Europees statistisch systeem



Martine Lefèvre



eurostat

De peer review is het voornaamste instrument dat wordt gebruikt om de vooruitgang bij de uitvoering van de [Praktijkcode voor Europese statistieken](#) te analyseren. Het wordt in het [Handbook of Statistical Organization \(2021\)](#) gedefinieerd als een "Evaluation of a statistical organization, statistical process and/or statistical output by one or more people with similar competencies to the producer. Often used for quality evaluation of a national statistical system, statistical process and/or statistical output by peers from within the same statistical organization or from another statistical organization" (p. 551).

Op ESS-niveau hebben sinds de goedkeuring van de praktijkcode in 2005 twee evaluatierondes plaatsgevonden en is een derde momenteel aan de gang. Zij hebben betrekking op alle lidstaten van de EU, de EVA-landen (IJsland, Liechtenstein,

Noorwegen en Zwitserland) en Eurostat.

De tekst hierna geeft een overzicht van de thematiek. Binnen het IIS zijn de statistische autoriteiten onderworpen aan deze code via art. 36 van het [Samenwerkingsakkoord van 15 juli 2014](#), ondertekend tussen de Federale Staat en de deelstaten.

De Praktijkcode van de Europese statistiek

Als onmisbaar onderdeel van het informatiesysteem van iedere democratische samenleving¹ worden officiële statistieken op Europees niveau gereguleerd door een praktijkcode. Alle statistische autoriteiten die verantwoordelijk zijn voor de ontwikkeling, productie en verspreiding van officiële statistieken, hebben zich geëngageerd om deze code na te leven.

¹ Verenigde Naties (2014), eerste fundamentele principe van de officiële statistiek.



Zoals aangeduid in art. 11 van de [Europese verordening betreffende de Europese statistiek](#): “De Praktijkcode heeft tot doel het publiek vertrouwen in de Europese statistiek te waarborgen en stelt daartoe vast hoe Europese statistieken in overeenstemming met de statistische beginselen als uiteengezet in artikel 2, lid 1, en de beste internationale statistische praktijken moeten worden ontwikkeld, geproduceerd en verspreid.” De beoogde statistische beginselen zijn professionele onafhankelijkheid, onpartijdigheid, betrouwbaarheid, statistische geheimhouding en kosteneffectiviteit.

Onder toezicht van het [Comité voor het Europees statistisch systeem](#) is deze code, die in 2005 is aangenomen, in 2011 en in 2017 herzien. Dit zelfreguleringsinstrument, dat de hoeksteen vormt van het gemeenschappelijk kwaliteitskader van het ESS, is gebaseerd op zestien beginselen, die betrekking hebben op het institutionele en organisatorische kader, de statistische processen en de resultaten, elk met een reeks indicatoren die worden gebruikt bij het onderzoeken van de uitvoering ervan.

Opvolging van de uitvoering van de code door middel van peer reviews

Deze peer reviews, die door Eurostat worden georganiseerd, verlopen in verschillende fasen: zelfbeoordeling door elke statistische autoriteit van haar naleving van de beginselen van de code, analyse van de documenten door een team van deskundigen dat vervolgens een bezoek aan het betrokken land brengt voor een diepgaand onderzoek naar de uitvoering van de code en een verslag met aanbevelingen voor verbetering opstelt. Dit verslag wordt aan de betrokken statistische autoriteit voorgelegd ter goedkeuring en voor het uitwerken van maatregelen die moeten worden genomen om de aanbevelingen op te volgen. Op het niveau van de landen wordt de opvolging van de uitvoering van de verbeteracties waargenomen door Eurostat; terwijl dit op het niveau van Eurostat een taak is van de [Europese Adviescommissie voor statistische governance](#), die als missie heeft «onafhankelijk toezicht uit te oefenen op het ESS wat de uitvoering van de Praktijkcode voor Europese statistieken betreft». Al deze werkzaamheden en de opvolging ervan worden gepubliceerd op de [website van Eurostat](#).

- 1e peer review (2006 – 2008)

Bij de eerste review werd nagegaan in hoeverre de nationale statistische instituten (NSI's) de beginselen van de code naleven die betrekking hebben op hun institutionele omgeving en verspreidingspraktijken. Het [verslag van de Commissie aan het Europees Parlement en de Raad \(oktober 2008\)](#) geeft een samenvatting van de resultaten hiervan en voorziet in een nieuwe ronde van peer reviews in de volgende vijf jaar.

In 2012 heeft het Comité voor het Europees statistisch systeem aanbevelingen voor het opstarten van een nieuwe onderzoeks cyclus gedaan, onder meer rekening houdend met het beperkte aantal beginselen van de code die in de eerste cyclus werden onderzocht, de herziening van de code in 2011, de gezamenlijke goedkeuring van een [kader voor kwaliteitsbeheer](#)², de ontwikkelingen op het vlak van governance van de Europese statistiek, waaronder de publicatie van de verordening betreffende de Europese statistiek in 2009, en de conclusies [van het speciaal verslag van de Europese Rekenkamer van 2012 over de Europese statistiek](#).

