



van Jaap Jansen en Arthur Provoost, Centraal Bureau voor de Statistiek
onderwerp Resultaten proefkoppeling data TenderNed
datum 13 april 2022

TenderNed beschikt over microdata van een groot deel van de aanbestedingen die door Nederlandse overheden worden gedaan. Omdat deze data potentieel interessante beleidsinformatie bevatten, die door koppeling met het ABR verder verrijkt kunnen worden, is een data assessment uitgevoerd. Daarin zijn een proefkoppeling en enkele eenvoudige analyses uitgevoerd. Deze notitie bespreekt achtereenvolgens de resultaten van de proefkoppeling en van de analyses die zijn uitgevoerd. Ten slotte wordt een conclusie getrokken over de bruikbaarheid van deze data voor beleidsanalyses.

Resultaten koppeling

Het CBS heeft een (proef)dataset ontvangen van TenderNed. Na het uitvoeren van een aantal kwaliteitscontroles op de aangeleverde data, is de eerste stap in het koppelingsproces met het ABR uitgevoerd. Het resultaat van deze proefkoppeling laat zien welk aandeel van de TenderNed gegevens gekoppeld kan worden aan het ABR. De proefkoppeling heeft ook als doel om tijdig een keuze te maken in de variabelen en inzicht te krijgen in wat die variabelen precies voor informatie bevatten. Hieronder volgen de bevindingen met betrekking tot de proefkoppeling.

Databestand van TenderNed

De levering van TenderNed bestaat uit drie bestanden:

- Aanbestedingen. Hierin worden de kenmerken van de aanbestedingen zelf en de betreffende aanbestedende dienst weergegeven.
- Gunningen. Hierin wordt van iedere aanbesteding per perceel aangegeven wanneer het gegund is en wat de waarde¹ van de gunning is. Dit bestand geeft ook de datum waarop de gunning is verleend. Dus dit bestand wordt gebruikt om te bepalen in welke periode een gunning valt.
- Ondernemingen. Hierin staan kenmerken van de bedrijven die een aanbesteding gegund hebben gekregen. Dit bestand is gebruikt om de koppeling met het ABR tot stand te brengen.

Het controleren en koppelklaar maken van deze bestanden is gebeurd in R. Tabel 1 geeft een korte beschrijving van de bestanden.

¹ TenderNed gaf aan dat de waarde van de gunning vaak niet goed ingevuld is, dus deze variabele is verder niet gebruikt.



Tabel 1. Beschrijving van de door TenderNed verstrekte bestanden.

	Aanbestedingen	Gunningen	Ondernemingen
Aantal records	37.997	54.195	73.993
Aantal records TenderNed	18.071	n.v.t.	26.653
Aantal records Tsender	19.926	n.v.t.	47.340
Unieke aanbestedingen	37.997	37.997	37.997
Unieke aanbestedende partijen (o.b.v. KvK-nummer)	1.597	n.v.t.	n.v.t.
Unieke ondernemingen (o.b.v. KvK-nummer)	n.v.t.	n.v.t.	9.995
Records met kloppende KvK-nummers (voor bewerking)	34.436	n.v.t.	38.813
Records met foutieve of missende KvK-nummers (voor bewerking)	3.561	n.v.t.	35.180
Records met kloppende KvK-nummers (na bewerking)	37.950	n.v.t.	38.900
Records met foutieve of missende KvK-nummers (na bewerking)	47	n.v.t.	35.093

Het KvK-nummer is de belangrijkste koppelsleutel met het ABR. De volledigheid en kwaliteit van de KvK-nummers beïnvloeden het koppelingsrendement met het ABR. Het CBS voert daarom een aantal kwaliteitscontroles op het KvK-nummer uit.

