



Centraal Bureau
voor de Statistiek

Rapport

Datakoppeling ETS-bedrijven

Cor Graveland

Albert Pieters

Kathleen Geertjes

CBS Den Haag
Henri Faasdreef 312
2492 JP Den Haag
Postbus 24500
2490 HA Den Haag
+31 70 337 38 00
www.cbs.nl

projectnummer 15215
ENR - Arbeid, Milieurekeningen en Regio
30 November 2016

kennisgeving De in dit rapport weergegeven opvattingen zijn die van de auteur(s) en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

Inhoudsopgave

Voorwoord	5
Samenvatting	6
1. Inleiding	10
1.1 Introductie en achtergrond project	10
1.2 Doel van het project	10
1.3 Producten en rapportage	11
2. Populatie ETS bedrijven	13
2.1 Populatie ETS-bedrijven en regels voor deelname EU ETS	13
2.2 Monitoring ETS-bedrijven populatie door de NEa	13
2.3 Populatie ETS-bedrijven in het CBS-bedrijvenregister ABR	14
3. Databronnen en organisatie	17
3.1 De databronnen en statistieken voor koppeling	17
3.2 De organisatie van de data	19
4. Methode datakoppeling	21
4.1 Processtappen voor realisatie van de koppeling	21
4.2 Koppelingssystematiek ETS-populatie aan het ABR	22
4.3 Mogelijkheden voor verdere verbetering en uitbreiding van de koppeling	27
5. Technisch resultaat van de koppeling	28
5.1 Methode en koppelingsresultaat	28
5.2 Koppelresultaten	29
6. Gebruik, beheer en onderhoud data-infrastructuur	31
6.1 Data-infrastructuur	31
6.2 Gebruik	31
6.3 Beheer en onderhoud	32
7. Conclusies en aanbevelingen	34
7.1 Conclusies	34
7.2 Aanbevelingen	37
Bronnen	39
Bijlagen	40
Bijlage I: Samenvatting pilot koppeling NEa- en CBS-eenheden 2015	40
Bijlage II: Praktische aspecten van data-organisatie en data-structurering	41
Bijlage III: Praktische illustratie koppelproblematiek	43

Bijlage IV: Mogelijke onderzoeksvragen vervolgonderzoek	44
Bijlage V: Project bijdragen	46

Voorwoord

Dit rapport beschrijft de uitkomsten van het project datakoppeling van de Emission Trading System - bedrijven (ETS - bedrijven) populatie dat het Centraal Bureau voor de Statistiek (CBS) heeft uitgevoerd. Van de bedrijven die deelnemen aan het EU ETS wordt de emissiehandel gemonitord door de Nederlandse Emissieautoriteit (NEa). De informatie die NEa over deze bedrijven heeft zijn in dit project gekoppeld aan de CBS-data-infrastructuur.

Het project is ontstaan uit de samenwerking en uitwisseling die de afgelopen acht jaar stapsgewijs is ontwikkeld en gegroeid tussen CBS en de NEa op het terrein van CO₂-emissierechten. De andere aanleiding van dit onderzoek is het 'Anders omgaan met data programma' van het ministerie van Infrastructuur en Milieu (IenM). In dit programma wordt gezocht naar aansprekende cases voor het verkrijgen van nieuwe inzichten uit nieuwe dataverbindingen. Dat wil zeggen het verbinden van databronnen die eerder niet met elkaar werden verbonden of door op een alternatieve wijze te verbinden, anders dan in het verleden werd gedaan. Bijvoorbeeld met gebruikmaking van recente inzichten en mogelijkheden van (big) data.

Eerder bestond al de uitwisseling van data tussen de emissieregistratie, waar CBS aan bijdraagt en de NEa over de fysieke (IPCC) emissies in Nederland. Dit onderzoek richt zich op de koppeling van de microdata van de ETS-bedrijven bij de NEa met de data van die bedrijven bij CBS. De ETS-bedrijven zijn de bedrijven die verplicht deelnemen aan het Europese systeem van CO₂-emissiehandel (EU ETS).