- 2e peer review (2013 – 2015)

Een tweede peer review werd aldus gerealiseerd binnen het ESS tussen 2013 en 2015. Die verschilde in verscheidene opzichten van de eerste, waaronder de uitbreiding van de evaluatie tot alle beginselen van de code en de opname in het onderzoek van een bepaald aantal [andere nationale statistische autoriteiten \(ONAs\)](#). Een [verslag van de review werd opgemaakt door de Commissie en overgemaakt aan het Europees Parlement en de Raad \(maart 2016\)](#).

In haar [jaarverslag van 2020](#) (p. 36), meldde de Europese Adviescommissie voor statistische governance dat als resultaat van deze tweede peer review 707 aanbevelingen werden voorgesteld, die hebben geleid tot de formulering van 910 verbeteringsacties door de betrokken statistische instellingen. De commissie merkte op dat er eind 2019 nog 118 acties moeten worden voltooid.

- 3e peer review (2021 – 2023)

Op basis van de resultaten van de vorige peer review en na de herziening van de code in 2017 hebben de verantwoordelijken van de Europese statistiek besloten om met ingang van 2018 de conformiteit van de nationale statistische autoriteiten, de andere nationale autoriteiten en Eurostat met de herziene code opnieuw te evalueren. Het toepassingsgebied van deze derde peer review breidt zich uit tot de andere nationale autoriteiten, die door het NSI geselecteerd werden, vóór het bezoek van de peer reviewers, op basis van een specifieke procedure.

Er wordt rekening gehouden met de impact van de Covid 19-pandemie op de officiële statistieken en met de toenemende complexiteit van de productieomgeving voor statistieken. De peer review heeft twee doelstellingen :

- aan de partners binnen het ESS en aan externe stakeholders van het ESS aantonen dat het ESS gebaseerd is op en werkt volgens fundamentele beginselen;
- de NSI's, de geselecteerde andere nationale autoriteiten (ONA's) en Eurostat helpen hun statistisch systeem te verbeteren en verder te ontwikkelen door toekomstgerichte aanbevelingen voor verbeteringen te doen die verder kunnen gaan dan de huidige code; maar ook door overheidsinstanties aan te moedigen de uitvoering van de aanbevelingen te ondersteunen.

Deze doelstellingen zijn gericht op zowel interne als externe stakeholders van het ESS. Deze externe dimensie is belangrijk gezien de behoefte aan betrouwbare statistieken en de noodzaak om het vertrouwen van het publiek in officiële statistieken te versterken. Om deze doelstellingen te bereiken, werd begin 2021 besloten een specifieke communicatiestrategie op te zetten rond de uitvoering van deze 3e peer review.

[Alle documentatie over de uitvoering en de modaliteiten van de organisatie van deze derde peer review](#) is gepubliceerd op de website van Eurostat, met de aanduiding van de planning van de evaluaties en de kalender die voorzien is voor de publicatie van de verslagen. Voor België vond de peer review plaats van 6 tot 10 december 2021 en het verslag wordt verwacht in april 2022.



2 Gids en aanvullingen voor de uitvoering van de Praktijkcode voor Europese statistieken.

Indicatoren voor de opvolging van de duurzame-ontwikkelingsdoelstellingen (SDG's).

Alain Henry

Een IIS-werkgroep coördineert de opvolging van de Sustainable Development Goals op basis van indicatoren. Er werd een set van 116 indicatoren opgesteld, waarvan er momenteel 81 gepubliceerd zijn op de [website « indicatoren » van het Federaal Planbureau](#). Deze set kan in de toekomst nog evolueren, met name na het overleg met de belangrijkste federale adviesraden in 2022.

De SDG's en hun opvolgingsindicatoren

In 2015 heeft de VN de [Sustainable Development Goals](#) (SDG's, duurzame-ontwikkelingsdoelstellingen) aangenomen, een reeks van 17 ambitieuze doelstellingen voor 2030. Deze 17 SDG's zijn onderverdeeld in [169 subdoelstellingen](#). De opvolging van de vooruitgang in de richting van deze doelstellingen is georganiseerd rond [231 indicatoren](#) die de Statistische Commissie van de VN heeft opgesteld.