Foutieve KvK-nummers zijn gecorrigeerd. Wanneer de lengte van het KvK-nummer geen 8 is, dan is er een voorloopnul toegevoegd wanneer de lengte 7 is. De overige lengtes zijn op missend (NA) gezet. Ook onlogische KvK-nummers, zoals 8 nullen of een non-numeriek KvK-nummer is op missend (NA) gezet. Bij Aanbestedingen was het mogelijk om relatief veel KvK-nummers van aanbestedende diensten te corrigeren, doordat de lengte voor bewerking 3.515 keer niet 8 was. Door een voorloopnul toe te voegen konden op één na alle KvK-nummers met een foutieve lengte kloppend gemaakt worden. Bij Ondernemingen was de lengte voor bewerking slechts 311 keer niet 8, waarbij ook lengtes groter dan 8 voorkwamen, waardoor er minder ruimte was voor verbetering van KvK-nummers van gegunde bedrijven.

Bij bijna de helft van de records in het bestand Ondernemingen ontbreekt het KvK-nummer of is dit foutief. Sommige KvK-nummers kunnen makkelijk correct gemaakt worden door hier een voorloopnul toe te voegen. Dit zorgt ervoor dat het aantal records met een ontbrekend of foutief KvK-nummer uitkomt op 35.093. Foutieve of missende KvK-nummers komen voornamelijk voor bij records die uit andere bronnen komen dan TenderNed, namelijk de Tsenders. Bij records met als bron TenderNed is het aantal foutieve of missende KvK-nummers 237 (0,9%) en bij Tsenders is dit 34.856 (73,6%). De foutieve of missende KvK-nummers afkomstig van TenderNed zijn in alle gevallen KvK-nummers van buitenlandse ondernemingen, die dus sowieso nooit gekoppeld kunnen worden met het ABR.



Bij het controleren komt verder het volgende naar voren:

- De variabele `ab_kenmerk` ziet er goed uit bij alle drie de datasets. Dit is belangrijk, omdat deze variabele fungeert als koppelsleutel tussen de drie datasets.
- De variabele `pc_nummer` ontbreekt in het bestand Aanbestedingen. Dit is logisch, omdat dit bestand over hele aanbestedingen gaat en niet over specifieke percelen. Daarnaast ontbreekt het `pc_nummer` 37.173 keer in het bestand Gunningen en 14.969 keer in het bestand Ondernemingen. Dit is logisch, omdat er niet altijd sprake is van verschillende percelen.
- De verschillende datums van aankondigingen, gunningen en publicaties (in het bestand Aanbestedingen en in het bestand Gunningen) zijn ook met elkaar vergeleken. Hieruit komt naar voren dat onmogelijke uitkomsten zelden voorkomen². De onmogelijke uitkomsten die er wel zijn, worden gezien als gebruikelijke ruis.

Proefkoppeling met het ABR

Het koppelingsproces met het ABR bestaat uit een aantal stappen. Voor de proefkoppeling voeren we alleen de eerste stap uit, omdat dit al een goede indicatie geeft van het koppelingsrendement. De proefkoppeling is uitgevoerd met het ABR van 2020.

Er zijn verschillende koppelingen gemaakt:

1. Tussen het bestand Aanbestedingen en het ABR.
2. Tussen het bestand Ondernemingen en het ABR.
3. Tussen het gekoppelde bestand Ondernemingen-ABR en het bestand Aanbestedingen.

Bij de koppeling tussen het bestand Aanbestedingen en het ABR komt het volgende naar voren:

- Er is een hoog koppelingsrendement. 37.397 van de 37.950 originele records koppelen (98,5%). Het gaat hier om de aanbestedende partijen. Het aantal records is wanneer alle records worden samengevoegd hoger dan in het originele bestand Aanbestedingen, doordat er KvK-nummers voorkomen met meerdere bedrijfseenheden (BEIDs)³.
- De gemeenschappelijke woordenlijst overheidsopdrachten (CPV) is een centraal classificatiesysteem voor overheidsopdrachten dat ervoor zorgt dat aanbestedende diensten en entiteiten dezelfde referenties gebruiken om hun opdracht te beschrijven. De CPV-codes zijn op basis van `ab_cpv_code` en de SBI-codes zijn op basis van `SbiGecoördineerd` gehercodeerd tot hun hoofdcategorieën. De combinaties SBI, CPV, omschrijving van de CPV-code en naam aanbesteder zien er logisch uit.