CBS dankt de NEa en het Ministerie van IenM onder andere met het programma 'Anders omgaan met data' voor de geboden mogelijkheden in dit project aangaande data, kennisuitwisseling en samenwerking.

Samenvatting

De onderzoeksvraag

Dit project onderzoekt of de microdata van de ETS-bedrijven bij de NEa gekoppeld kan worden met de microdata van die bedrijven bij CBS. De ETS-bedrijven zijn de bedrijven die verplicht deelnemen aan het Europese systeem van CO₂-emissiehandel, het EU ETS.

Het datakoppelingsproject is als innovatief aangemerkt en als voorbeeld hoe technische dataverbindingen kunnen worden ingezet voor beleidsontwikkeling. Het project maakt daarom onderdeel uit van het IenM programma 'Anders omgaan met data' waarbij het streven is anders om te gaan met de data en datakoppeling dan in de praktijk gebruikelijk.

De eerste onderzoeksvraag luidt of een werkende data-infrastructuur kan worden opgezet tussen enerzijds de NEa-data en anderzijds bedrijfseenheden in het ABR bij CBS. Vervolgens of dat kan worden gekoppeld met andere statistieken voor 2013 en 2014.

De tweede onderzoeksvraag was hoe de koppelingssystematiek verder kon worden ontwikkeld met het oog op de mogelijkheden voor toekomstige analyse en voortgaande verbetering daarvan, dat ten opzichte van het pilot project dat voorafging.

Door in dit project de gegevens van de ETS-bedrijven te koppelen aan de data-infrastructuur van CBS ontstaan mogelijkheden voor onderzoek, beleid en praktijk om de economische prestaties en de milieuprestaties van de ETS-bedrijven met elkaar in verband te brengen en te analyseren. Het project test de haalbaarheid van een data-infrastructuur waarmee de koppeling tussen de NEa-data en CBS-data aangaande de ETS-bedrijven kan worden gerealiseerd.

De datakoppeling

De data-infrastructuur wordt opgezet vanuit de populatie van 450 emissievergunningen, ook wel aangeduid als 'inrichtingen'. Deze zijn uitgegeven aan een kleiner aantal bedrijven die allen deelnemen aan het EU ETS in de verslagjaren 2013 en 2014.

Als eerste koppelonderdeel wordt getracht de populatie ETS-vergunningen zo goed mogelijk aan te sluiten op de bedrijfseenheden in het Algemeen Bedrijven Register (ABR) van CBS. Het ABR is voor CBS het basisbestand met de gehele Nederlandse bedrijvenpopulatie. Om daar de ETS bedrijven met hun 450 emissievergunningen aan te koppelen vormt de *kernvraag* van dit onderzoek. Kortom zijn de statistische eenheden, hier specifiek de individuele Bedrijfseenheid (BEID), en in mindere mate de Ondernemingsgroep (OG), eenduidig te identificeren en af te bakenen zodanig dat ze dezelfde activiteit ofwel bedrijvigheid beschrijven als de vergunningen en aldus verantwoord aan elkaar kunnen worden gekoppeld. Hier worden namelijk de uitkomsten van dit en mogelijk vervolgonderzoek steeds aan gerelateerd.

Per vergunning zijn van de NEa de relevante openbare gegevens ontvangen als vergunningsnummer, inrichtingsnaam, locatie, adres en KvK-nummer van de 'drijver' of vergunningshouder als geregistreerd bij de NEa en aanvullende gegevens. Dit wordt gekoppeld aan het ABR en vormt zo de ruggengraat voor de datakoppeling.

Het tweede koppelonderdeel bestaat uit het doorverbinden, het koppelen van deze vergunningen – BEID ruggengraat aan de andere CBS statistieken. Dat onder andere aan de volgende statistieken: Productiestatistiek (economische gegevens bedrijven), Internationale Handel, Investerings, Arbeidsrekeningen, Polis administratie, zeggenschap bedrijven, verkopen industriële producten (PRODCOM), onderzoek en innovatie statistieken, Research en Development enquête (R&D), Innovatie enquête (CIS), Milieukosten Bedrijven,

Energiestatistieken en milieustatistieken. Bevindingen hiervan zijn te vinden in het belendende rapport met beschrijvende statistiek.