Met de goedkeuring van de SDG's, roept de VN landen op om de vooruitgang naar deze doelstellingen op te volgen aan de hand van indicatoren. In België kreeg het IIS de opdracht om de opvolging van deze indicatoren te organiseren en ze heeft daartoe in 2016 de werkgroep 'Indicatoren voor SDG's' opgericht. Het is een van de vier permanente werkgroepen, belast met de voorbereidende werkzaamheden voor de Raad van Bestuur van het IIS. De werkgroep is samengesteld uit vertegenwoordigers van Statistics Belgium, de Nationale Bank van België, het Federaal Planbureau (FPB) en de gewestelijke statistische diensten (Statistiek Vlaanderen, IWEPS en BISA). Het FPB fungeert als voorzitter en secretariaat.

Deze werkgroep heeft als taak de werkzaamheden over de indicatoren voor de opvolging van de SDG's tussen de partners van het IIS te coördineren. Ze formuleert voorstellen voor de ontwikkeling van een databank met deze indicatoren en voor het opstellen van nieuwe indicatoren. Ze volgt ook de werkzaamheden van de VN en Eurostat op dit gebied op en formuleert, zo nodig, aanvullende adviezen en internationale standpunten over statistische vraagstukken in verband met de indicatoren voor de monitoring van de SDG's.

Methode om de indicatorenlijsten op te stellen

Het initiële mandaat van de werkgroep betrof het identificeren van de in België al beschikbare indicatoren van de VN-lijst, waarvoor bovendien de opvolging snel kon worden georganiseerd. De werkgroep heeft de indicatoren opgesplitst volgens verschillende categorieën. Ze bouwde daarbij voort op de Eurostat-methode om de indicatoren op Europees niveau te definiëren rekening houdend met de in België bestaande gegevens. Overeenkomstig de consensus onder de deelnemende experts werden sommige indicatoren uit de werkzaamheden van de werkgroep uitgesloten, omdat zij van niet-statistische aard waren (bijvoorbeeld, beschikt België over een strategie om de risico's bij natuurrampen te beperken - VN-indicator 1.5.3) of niet van toepassing waren op België (bijvoorbeeld, wereldwijd gewogen gemiddelde douanerechten, indicator 17.10.1). De overige indicatoren werden vervolgens opgesplitst in indicatoren waarvoor gegevens beschikbaar waren en indicatoren waarvoor nog verder onderzoek vereist was.

Resultaten: indicatorensets

Als resultaat van deze werkzaamheden heeft de Raad van Bestuur van het IIS in februari 2021 een set van 116 indicatoren goedgekeurd om de SDG's in België te monitoren.

Van die 116 indicatoren zijn er momenteel 81 gepubliceerd op de [website « indicatoren van duurzame ontwikkeling » van het FPB](#). Voor tal van indicatoren worden ook opsplitsingen naar bevolkingscategorie voorgesteld. Zo zijn 37 indicatoren opgesplitst volgens geslacht, 40 volgens gewest, 19 volgens inkomen, 11 volgens opleiding, en zo verder. Voor 73 indicatoren wordt een vergelijking gemaakt met het Europese gemiddelde of met de buurlanden.

Van die 81 indicatoren zijn er 37 niet opgenomen in de VN-lijst. Ze behandelen vraagstukken die relevant zijn in België maar niet op mondiaal niveau zijn opgenomen. Het gaat bijvoorbeeld om obesitas bij volwassenen (subdoelstelling 2.2 over slechte voeding), uitstel of afstel van medische zorg om financiële redenen (subdoelstelling 1.4 over de toegang tot basisvoorzieningen), primair energieverbruik (subdoelstelling 7.3 over energieproductiviteit) of verkeersopstoppingen (subdoelstelling 9.1 over de toegang tot kwaliteitsinfrastructuur).

Deze set van 81 indicatoren zou kunnen worden uitgebreid tot de 116 hiervoor vermelde indicatoren. Naast deze 116 indicatoren publiceert de VN 31 bijkomende indicatoren voor België. Deze indicatoren zouden ook kunnen worden toegevoegd als de VN-methodes een directe link zouden kunnen leggen met de Belgische databronnen. Ten slotte heeft de werkgroep 87 andere indicatoren geïdentificeerd die nader onderzoek vereisen voordat zij aan deze set kunnen worden toegevoegd. Voor de uitvoering van deze werkzaamheden moeten echter extra personeelsmiddelen ter beschikking worden gesteld.

Gebruik van deze indicatoren

In 2017 werd een selectie van 34 indicatoren ook gepresenteerd in de statistische bijlage van de [Voluntary National Review](#) die België aan de VN heeft voorgesteld op het Politiek forum op hoog niveau voor duurzame ontwikkeling.

De opvolgingsindicatoren van de SDG's worden regelmatig gebruikt op het FPB. De vooruitgang van België naar de SDG's wordt gemeten aan de hand van een selectie van 51 indicatoren. Voor elke indicator evalueert het FPB of de doelstelling, gebaseerd op de SDG's en/of andere door België aangegane verbintenissen, zal kunnen worden bereikt met de voortzetting van de huidige trends.