Bij de koppeling tussen het bestand Ondernemingen en het ABR komt het volgende naar voren:

- Er is een laag koppelingsrendement. 38.144 van de 74.111 records koppelen (51,5%). Dit is logisch, want het KvK-nummer ontbreekt in bijna de helft van de gevallen in het bestand Ondernemingen doordat Tsenders deze niet (goed) doorgeven.

² Bij 1 van de 37.997 aanbestedingen valt de aankondiging later dan de gunning. Bij 71 van de 54.195 gunningen vindt de gunning zelf later plaats dan de publicatie hiervan.

³ 32.183 records koppelen aan 1 BEID en 15.879 records koppelen aan meer dan 1 BEID.



- Bij een nieuwe dataset Ondernemingen zonder de records waarbij het KvK-nummer ontbreekt, is er sprake van een hoog koppelingsrendement. 37.996 van de 38.900 originele records koppelen (97,7%). Het aantal records is wanneer alle records worden samengevoegd hoger dan in het originele bestand Ondernemingen waarbij het KvK-nummer klopte, doordat er KvK-nummers voorkomen met meerdere BEIDs⁴. Dat sommige KvK-nummers niet koppelen, kan veroorzaakt worden doordat bijvoorbeeld bedrijven die een aanbesteding kregen in 2017, niet meer voorkomen in het ABR van 2020. Zoals verwacht, stijgt het koppelingsrendement per jaar.

Tabel 2. Koppelingsrendement.

Jaar	Koppelingsrendement in %
2016	94,7
2017	95,9
2018	97,4
2019	98,5
2020	99,1
2021	99,2

- Voor bewerking hebben 37.955 records een andere naam in het ABR. Alle namen in hoofdletters zetten leidt ertoe dat nog 29.611 records een andere naam hebben. De resterende verschillen zien er echter niet problematisch uit. In het bestand Ondernemingen is de naam bijvoorbeeld (fictief voorbeeld) "Autobedrijf Nederland" en in het ABR is de naam "Autobedrijf Nederland N.V.". Soms komt het wel voor dat namen helemaal verschillen. Uit de handmatige controles blijkt dat er dan vaak sprake is van dochterondernemingen.

Bij de koppeling tussen het gekoppelde bestand Ondernemingen-ABR en het bestand Aanbestedingen komt het volgende naar voren:

- De combinatie SBI, CPV, omschrijving van de CPV-code en de naam van de onderneming zijn in de meeste gevallen logisch.
- De handmatige controles zien er goed uit.

Analyses

Na de koppeling van de data van TenderNed met het ABR is onderzocht of enkele eenvoudige analyses uitgevoerd kunnen worden op deze data. Eerst is een kwaliteitscheck gedaan door de CPV-codes te vergelijken met de SBI-codes van de bedrijven waaraan een aanbesteding is gegund. Vervolgens is een overzicht gemaakt van de bedrijfstakken waar aanbestedingen door EZK terecht zijn gekomen. Als laatste is een tabel gemaakt met het aantal aanbestedingen per bedrijfstak per jaar.

⁴ 37.898 records koppelen aan 1 BEID en 216 records koppelen aan meer dan 1 BEID.