Bovendien is ten behoeve van check op de kwaliteit van de koppelingen ook zoveel mogelijk gekoppeld aan individuele bedrijfsgegevens bij de Emissieregistratie en de (elektronische) milieujaarverslagen (e-MJV) zoals die worden opgesteld voor grote bedrijven.

Deze koppelingen zijn gerealiseerd voor de verslagjaren 2013 en 2014. Voor de toekomst kan dit worden uitgebreid met de nieuwe verslagjaren 2015 en verder evenals terug in de tijd. De benodigde brondata hiervoor is reeds deels georganiseerd. Tevens is een data-infrastructuur opgezet die geschikt is om het systeem van koppelingen verder uit te breiden, correct om te gaan met vertrouwelijke informatie in de CBS omgeving en geschikt om (inhoudelijke) analyses te verrichten op de data. Alle data is op een afzonderlijke locatie op het ICT-netwerk opgeslagen. Vanwege de gevoelige en vertrouwelijk beschikbaar gestelde informatie op bedrijfsniveau worden toekomstige analyses volgens de regels en binnen de beveiligde omgeving uitgevoerd.

Voor beheer en onderhoud van de opgezette data infrastructuur zal op verzoek van externe partij of partijen jaarlijks een nieuw koppelbestand worden gemaakt. Dat gebaseerd op de laatste set met vergunningsnummers met bijgeleverde KVK-nummer en de laatste stand van het ABR. Voor de verschillende gekoppelde statistieken wordt bij elk proces gezorgd voor validatie en opslag van de (jaar)gegevens.

De technische datakoppeling met de koppelroute: 1. emissievergunningen – 2. BEID ruggengraat (ABR) – 3. CBS statistieken vormt de kern van dit datakoppelingsproject. Hierbij gaat het er primair om individuele bedrijfseenheden in beide registratiesystemen te identificeren en af te bakenen zodat ze beiden zo goed mogelijk dezelfde activiteit beschrijven en daarmee verantwoord kunnen worden gekoppeld. Dit vindt plaats in een aantal duidelijk omschreven processtappen, die deels geautomatiseerd, deels in combinatie, en deels handmatig moeten worden uitgevoerd. Hierbij spelen een viertal koppelrelaties, uiteenlopende combinaties van Vergunning – OG – BEID een rol. Afhankelijk van vereisten bij het vervolgonderzoek zijn keuzen ingebouwd voor de manier van koppeling met bijvoorbeeld verschil in accent op de geautomatiseerde versus handmatige koppeling. De handmatige koppelingen zijn ondersteund door gebruik te maken van publiek beschikbare bronnen alsook van meer vertrouwelijk data binnen CBS en elektronische milieujaarverslagen. CBS heeft overigens hierbij geen gebruik gemaakt van (vertrouwelijke) achtergrondinformatie van de NEa. Dit koppelen, deels handmatig blijkt een arbeidsintensief proces maar is essentieel voor de kwaliteit van de koppelingen nu en bij uitbreiding van de tijdreeks. De handmatige check is toegepast bij de 150 grootste ETS bedrijven die tezamen 95 procent van de emissie veroorzaken. We onderscheiden in de koppeling van Vergunning – OG – BEID een zogenaamde '*ruime benadering*' waarbij van de gekoppelde OG's via een of meerdere BEIDs, alle onderliggende BEID's worden meegenomen. Daarnaast onderscheiden we de '*enge benadering*' waarbij diverse onderliggende BEID's wel worden uitgeselecteerd omdat ze niet aantoonbaar bijdragen aan de ETS-emissies, dan wel omdat ze geheel andere, geen emissie-intensieve activiteiten hebben zoals in geval van de groot- en detailhandel of bijvoorbeeld holdings- en concerndiensten of bepaalde andere diensten. Diverse koppelproblemen moesten overwonnen worden, waaronder het zo goed mogelijk omgaan met de grootschalige industriële bedrijvensites, waar emissies van verschillende bedrijven worden gecombineerd in een of enkele vergunning(en) in een bedrijfseenheid. Om de koppeling zo goed mogelijk uit te voeren is alvast gebruik gemaakt van enkele kernvariabelen uit de gekoppelde statistieken, dit biedt extra houvast bij de koppeling.