De laatste evaluaties zijn in juni 2019 gepubliceerd in het [Federaal rapport inzake duurzame ontwikkeling](#), in juli 2021 in een artikel [Trendbreuk nodig om alle SDG's tegen 2030 te bereiken](#), en in februari 2022 in [het rapport Indicatoren van duurzame ontwikkeling](#). De titel van de publicatie uit 2021 vat samen welke besluiten getrokken kunnen worden uit de opeenvolgende evaluaties die sinds 2005 zijn uitgevoerd.

Het FPB gebruikt deze indicatoren ook in de voorbereiding van de Federale rapporten inzake duurzame ontwikkeling, zowel in het deel 'evaluatie' als in het deel 'toekomstverkenning' van deze rapporten.

Toekomstige werkzaamheden

Deze indicatorenlijsten evolueren voortdurend naarmate de VN wijzigingen aanbrengt, nieuwe gegevens beschikbaar worden, nieuwe vraagstukken moeten worden behandeld (bijvoorbeeld geestelijke gezondheidsproblemen ten gevolge van de covid-19-pandemie), en zo verder. Het is ook de taak van de werkgroep om deze evoluties op te volgen en de Raad van Bestuur van het IIS te informeren.

Op Belgisch niveau wordt verwacht dat twee gebeurtenissen deze indicatorenlijsten zullen beïnvloeden. Ten eerste voorziet het in oktober 2021 door de federale regering goedgekeurde Federaal plan inzake duurzame ontwikkeling in een debat over de keuze van de opvolgingsindicatoren van de SDG's. Een voorstel van de IIS-experten en de federale overheidsdiensten zal worden voorgelegd aan de belangrijkste federale adviesraden.

Vervolgens moet België tegen medio 2023 een tweede Voluntary National Review opstellen. Net als in de editie van 2017 zal dit rapport waarschijnlijk een statistische bijlage bevatten. De geformuleerde vragen in verband met deze bijlage zullen ook als leidraad dienen voor de werkzaamheden van het IIS met betrekking tot opvolgingsindicatoren



Census Woningen - Samenwerken voor meer en betere statistieken

Saskia Hin

De IIS Census Woningen werd eind 2018 opgericht om de statistieken rond wonen te verbeteren voor de Census 2021. Eerder geobserveerde verschillen tussen het aantal woningen en het aantal huishoudens werden nader bestudeerd om oorzaken te achterhalen en metingen te kunnen verbeteren. Ook werd de behoefte aan nieuwe indicatoren en variabelen geïnventariseerd. Die leidde tot een lijst met aanbevelingen voor de aankomende census.

Doelstellingen

Het IIS zette eind 2018 een samenwerkingsverband op over de Census Woningen ter voorbereiding van de implementatie van de Belgische Census 2021. De Census 2021 vindt doorheen de EU plaats en Eurostat coördineert de werkzaamheden. Zo voorziet Eurostat homogene definities en metadata en verzamelt het de [resultaten van de census](#). Naast de verplichte indicatoren staat het landen vrij om aanvullende indicatoren te verzamelen volgens eigen definities.

De IIS-samenwerking concentreerde zich specifiek op het thema wonen, één onderdeel van de census. Deze samenwerking vond plaats tussen Statbel, VSA, IWEPS en BISA onder voorzitterschap en secretariaat van het IWEPS en had twee doelstellingen:

- Vooreerst moest uitgeklaard worden waarom er bij de Census 2011 in sommige statistische sectoren of gemeenten aanzienlijk minder woningen waren dan huishoudens, met als doel de methode om woningen te identificeren te verbeteren;
- Ten tweede onderzocht de werkgroep welke nieuwe aanvullende statistische indicatoren, en/of variabelen idealiter ontwikkeld zouden worden in het kader van de Census. Er was in het bijzonder behoefte aan indicatoren die informatie geven over het eigenaarschap en de profielen van eigenaars, alsook de kenmerken van woningen.

Eerste fase: gegevensaanvragen voor analyse

Een eerste stap bestond uit het verwerven van de benodigde gegevens. Hiervoor werden verschillende wegen verkend. Uiteindelijk bleek de gewenste aanpak van een gezamenlijke aanvraag van alle projectpartners juridisch te complex en dienden de verschillende partners elk afzonderlijk een aanvraag in. Eind 2019 konden de verschillende partners aan de slag met de gegevens. Die bestonden uit gekoppelde informatie uit het kadaster, het rijksregister en de Kruispuntbank van Ondernemingen (KBO).