Vergelijking SBI en CPV

Om de kwaliteit van de gegevens te beoordelen zijn na koppeling de CPV-codes uit het bestand Aanbestedingen en de SBI-codes uit het ABR vergeleken. Per record is onderzocht of de CPV-code logisch overeenkwam met de SBI-code. Tabel 3 geeft per SBI-lettercode weer welk percentage bijbehorende CPV-codes logisch, twijfelachtig of onlogisch is. Hieronder volgen wat voorbeelden van wat als logisch, twijfelachtig en onlogisch beschouwd wordt⁵:

- **Logisch:** Een voorbeeld waarbij de SBI-code logisch samenhangt met de CPV-code is SBI-code F (Bouwnijverheid) waarvan 2851 bedrijven een opdracht hebben gekregen met een CPV-code die begint met 45 (Bouwwerkzaamheden).
- **Twijfelachtig:** Een voorbeeld waarbij er twijfel is of de SBI-code logisch samenhangt met de CPV-code is SBI-code Q (Gezondheids- en welzijnszorg) waarvan 150 bedrijven een opdracht hebben gekregen met een CPV-code die begint met 79 (Zakelijke dienstverlening: juridisch, marketing, consulting, drukkerij en beveiliging).
- **Onlogisch:** Een voorbeeld waarbij de SBI-code onlogisch samenhangt met de CPV-code is SBI-code I (Logies-, maaltijd- en drankverstrekking) waarvan 12 bedrijven een opdracht hebben gekregen met een CPV-code die begint met 42 (Bedrijfsmachines). Bij deze SBI-code is het percentage onlogische CPV-codes het hoogst (10,6%). Dit komt doordat een aantal CPV-codes, zoals 'Bedrijfsmachines' en 'Zakelijke dienstverlening: juridisch, marketing, consulting, drukkerij en beveiliging' onlogisch zijn. Dit aantal is klein, maar doordat het totale aantal observaties in deze SBI-code ook relatief klein is, heeft zo'n klein aantal wel relatief veel effect op het percentage onlogische CPV-codes.

Tabel 3. Verdeling logica CPV-codes, naar SBI-lettercode.

SBI	CPV logisch	CPV twijfelgeval	CPV onlogisch
A	81,6%	14,1%	4,3%
B	83,3%	16,7%	0,0%
C	92,6%	6,5%	0,9%
D	90,8%	9,2%	0,0%
E	86,4%	11,6%	2,0%
F	98,9%	0,8%	0,3%
G	92,0%	6,9%	1,0%
H	94,6%	1,9%	3,5%
I	86,6%	2,8%	10,6%
J	96,2%	2,8%	0,9%
K	96,7%	0,4%	2,9%
L	81,3%	15,3%	3,5%
M	98,5%	1,5%	0,0%

⁵ De CPV-codes hebben een hiërarchische structuur. Hier is alleen gekeken naar de hoofcategorie. Als binnen een hoofcategorie naar een subcategorie gekeken wordt, kan iets wat op het eerste gezicht onlogisch toch logisch blijken te zijn. Daarnaast zijn alleen de hoofd-cpv-codes van een aanbesteding meegegeven, maar een aanbesteding kan meerdere cpv-codes hebben. Dat zou ook nog een deel kunnen verklaren.



N	96,9%	3,1%	0,0%
O	89,3%	7,1%	3,6%
P	90,5%	7,1%	2,4%
Q	95,4%	4,3%	0,3%
R	81,1%	14,2%	4,7%
S	82,6%	12,0%	5,4%

Door EZK gegunde aanbestedingen

Er is ook onderzocht waar de gegunde opdrachten van het Ministerie van EZK terecht komen. Daarbij is gekeken naar de bedrijfstak op SBI-letterniveau van bedrijven die een gunning van het Ministerie van EZK hebben gekregen. Tabel 4 geeft voor de hele dataset aan bij welke bedrijfstakken gunningen door EZK terecht zijn gekomen. De volgende tabel geeft per jaar (2016 – 2021) aan bij welke bedrijfstakken gunningen terecht zijn gekomen. Op deze manier is het mogelijk om hier eventuele trends en ontwikkelingen in te signaleren.

