

Methodische bijlage: inzet op maatschappelijke thema's

In opdracht van het Ministerie van Economische Zaken en Klimaat (EZK) heeft RVO in beeld gebracht in hoeverre projecten die ondersteund zijn vanuit de innovatie-instrumenten bijdragen aan vier maatschappelijke thema's: 1) Energietransitie & Duurzaamheid; 2) Landbouw, Water & Voedsel; 3) Gezondheid & Zorg, en 4) Veiligheid. Dit document zet de methode uiteen hoe tot dit resultaat gekomen is.

Let op: dit document is een bijlage van de website

<https://www.bedrijvenbeleidinbeeld.nl/bouwstenen-bedrijvenbeleid/innovatie>, van het ministerie van EZK. Ook wordt een deel van de resultaten gebruikt op <https://www.bedrijvenbeleidinbeeld.nl/bouwstenen-bedrijvenbeleid/missiegedreven-innovatiebeleid/maatschappelijke-themas>.

Methodische achtergrond

Scope

RVO heeft de analyse uitgevoerd over de volgende innovatieprogramma's en -regelingen van het ministerie van EZK:

- de Wet Bevordering Speur- en Ontwikkelingswerk (WBSO);
- inzetprojecten binnen de regeling PPS-toeslag Onderzoek en Innovatie (waarbij PPS staat voor privaat-publieke samenwerking);
- Mkb-innovatiestimulering Regio en Topsectoren (MIT)¹;
- Vroegefasefinanciering (VFF)²;
- SEED Capital;
- Small Business Innovation Research (SBIR);
- Innovatiekrediet;
- Eureka (inclusief JTI's);
- Eurostars;
- Globalstars;
- het Nationaal Groeifonds; en
- subsidies gericht op energie-innovatie binnen de Topsector Energie (EI)

Toedeling

RVO heeft de in 2020 gecommitteerde projecten³ binnen genoemde programma's en regelingen op basis van projectomschrijvingen toebedeeld aan een of meer van de vier maatschappelijke thema's: 1) Energietransitie & Duurzaamheid; 2) Landbouw, Water & Voedsel; 3) Gezondheid & Zorg; en 4) Veiligheid. Met uitzondering van de WBSO zijn de projecten handmatig geclassificeerd op basis van een uitgebreide omschrijving van de vier maatschappelijke thema's en de missies die binnen deze thema's vallen (zoals uiteengezet in de kamerbrief '[Missiegedreven topsectoren en innovatiebeleid](#)' van 26 april 2019).

Voor alle instrumenten zijn de uiteindelijke labels toegekend door RVO-adviseurs. Hierbij zijn projecten handmatig gecategoriseerd naar de vier maatschappelijke thema's op basis

¹ Betreft de MIT R&D-samenwerkingsprojecten en Haalbaarheidsstudies. De MIT Kennisvouchers zijn niet gelabeld.

² Voor zover in uitvoering bij RVO.

³ Met uitzondering van SEED Capital en (deels) SBIR, waar is gelabeld op respectievelijk fonds- en tenderniveau.

van de projectomschrijvingen. Voor de PPS-toeslagregeling hebben de Topconsortia voor Kennis en Innovatie (TKI's) zelf een voorzet gedaan voor een categorisatie, en voor de MIT-regeling is dit door indieners gedaan. Om tot een consistente beoordeling te komen, is de beoordeling van de PPS-toeslag-inzetprojecten en MIT-projecten gecheckt en waar nodig aangepast door adviseurs van RVO. Voor alle instrumenten geldt dat eventuele twijfelgevallen binnen het team van projectbeoordelaars zijn besproken en onderling zijn afgestemd.

Voor de instrumenten binnen het missiegedreven innovatiebeleid, de PPS-regeling, de MIT en de SBIR-regeling, is er gelabeld op zes mogelijke Kennis- en Innovatie Agenda's (KIA's) en onderliggende Meerjarige Missiegedreven Innovatieprogramma's (MMIP's). De KIA's bestaan uit vier thematische KIA's die overeenkomen met de maatschappelijke thema's, en daarnaast de KIA Sleuteltechnologieën en de KIA Maatschappelijk Verdienvermogen. Deze laatste twee zijn dwarsdoorsnijdend en van belang voor het zowel realiseren van innovatie als het creëren van impact. Elke KIA heeft een aantal missies. Enkele grafieken in de analyse tonen, voor de PPS-toeslagregeling, per maatschappelijk thema het aantal projecten dat aan deze missie bijdraagt, en de totale omvang van de PPS-toeslag van die projecten. Voor de toedeling naar missies is gebruik gemaakt van de toedeling naar de MMIP's. Deze zijn op dezelfde wijze door RVO gecontroleerd als de toedeling naar de KIA's.