Identificatie van de woningen

Als de data uit het rijksregister en deze uit het kadaster met elkaar gekoppeld worden, ziet men regelmatig dat het aantal huishoudens op een adres groter is dan het aantal woningen volgens het kadaster op datzelfde adres. De oorzaak hiervan kan bijvoorbeeld liggen in huishoudens die zijn geëmigreerd maar nog in het bevolkingsregister ingeschreven staan (overregistratie). Het kan ook gaan om woningen die niet in het kadaster geregistreerd staan (onderregistratie). Hierbij kan het gaan om woningen die werden verbouwd tot meerdere appartementen, maar waarbij deze gewijzigde situatie niet in het kadaster is opgenomen.

Om te voorkomen dat het aantal woningen hierdoor onderschat zou worden, werden in de census "virtuele woningen" aangemaakt. De methodologie die in de census 2011 gebruikt werd om deze virtuele woningen aan te maken werd door de IIS-werkgroep onder de loep genomen.

De vastgestelde problematiek waarbij er minder woningen zijn dan huishoudens verschilt sterk tussen de gewesten. In het Vlaamse Gewest deed het probleem zich alleen voor in Antwerpen, Leuven en Gent. In het Brusselse Hoofdstedelijke Gewest gold het voor alle gemeenten, al was het meer uitgesproken in de centrumgemeenten. In het Waalse Gewest betrof het een mix van gemeenten met onderling sterk verschillende kenmerken (bijvoorbeeld zowel stedelijk als ruraal).

De geografische verschillen in de aanmaak van deze virtuele woningen werden in detail bestudeerd en het proces en de hypothesen achter de identificatie van virtuele woningen werden geëvalueerd. Op basis van deze studie heeft de werkgroep aanbevelingen geformuleerd om de methode voor het bepalen van deze virtuele woningen in de volgende census nog te verbeteren.

Er werd tevens een vragenlijst gestuurd naar de collega's van andere Europese landen die eveneens een administratieve woningencensus organiseren. In andere Europese landen waren de verschillen tussen het kadaster en het bevolkingsregister beduidend kleiner dan bij ons. De relatie tussen de administraties die gegevens produceren zoals het kadaster en de statistische instituten is in andere landen over het algemeen wederkerig: de statistische instituten gebruiken niet alleen de gegevens van de administraties, maar de administraties houden ook rekening met de suggesties van de statistische instituten. In België gebruiken de statistische instituten de gegevens van de administraties. De administraties verbeteren echter hun werking niet op basis van mogelijke suggesties van de statistiekinstellingen. De praktijkcode voor Europese statistieken waartoe de verschillende regionale overheden en de federale overheid zich engageerden, vermeldt hieromtrent volgende richtlijn: "Wanneer Europese statistieken gebaseerd zijn op administratieve en andere gegevens, moeten de definities en begrippen die voor niet-statistische doeleinden worden gebruikt, goed aansluiten bij de voor statistische doeleinden gehanteerde definities en

begrippen." Een onderling overleg in het kader van een betere afstemming is dan ook zinvol en kan een win-win opleveren voor zowel statistische instellingen als publieke administraties..

Nieuwe indicatoren en variabelen

Eind 2020 werd de samenwerking tussen dezelfde partners verlengd om prioritaire nieuwe indicatoren en variabelen rond het thema wonen te identificeren in de regio's. De thema's waarrond werd gewerkt zijn het eigenaarschap en de profielen van eigenaars (zoals het aantal woningen in bezit van de eigenaar of de aanwezigheid van vruchtgebruik), alsook de kenmerken van woningen (zoals het energieprestatiecertificaat en de leegstand). De werkgroep besprak deze prioriteiten en maakte een gemeenschappelijke ranking van prioriteiten op. De definities werden gedetailleerd uitgewerkt om zo een éénduidige omschrijving te verzekeren en de vergelijkbaarheid van indicatoren tussen de regio's te waarborgen. Eind september 2021 werd een lijst met gewenste indicatoren en variabelen geconsolideerd en opgenomen in een intern rapport ten behoeve van de Raad van Bestuur van de IIS. De haalbaarheid van de verschillende indicatoren en variabelen

werd ingeschat door het team van Statbel en resulteerde in een haalbaarheidsranking.

En verder?

De doelstellingen van het samenwerkingsproject werden hiermee bereikt en het project werd formeel afgesloten. Toch willen de IIS-partners het platform behouden voor verdere kennisuitwisseling rond het thema huisvesting.

Statbel zal - na de aanlevering van de voor de census verplichte indicatoren en variabelen aan Eurostat - in functie van de beschikbare tijd en middelen bepalen wanneer welke door het IIS gewenste aanvullende indicatoren en variabelen effectief door hen geproduceerd kunnen worden. Daarbij is Statbel vragende partij voor een mogelijke detachering van medewerkers van de regionale partners. Er zal in 2022 opnieuw een IIS-werkgroep met dezelfde partners opgericht worden om dit proces verder op te volgen in een gecontinueerde samenwerking tussen de regionale partners en Statbel.