Tabel 4. Gunningen door EZK, naar bedrijfstak, 2016-2021

SBI	Aantal	Percentage
A. Landbouw, bosbouw en visserij	6	0.9%
C. Industrie	3	0.5%
E. Winning en distributie van water; afval- en afvalwaterbeheer en sanering	6	0.9%
F. Bouwnijverheid	2	0.3%
G. Groot- en detailhandel; reparatie van autos	25	3.8%
H. Vervoer en opslag	7	1.1%
J. Informatie en communicatie	168	25.4%
K. Financiële instellingen	16	2.4%
L. Verhuur van en handel in onroerend goed	2	0.3%
M. Advisering, onderzoek en overige specialistische zakelijke dienstverlening	226	34.2%
N. Verhuur van roerende goederen en overige zakelijke dienstverlening	158	23.9%
P. Onderwijs	12	1.8%
Q. Gezondheids- en welzijnszorg	7	1.1%
R. Cultuur, sport en recreatie	2	0.3%
S. Overige dienstverlening	21	3.2%

Aanbestedingen per jaar

De tabellen 5a en 5b geven een overzicht van het aantal aanbestedingen per bedrijfstak (SBI-letter) en jaar in aantallen en in percentages.



Tabel 5a. Aantal aanbestedingen per jaar, naar bedrijfstak.

SBI	2016	2017	2018	2019	2020	2021
	Aantal	Aantal	Aantal	Aantal	Aantal	Aantal
Advisering, onderzoek en overige specialistische zakelijke dienstverlening	125	334	454	479	495	561
Bouwnijverheid	146	494	576	642	707	618
Cultuur, sport en recreatie	2	8	7	13	9	11
Financiële instellingen	45	67	135	167	180	129
Gezondheids- en welzijnszorg	35	103	107	119	106	164
Groot- en detailhandel; reparatie van autos	117	515	585	533	604	569
Industrie	40	165	201	220	191	233
Informatie en communicatie	84	203	349	297	365	428
Landbouw, bosbouw en visserij	5	43	48	56	74	83
Logies-, maaltijd- en drankverstreking	7	16	22	28	27	17
Onderwijs	23	18	45	32	27	82
Openbaar bestuur, overheidsdiensten en verplichte sociale verzekeringen	0	4	0	1	0	1
Overige dienstverlening	2	24	18	16	26	34
Productie en distributie van en handel in elektriciteit, aardgas, stoom en gekoelde lucht	2	33	37	28	32	18
Verhuur van en handel in onroerend goed	2	7	7	6	17	7
Verhuur van roerende goederen en overige zakelijke dienstverlening	88	428	620	573	612	575
Vervoer en opslag	45	154	107	112	106	118
Winning en distributie van water; afval- en afvalwaterbeheer en sanering	56	103	137	112	128	119
Winning van delfstoffen	1	2	2	1	4	0

Tabel 5b. Percentage aanbestedingen per jaar, naar bedrijfstak.