Voor de WBSO is de classificatie naar maatschappelijke thema's uitgevoerd door middel van een gewogen-trefwoorden-analyse. Hoe dit precies werkt wordt hieronder bij de sectie 'Uitgelicht' toegelicht. Gezien het grote aantal projecten (139.000 niet-unieke projecten in 2021) is het niet haalbaar om met beperkte middelen alle WBSO-projecten te lezen en handmatig te classificeren.

Bedragen

In de analyses wordt afwisselend gesproken over middelen, budgetten en committeringen. Hier wordt in principe altijd het gecommiteerde bedrag bedoeld voor de projecten, met de volgende toelichting of uitzonderingen:

- Bij de MIT betreft het enkel de landelijke committeringen, vanuit het EZK-budget, en niet van de regio's.
- Bij de PPS-toeslag betreft het de door de TKI's gerapporteerde ingezette PPS-toeslag.
- Bij de SBIR betreft het budgetten van SBIR-calls. Het betreft ook SBIR's waar het ministerie van EZK geen opdrachtgever van is. Het ministerie van EZK is wel verantwoordelijk voor het SBIR-instrument als geheel.
- Bij de SEED Capital-regeling betreft het een participatiebedrag op fondsniveau.

Uitgelicht: gewogen-trefwoorden-analyse WBSO

Achtergrond

Via de WBSO stimuleert het ministerie van EZK ondernemers om meer te investeren in spur- en ontwikkelingswerk (S&O), ook wel bekend als Research en Development (R&D). Ondernemers die gebruik maken van de WBSO kunnen hiermee hun R&D-uitgaven verlagen. Jaarlijks worden door RVO ongeveer 130.000 WBSO-projecten (niet

uniek) toegekend. Van deze projecten zijn alle omschrijvingen opgeslagen in een database van RVO⁴.

Het ministerie van EZK en andere stakeholders vragen regelmatig om verschillende thema's in kaart te brengen en inzicht te geven in welke mate deze worden ondersteund binnen de WBSO. In zo'n thema-analyse wordt onderzocht hoeveel WBSO-projecten betrekking hebben op één of meerdere specifieke thema's.

Methodie

Voor deze analyses wordt gebruik gemaakt van een zogenaamde 'gewogen-zoektermenmethodiek'. De gewogen-zoektermenmethodiek is door RVO ontwikkeld en kan worden gebruikt om WBSO-projecten te classificeren. Bij deze methode worden de projecten niet handmatig gelezen en geclassificeerd, maar wordt er in de projectomschrijvingen gezocht naar gerelateerde zoektermen, waarbij er een bepaalde geweging aan deze zoektermen wordt gegeven.

De methodiek bestaat uit vier primaire stappen die in figuur 1 worden weergegeven.



Figuur 1 - Schematische weergave van de gewogen zoektermenmethodiek van RVO binnen de WBSO, een iteratief proces

Zoektermen & weging

Om thema's af te bakenen worden voorafgaand aan de analyse voor elk thema een lijst met zoektermen opgesteld door adviseurs van RVO, met per zoekterm een wegingsfactor. Dit is een zeer belangrijke stap in het totale proces die in grote mate de kwaliteit van de uitkomst bepaalt. Deze stap wordt dan ook in samenwerking met een vakinhoudelijk-expert op het gebied van het gevraagde thema uitgevoerd, daardoor kan zoveel mogelijk naar de juiste (technische) vaktermen worden gezocht. Ieder trefwoord krijgt een geweging op basis van relevantie in relatie tot het gezochte thema. Zoektermen die kenmerkender zijn voor het thema dat wordt gezocht, krijgen een hogere waarde dan meer algemene woorden die wellicht raakvlak hebben met het gezochte thema, maar mogelijk ook in een andere context een andere betekenis kunnen hebben. In de lijst met zoektermen wordt rekening gehouden met linguïstiek (taal, spellingsvarianten en lemmatisering - het gebruik van woordstammen). Zoektermen die aangrenzende thema's kunnen identificeren die juist niet meegenomen mogen worden, kunnen een (sterk) negatieve geweging meekrijgen om bepaalde projecten uit te sluiten.