Europese Statistiekolympiade daagt jongeren uit

Wendy Schelfaut

In de Europese Statistiekolympiade worden leerlingen uit de tweede en derde graad secundair uitgedaagd in hun kennis over en interpretatie van officiële statistieken. De Olympiade wordt sinds het schooljaar 2018-2019 in België georganiseerd door Statbel, Statistiek Vlaanderen en IWEPS, in samenwerking met Eurostat en INE Spanje.

782 leerlingen nemen deel

De Europese Statistiekolympiade is een atypische wedstrijd. Anders dan in andere olympiades werken de leerlingen eerder in teams. Eén leerkracht begeleidt zijn of haar leerlingen doorheen de wedstrijd. Samen beslissen ze of leerlingen individueel werken, in duo of toch in teams van drie.

In het huidige schooljaar 2021-2022 schreven 322 teams zich in voor de Europese Statistiekolympiade: 246 uit de derde graad, 76 teams uit de tweede graad secundair.

Twee derde van de teams neemt deel in het Nederlands (213 teams), een derde in het Frans (109 teams).

Theorie, onderzoek en creatieve vaardigheden

In de Europese Statistiekolympiade worden verschillende vaardigheden getest bij de deelnemende leerlingen: hun theoretische kennis van statistiek, hun onderzoekscompetenties en hun creatieve vaardigheden. Alle proeven verlopen online. De leerlingen worden doorheen alle proeven begeleid door een leerkracht. Die kan uit eender welk vak komen: wiskunde, economie, aardrijkskunde, maar ook taalvakken of meer creatieve vakken. Uit al die vakken komen namelijk vaardigheden aan bod, en meer dan eens werken vakleerkrachten binnen de school ook samen om hun teams zo goed mogelijk te ondersteunen.

In België verloopt de Statistiekolympiade in twee fasen. Tijdens een **eerste online proef (januari 2022)** worden de teams getest op 3 vaardigheden: basiskennis statistiek, analysevaardigheden en statistische rapporten interpreteren. Deze test is eliminerend: teams slagen als ze minstens 50% halen. Van de geslaagde teams gaat enkel de top 20 per gemeenschap en graad door naar de tweede proef. Dus maximaal 20 Nederlandstalige teams uit de 2de graad en 20 Nederlandstalige teams uit 3de graad, en idem voor de Franstalige en Duitstalige gemeenschap.

In de **tweede proef (februari – maart 2022)** wordt gepeild naar creatieve competenties: kunnen de leerlingen in een video correct en creatief communiceren over statistieken? Alle geselecteerde teams krijgen de opdracht om een video te creëren van maximum 2 minuten over een vooraf bepaald onderwerp, in het Engels, of minstens met Engelstalige ondertitels. Dat thema is hetzelfde voor alle Europese landen, aangezien de winnaars van alle landen tegen elkaar strijden op Europees niveau.

Begin mei 2022 is er de **Belgische prijsuitreiking** en bekendmaking van de vier Belgische teams (twee per graad) die met hun video deelnemen aan de Europese ronde.

Statistische geletterdheid

Statistiek is nauw verbonden met het dagelijks leven. We worden overspoeld door cijfers en gegevens uit de meest uiteenlopende bronnen. Er zijn officiële statistiekbureaus, zoals Eurostat voor Europa, Statbel voor België, en Statistiek Vlaanderen, Iweps en BISA voor respectievelijk Vlaanderen, Wallonië en het Brusselse Hoofdstedelijke Gewest.

Als openbare statistiekinstellingen zetten zij zich in voor de statistische geletterdheid

van een breed publiek. De Europese Statistiekolympiade daagt leerlingen uit om aan de slag te gaan met statistieken, en te leren om gegevens te interpreteren en te presenteren.



Internationale en nationale partners

Als wedstrijd is de Europese Statistiekolympiade uniek in zijn soort. Er wordt over en binnen de landsgrenzen samengewerkt om de competitie te doen slagen. Internationaal zijn Eurostat en INE Spanje de centrale partners. Zij staan in voor de internationale coördinatie, zorgen voor de technische ondersteuning en organiseren de Europese fase van de wedstrijd. Daarnaast wordt ook geregeld overlegd met de statistiebureaus uit alle deelnemende landen: Bulgarije, Cyprus, Finland, Frankrijk, Griekenland, Hongarije, IJsland, Ierland, Italië, Kroatië, Liechtenstein, Litouwen, Luxemburg, Noorwegen, Polen, Portugal, Slovenië, Spanje en uiteraard België.

Binnen België ontstond de samenwerking rond de Europese Statistiekolympiade in het Interfederaal Instituut voor de Statistiek (IIS). In 2018 besloten Statbel, Statistiek Vlaanderen en IWEPS om de handen in elkaar te slaan en de Olympiade te organiseren in het hele land. De drie instellingen werken samen om de communicatie over de Olympiade zo breed mogelijk te laten verlopen. De drie partners zijn bovendien betrokken bij de inhoudelijke evaluaties via gemeenschapsjury's en één nationale jury, en bij de nationale prijsuitreiking op het einde van het proces.