SBI	2016	2017	2018	2019	2020	2021
	Percentage	Percentage	Percentage	Percentage	Percentage	Percentage
Advisering, onderzoek en overige specialistische zakelijke dienstverlening	15,2%	12,3%	13,1%	13,9%	13,3%	14,9%
Bouwnijverheid	17,7%	18,2%	16,7%	18,7%	19,1%	16,4%
Cultuur, sport en recreatie	0,2%	0,3%	0,2%	0,4%	0,2%	0,3%
Financiële instellingen	5,5%	2,5%	3,9%	4,9%	4,9%	3,4%
Gezondheids- en welzijnszorg	4,2%	3,8%	3,1%	3,5%	2,9%	4,4%
Groot- en detailhandel; reparatie van autos	14,2%	18,9%	16,9%	15,5%	16,3%	15,1%
Industrie	4,8%	6,1%	5,8%	6,4%	5,1%	6,2%
Informatie en communicatie	10,2%	7,5%	10,1%	8,6%	9,8%	11,4%
Landbouw, bosbouw en visserij	0,6%	1,6%	1,4%	1,6%	2,0%	2,2%
Logies-, maaltijd- en drankverstreking	0,8%	0,6%	0,6%	0,8%	0,7%	0,5%
Onderwijs	2,8%	0,7%	1,3%	0,9%	0,7%	2,2%
Openbaar bestuur, overheidsdiensten en verplichte sociale verzekeringen	0,0%	0,1%	0,0%	0,0%	0,0%	0,0%
Overige dienstverlening	0,2%	0,9%	0,5%	0,5%	0,7%	0,9%
Productie en distributie van en handel in elektriciteit, aardgas, stoom en gekoelde lucht	0,2%	1,2%	1,1%	0,8%	0,9%	0,5%
Verhuur van en handel in onroerend goed	0,2%	0,3%	0,2%	0,2%	0,5%	0,2%
Verhuur van roerende goederen en overige zakelijke dienstverlening	10,7%	15,7%	17,9%	16,7%	16,5%	15,3%
Vervoer en opslag	5,5%	5,7%	3,1%	3,3%	2,9%	3,1%
Winning en distributie van water; afval- en afvalwaterbeheer en sanering	6,8%	3,8%	4,0%	3,3%	3,5%	3,2%
Winning van delfstoffen	0,1%	0,1%	0,1%	0,0%	0,1%	0,0%

Conclusie en vervolgonderzoek

Het is goed mogelijk om een koppeling te maken tussen de data van TenderNed en het ABR. Een nadeel hierbij is wel dat bij bijna de helft van de records de KvK-nummers ontbreken in het bestand met gegunde ondernemingen. De records met ontbrekende KvK-nummers komen van de Tsenders. Dit is dus niet iets wat door TenderNed verbeterd kan worden. Dat zorgt ervoor dat de populatie waarover geanalyseerd kan worden, versmald wordt tot de bedrijven die een gunning hebben gekregen die door TenderNed zelf geregistreerd is. Of dit representatief is voor alle overheidsaanbestedingen in Nederland zal moeten blijken uit vervolgonderzoek.

De KvK-nummers die niet ontbreken hebben over het algemeen een hoge kwaliteit. In 97 tot 99% van de gevallen waar een zinvol KvK-nummer beschikbaar is kan worden gekoppeld aan het ABR. Bij vergelijking tussen SBI-codes en CPV-codes blijkt ook dat grofweg 90 procent van de koppelingen er logisch uitzien. De proefkoppeling met het ABR van de bestanden van TenderNed kan daarom ook gezien worden als succesvol.

Het is daarbij mogelijk om eenvoudige analyses te doen op de data, zoals analyseren waar de gegunde opdrachten van het Ministerie van EZK terechtkomen.



De volgende zaken kunnen in de toekomst nog nader worden onderzocht:

- Zoals in het projectvoorstel stond, is onderzocht waar gegunde opdrachten met een innovatiedoel terecht komen. In de huidige dataset zijn daar echter geen geschikte variabelen voor aanwezig. Binnen TenderNed wordt op dit moment onderzocht in hoeverre er een variabele te ontwikkelen is, die hier wel inzicht in geeft. Wellicht is het in de nabije toekomst mogelijk om specifiek innovatieve aanbestedingen te analyseren.
- Er zijn ook KvK-nummers die aan meerdere bedrijfseenheden (BEIDs) koppelen. Deze BEIDs kunnen worden geïdentificeerd en vervolgens kan er een logische keuze worden gemaakt aan welke BEID een KvK-nummer moet koppelen o.b.v. bepaalde beslisregels. Voor dit data assessment is een willekeurige keuze voor een BEID gemaakt. Wanneer dit in productie genomen wordt, moet de meest passende BEID handmatig gekozen worden. In de huidige dataset is dit 216 keer het geval (0.6%).
- Enkel de openbare data zijn aangeleverd door TenderNed. Bij een vervolg zou TenderNed eventueel ook data over onderhandse aanbestedingen kunnen leveren.