⁴ Deze informatie wordt gebruikt om de WBSO uit te voeren en valt onder de algemene wet inzake Rijksbelastingen (AWR).

Ter illustratie: Bij het zoeken naar het thema Cybersecurity wordt de zoekterm 'virus' gebruikt. Door te zoeken naar deze term zullen ook veel medische projecten worden gevonden die niet geclassificeerd mogen worden tot het thema Cybersecurity. Door zoektermen uit de medische technologie toe te voegen met een sterk negatieve weging, wordt er onderscheid gemaakt tussen het bestrijden van een 'medisch virus' en een 'ICT-virus'.

Doorzoeken & 'scoren'

Nadat de zoektermenlijst en wegingen zijn bepaald, worden met een query alle WBSO-projecten in de statistiek-database doorzocht op deze termen. In de query kan worden aangegeven welke WBSO-jaren worden meegenomen en welke projectvelden worden doorzocht. Komt een zoekterm voor in één van de projectvelden, dan krijgt het project de weging mee. Komt dezelfde term in hetzelfde project nogmaals voor, dan telt de weging slechts één keer mee. Het script telt vervolgens alle wegingen per project op en rangschikt alle WBSO-projecten met de gevonden trefwoorden op projectscore van hoog naar laag.

Valideren uitkomsten

Om de uitkomsten te valideren, wordt door middel van steekproeven bij verschillende projectscores de inhoud van de WBSO-projecten getoetst of de gevonden WBSO-projecten tot het gezochte thema behoren en welke niet. De validatie wordt uitgevoerd door een vakinhoudelijk expert. De steekproeven vinden daarbij plaats van een hoge projectscore naar een steeds lagere projectscore. Op deze manier wordt gezocht naar een zekere grenswaarde, waar tenminste $2/3$ van het aantal WBSO-projecten uit de steekproef tot het thema behoren. Behoren minder dan $2/3$ van het aantal projecten met een bepaalde projectscore tot het thema, dan vallen deze buiten de analyseresultaten. Dit is geen wetenschappelijk vastgesteld aandeel, maar ervaring leert inmiddels dat het een grens is die de thema's over het algemeen correct schat in aantal, mits de trefwoordenlijst zorgvuldig is samengesteld en wanneer er intensief contact is met een vakinhoudelijk-expert. Indien er bijvoorbeeld van de WBSO-projecten met een projectscore van 11 meer dan $2/3$ tot het gezochte thema behoort, maar bij een projectscore van 10 nog maar $1/3$ tot het thema behoort dan wordt de grenswaarde van de projectscore vastgesteld op 11. Alle WBSO-projecten met een projectscore van 11 en hoger worden dan meegeteld in de analyseresultaten, de projecten met een lagere projectscore niet. Deze methode wordt toegepast op uiteenlopende thema-analyses binnen de WBSO. Voor elke nieuwe thema-analyse wordt dit proces herhaald en heeft de grenswaarde per thema een andere waarde.

Op basis van verschillende steekproeven per projectscore wordt bepaald of projecten correct zijn geclassificeerd. Doordat het steekproeven betreft, worden niet alle projecten op deze manier handmatig gevalideerd.

Een nadeel van deze methode is dat de er geen exacte getallen uitgerekend kunnen worden voor de gangbare indicatoren die iets zeggen over de kwaliteit van het model, te weten de 'Precision & Recall'. Een voordeel is dat er zonder een trainingsset toch een inschatting gemaakt kan worden van een specifiek thema.

Door de methode, zoektermenlijst, weging én de grenswaarde per thema identiek te houden kan er een vergelijking worden gemaakt tussen verschillende jaargangen binnen de statistiekdatabase. Wordt ervoor gekozen om op een bepaald moment de methode aan te passen (door bijvoorbeeld zoektermenlijsten aan te passen of andere

projectvelden te doorzoeken), dan zullen reeds gepubliceerde cijfers gereviseerd moeten worden, omdat de resultaten niet meer vergelijkbaar zijn met eerdere publicaties.