Naast de samenwerking binnen het IIS wordt de Europese Statistiekolympiade in België ook ondersteund door andere partners: het Departement Economie, Wetenschap en Innovatie en het Departement Onderwijs van de Vlaamse Gemeenschap, het ministerie van de Duitstalige Gemeenschap, het departement onderwijs van de Federatie Wallonië-Brussel, de Vlaamse Vereniging WiskundeLeraars en de Koninklijke Belgische Vereniging voor Statistiek.



Meer info?

De communicatie over de Europese Statistiekolympiade verloopt via meerdere kanalen:

De website: www.statistiekolympiade.be

Een tweewekelijkse nieuwsbrief; inschrijven kan via de website

Een Facebook-pagina: <https://www.facebook.com/EuropeanStatisticsCompetitionBelgium>

Het team dat de organisatie coördineert, kan je ook rechtstreeks bereiken via

statolympiade@economie.fgov.be

Machine Learning voor openbare statistieken

Michael Reusens

Introductie

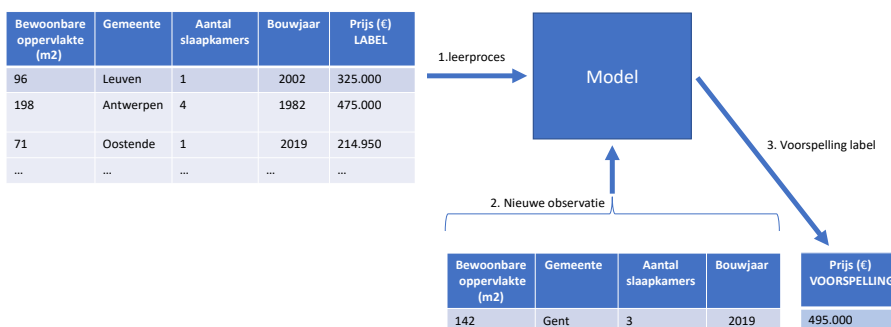
Machine Learning (ML) is een hot topic in de wereld van openbare statistiek. Dit artikel geeft een inzicht in één vorm van ML, namelijk 'supervised learning' en bespreekt de opportuniteiten en uitdagingen van ML voor openbare statistieken.

Machine Learning

ML betreft **computer algoritmen** die automatisch verbeteren door te leren uit ervaring en **data**¹. Deze algoritmen bouwen op basis van data zelf een **model** op dat steeds betere voorspellingen of beslissingen maakt. ML is een onderdeel van 'Artificiële Intelligentie'.

ML bestaat uit verschillende subtakken die opportuniteiten bieden voor openbare statistieken: supervised, unsupervised en reinforcement learning. De meest toegepaste daarvan is 'Supervised learning'. Supervised learning is een vorm, waarbij het model getraind wordt op basis van gelabelde data. Dat model wordt vervolgens ingezet om kenmerken (labels) van nieuwe observaties te voorspellen. Een voorbeeld: men wil een model ontwikkelen dat de prijs van een huis voorspelt op basis van enkele karakteristieken van dat huis, waaronder de woonbare oppervlakte, de gemeente, het aantal slaapkamers en het bouwjaar. Figuur 1 toont hoe dergelijke dataset (zie tabel) voor supervised learning er typisch uitziet.

Op basis van heel wat records wordt een model opgebouwd dat de prijs voor een nieuwe woning voorspelt op basis van de beschrijvende eigenschappen. Figuur 1 geeft tevens aan dat het model voortdurend verbeterd wordt op basis van nieuwe observaties: eens de prijs voor een nieuwe woning gekend is, kan deze nieuwe observatie automatisch gebruikt worden om het computer algoritme te verbeteren. Het model dat we hier schematisch bespreken zou bijvoorbeeld kunnen worden toegepast als automatische schatter van huisprijzen.



Figuur 1. Schematische voorstelling training en toepassing ML-model

Er bestaan veel verschillende supervised learning modellen, waaronder logistische regressie, beslissingsbomen en neurale netwerken. Deze modellen verschillen van elkaar op vlak van complexiteit van de patronen die ze kunnen leren, de mate van inzicht dat je kan hebben in een getraind model, hoeveel data ze nodig hebben om te leren en de hoeveelheid benodigde rekenkracht.

¹ Mitchell, Tom (1997). *Machine Learning*. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.

Opportunities voor openbare statistieken

Bovenstaand voorbeeld bevat enkel gestructureerde variabelen. De statistiekwereld heeft al decennia ervaring met het verwerken en analyseren van dat soort data. Openbare statistieken worden traditioneel gemaakt op basis van bevragingen en gestructureerde administratieve databronnen, zoals rijksregistergegevens. Ook ML-technieken kunnen deze data verwerken.