Het resultaat

Het technische resultaat is dat de ETS-vergunningen voor 2013 en 2014 voor 100 procent te koppelen zijn aan het ABR. In meer detail leverde koppeling emissievergunning – ABR het volgende resultaat. Voor 109 vergunningen bleek één unieke BEID te koppelen voor verslagjaar 2013. Vervolgens is er een categorie waarin in 95 gevallen meerdere vergunningen koppelden aan één unieke BEID, dat leverde 204 unieke combinaties, waarbij de unieke BEID één of meerdere vergunningen heeft. De rest 465 minus 204 betekent 261 BEID'S, waarbij één of meerdere vergunningen bij meerdere BEID's kunnen horen. Hierbij zitten ook enkele speciale gevallen zoals een enkele grote bedrijvensite. Dit via de OG-route met alle daaronder vigerende BEIDs, grofweg een aantal van 900 BEID's, dat wordt aangeduid als de '*ruime benadering*'. Dit betekent dat in geval behoefte is aan een totaal koppeling aan de OG met bijbehorende BEID's die dan volledig koppelen. Selecteert men scherper, dat is naar de exacte BEID die de ETS – emissies zelf veroorzaakt dan verliezen we BEIDs, dat wordt aangeduid als de '*enge benadering*'. Het aantal gekoppelde BEID's loopt dan terug van 900 naar grofweg 550.

De koppelmogelijkheden van de NEa-data aan het ABR bij CBS blijkt toereikend. In elk geval is alle bedrijvigheid die hoort bij de emissievergunningen te identificeren in het ABR en kan die worden verbonden met de CBS-data. Dat tenminste volledig op het niveau van de ondernemingsgroep. Voor de koppeling tussen de vergunningen en de individuele BEID's in het ABR, daarvan zijn de bevindingen genuanceerder. Een deel van de vergunningen kan naadloos worden gekoppeld. Voor een ander deel is voor een kwalitatief goede koppeling, een aanzienlijke hoeveelheid research nodig voor beoordeling en handmatige correcties op een eerste bod koppelingen die geautomatiseerd tot stand kwamen. Daarbij gebruik makend van informatie van zowel publieke bronnen als beveiligde bronnen binnen en deels buiten CBS. Bij mogelijk vervolgonderzoek kan zowel gebruik worden gemaakt van de data volgens de '*ruime benadering*' als volgens de '*enge benadering*'. Dat biedt extra mogelijkheden.

Het 'resultaat' uit de tweede stap van de koppeling, de stap ABR - overige statistieken, is mede afhankelijk van de eerste koppeling. Van of er gekozen is voor de '*ruime benadering*' of voor de '*enge benadering*' en deels ook van het gekozen scenario. Hoe ruimer de gekoppelde BEID populatie, beïnvloed weliswaar niet of nauwelijks de absolute aantallen gekoppelde BEIDs naar de overige statistieken, maar beïnvloedt wel het aandeel dat gekoppeld kan worden. De grote BEIDs worden vooral gekoppeld, de kleine BEID's die er met name bijkomen in de ruime benadering minder goed. Dit verlaagt het percentage gekoppelde BEID's. Zelfs bij de '*ruime benadering*' wordt een groot deel van de BEID's gekoppeld naar de verschillende statistieken en dus naar reeksen van variabelen. Variabelen die stuk voor stuk mogelijkheden bieden voor vervolgonderzoek en analyse.

Hiermee komen we tot het inhoudelijke doel: Enerzijds het maken van gedetailleerde beschrijvende statistieken van de ETS bedrijven op het niveau van de vergunningen. En anderzijds het analyseren van de ETS bedrijven in termen van hun economische prestaties en in samenhang daarmee hun energie- en milieutechnische prestaties. Waarbij op microniveau de mate van ont koppeling tussen economische en milieuprestaties kan worden geanalyseerd en daarmee de dynamiek in beeld kan worden gebracht die onderligt aan de ont koppeling van de ETS sectoren op bedrijfstakniveau en op nationale schaal volgens de milieurekeningen.

Concluderend kan worden gesteld dat de eerste belangrijke onderzoeksvraag, of een werkende data-infrastructuur kan worden opgezet, bevestigend kan worden beantwoord. De NEa-data kan worden gekoppeld aan de BEID's in het ABR. Vervolgens konden de data van de andere statistieken aan deze dataruggengraat worden gekoppeld.

De tweede onderzoeksvraag was of de koppelingssystematiek, ten opzichte van het voorafgaande pilotproject verder kon worden ontwikkeld met het oog op de vereisten voor toekomstige analyse. CBS heeft mede vanwege het geboden projectkader door het Ministerie van IenM van het *'Anders omgaan met data'* hierbij voor een andersoortige aanpak gekozen. Verschillend van de aanpak bij de pilot en met een andere aanpak dan die doorgaans wordt gehanteerd bij CBS voor koppeling van microdata. De koppelingssystematiek is aanzienlijk doorontwikkeld en in staat gebleken passende resultaten op te leveren. Daarmee is het project een succesvolle exercitie gebleken waarbij de uiteindelijke koppelresultaten de verwachtingen bij de start van het project overtreffen.

De aanbevelingen voor vervolg

Het project heeft een aantal inzichten en producten opgeleverd die reden geven om samen met belanghebbende en geïnteresseerde organisaties te bekijken of het koppelproduct kan worden uitgebreid in scope. Daarbij zal blijvend passende aandacht besteed moeten worden aan de kwaliteit van de koppelingen. Dat is tegelijk een belangrijke aanbeveling, om in een vervolgproject passend capaciteit te besteden aan consolidering en verbetering van de koppeling met een scherp oog voor afbakening van de ETS - bedrijven.

Een andere aanbeveling is om een uitgebreidere tijdreeks te maken met daarbij het recentste jaar, in dit geval 2015. Daarnaast om de data samen te stellen terug in de tijd. Dat mogelijk terug tot aan 2012 of zo mogelijk terug tot aan 2008 vanwege de start van de tweede fase van het EU ETS. Dat biedt dan mogelijkheden voor monitoren van de ETS-bedrijvendynamiek en voor het meten en analyseren van de economische en milieuprestaties in relatie met het EU ETS. Het project biedt ruim voldoende aanknopingspunten voor vervolgonderzoek. De prioriteiten bij de relevante stakeholders zullen bepalen of hier invulling aan kan worden gegeven.

Vanuit het perspectief van het programma *'Anders omgaan met data'* door het Ministerie van IenM lijkt er reden om de in dit project gehanteerde aanpak om met een frisse en alternatieve aanpak naar (bestaande) data te kijken vaker te hanteren. Voor een dergelijke, meer creatieve omgang met data is vaak geen ruimte. Indien expliciet tijd en capaciteit wordt besteed om met zowel een creatieve als deskundige blik naar nieuwe combinaties van bestaande datasets te kijken, kan waardevolle nieuwe informatie wordt gegenereerd voor praktijk en beleid.

1. Inleiding

1.1 Introductie en achtergrond project

De Nederlandse Emissieautoriteit borgt als uitvoeringsorganisatie en toezichthouder dat bedrijven die deelnemen aan het EU ETS voldoen aan hun verplichtingen. Gegevens over deelnemende bedrijven worden bijgehouden in een dataregister. Het EU ETS is het Europees brede systeem voor verhandelbare emissierechten (CO₂-rechten). De NEa heeft vanuit deze functie de beschikking over emissiedata van de ETS-bedrijven, dit betreft de 'operators', in het vervolg van rapport noemen we dit de ETS-bedrijven. De bedrijven die alleen handelen in emissierechten (de handelsbedrijven) vallen buiten de scope van het project.

Door in dit project deze gegevens te koppelen aan de data-infrastructuur van Centraal Bureau voor de Statistiek ontstaan mogelijkheden voor onderzoek, beleid en praktijk om de economische prestaties en de milieuprestaties van de ETS-bedrijven te monitoren.