Hiernaast ontwikkelde de ML-wereld zeer **performante methodes voor 'niet gestructureerde data'**, zoals afbeeldingen ('computer vision'), vrije tekst ('natural language processing') en geluid ('computer audition'). Dergelijke statistieken duiken almaar vaker op:

- Het departement landbouw en visserij (Vlaamse overheid) analyseert de [groei van soorten gewassen](#) via 'computer vision'- algoritmen.
- Het Bureau of Labor Statistics (USA) categoriseert [arbeidsongevallen](#) beschreven in vrije tekst automatisch en maakt hier statistieken van.
- Het Centraal Bureau voor Statistiek (NL) voorspelt voor elke inwoner de [kans op verhuis](#) op basis van registergegevens.

ML-technieken bieden de volgende voordelen t.o.v. de 'klassieke statistiekproductie' o.b.v. administratieve data:

- ML-voorspellingen laten toe om **meer inzichten** te halen uit de **beschikbare (administratieve) data**.
- Methodes om te werken met ongestructureerde data laten toe **nieuwe fenomenen in kaart te brengen** en **bestaande statistieken te verbeteren**.

Ook t.o.v. bevragingen kunnen ML-technieken enkele voordelen bieden:

- **Verlaging van de responslast** omdat de eigenschappen niet bevraagd moeten worden, maar afgeleid worden uit beschikbare data.
- **Meer tijdige en frequente** statistieken omdat de data vaak in real time beschikbaar zijn, terwijl bevragingen periodiek gebeuren.
- **Goedkopere** statistieken tegenover de hoge kostprijs bij bevragingen in gevallen waarbij de nodige data goedkoop verkregen kunnen worden.

Uitdagingen voor openbare statistieken

De vermelde opportuniteiten van ML en nieuwe databronnen zijn afhankelijk van het succesvol aanpakken van volgende uitdagingen.

- **Bias** in de nieuwe databronnen. Surveys worden zo opgezet dat ze maximaal representatief zijn voor de bedoelde groep, bv. volwassen van 18 tot 45 jaar. Nieuwe databronnen zijn vaak een afgeleide van bestaande, niet-statistische processen/activiteiten zoals berichten op sociale media, websites van bedrijven
- **Transparantie** van de methodes. ML-modellen hebben vaak een hoog black box-gehalte, zodat het moeilijk te communiceren valt waarom een bepaald label toegekend werd aan een observatie.
- **Privacy en ethische bekommernissen** steken snel de kop op wanneer het gaat over het grootschalig analyseren van persoonsgegevens, bijvoorbeeld als bepaalde bevolkingsgroepen benadeeld kunnen worden door het model. Voor het correct toepassen van persoonsgegevens is de algemene verordening gegevensbescherming een belangrijke leidraad.
- **ML-vaardigheden**. Het optimaliseren van de opportuniteiten van deze nieuwe technieken veronderstelt het aanscherpen van de expertise binnen statistiekinstellingen.
- **Toegang** tot de nodige data. In veel gevallen zijn de benodigde data niet in handen van statistiekdiensten, maar wel bijvoorbeeld van andere entiteiten en private organisaties. Toegang krijgen tot deze data voor statistiekproductie is meestal zeer uitdagend.

Statistiek Vlaanderen data science hub

Statistiek Vlaanderen bouwt sinds 2019 aan een data science hub. Deze zet in op het ontsluiten van nieuwe databronnen en het toepassen van ML voor het maken van nieuwe en betere statistieken over Vlaanderen. Zo werkt de hub onder andere aan nieuwe statistieken rond industrie 4.0 op basis van web scraping, en over levenslang leren op basis van Twitter-data. Voorts organiseert de hub opleidingen om de ML-vaardigheden binnen Statistiek Vlaanderen te verbeteren. Die activiteiten gebeuren in samenwerking met partners binnen de Vlaamse overheid, kennisinstellingen en andere statistiekdiensten die met dezelfde uitdagingen kampen.

COLOFON

Werkten mee aan dit nummer:

Alain Henry – Expert - Federaal Planbureau

Saskia Hin – Expert - BISA

Martine Lefèvre – Attachée scientifique - IWEPS

Philippe Mauroy – Directeur generaal a.i. - Statbel

Michael Reusens – Data Science Coördinator - Statistiek Vlaanderen

Wendy Schelfaut – Diensthoud « Communicatie & Verspreiding » - Statbel

Het IIS is het resultaat van een samenwerkingakkoord uit 2014.

Het IIS moet toelaten om de kwaliteit van openbare statistieken, ter ondersteuning van het openbaar beleid, continu te bevorderen.



www.iis-statistics.be