De eerste ideeën voor een dergelijke koppeling zijn een aantal jaren geleden ontstaan bij de NEa en CBS. Dit heeft verder gestalte gekregen tijdens en na een kennismakingsgesprek van de Directeur Generaal van CBS met de directeur en bestuursvoorzitter van de NEa in september 2014. Het idee ontwikkelde zich dat het combineren van data en informatie van de NEa met die bij CBS een aantal mogelijkheden voor analyse van het functioneren van de ETS-bedrijven en het ETS-systeem zouden kunnen opleveren. Afgesproken is vervolgens dat de NEa en CBS de mogelijkheden voor een concreet samenwerkingsproject zouden verkennen.

Uit een eerste verkennend pilotonderzoek van CBS dat in 2015 werd uitgevoerd (CBS, 2015a) bleek dat er nog een aantal voorbereidende werkzaamheden nodig waren om de CBS-data goed koppelbaar te maken aan de NEa-data.

Het datakoppelingsproject wordt als innovatief gekenmerkt en als een voorbeeld van hoe technische dataverbindingen kunnen worden benut voor beleidsontwikkeling. Het is uitgekozen als een van de voorbeeldprojecten voor 'Anders omgaan met data' en ontvangt vanuit die titel financiering voor de benodigde personele inzet bij CBS. De NEa treedt op als formeel gedelegeerd opdrachtgever en CBS brengt in een aantal fasen de datakoppeling tot stand met de grofweg deelnemende bedrijven met 450 vergunningen (inrichtingen) aan het EU ETS. Voor deze koppeling moest een data-infrastructuur worden opgezet bij CBS.

1.2 Doel van het project

Met dit project is de haalbaarheid onderzocht van een data-infrastructuur waarmee de koppeling tussen de NEa-data en CBS-data aangaande de ETS-bedrijven kan worden gerealiseerd. Vervolgens is die data-infrastructuur ontwikkeld. Startpunt van deze data-infrastructuur vormt de populatie van 450 vergunningen bij bedrijven die deelnemen aan het EU ETS in de verslagjaren 2013 en 2014. Hiervan beheert de NEa het Nederlandse deel van het Europese ETS-register. We beogen met dit project om de informatie van de ETS-vergunningshouders bij de NEa zo goed mogelijk aan te sluiten op de bedrijfseenheden in het Algemeen Bedrijven Register (ABR) van CBS. Bij de koppeling wordt een consistente indeling naar bedrijfstakken (sectoren) nagestreefd. Het principale doel van het project is het tot stand brengen van de datakoppeling en de benodigde technische infrastructuur.

Het eerste inhoudelijke doel van het project is het maken van beschrijvende statistiek over de ETS bedrijven. Dat wil zeggen het beschikbaar maken van economische en milieutechnische informatie zoals over werkgelegenheid, productie, investeringen, CO₂-emissies en energiegebruik van de ETS bedrijven en de relevante ETS- sectoren. Deze informatie zal de kennis over het EU ETS en over het gedrag bij de deelnemende ETS-bedrijven in de Nederlandse economie vergroten. Verkregen inzichten kunnen ingezet worden voor het verrijken van publicaties over emissiehandel, bij analyses over het functioneren van het ETS-systeem en bij de advisering inzake het beleid over het EU ETS. De economische en milieutechnische informatie die met de koppeling wordt verkregen, kan bijvoorbeeld (op termijn) opgenomen worden in het jaarlijkse voortgangsrapport 'Voortgang Emissiehandel' van de NEa.

Het tweede inhoudelijke doel dat met de datakoppeling wordt beoogd is het verrichten van (econometrisch) onderzoek naar de effecten van het EU ETS op de economische en milieuprestaties van de Nederlandse ETS-bedrijven. Hierbij kunnen verschillende hypothesen worden getoetst. Bijlage IV vermeldt een aantal potentiële onderzoeksvragen voor vervolgonderzoek. Bij de keuzen gemaakt bij het realiseren van de data-infrastructuur, was het van belang deze mogelijke onderzoeksvragen als leidraad te nemen. De koppeling zou de beantwoording van genoemde vragen moeten faciliteren.

Als onderdeel van het project is het eerste inhoudelijke doel in 2016 getest (zie: CBS, 2016b). In relatie tot het tweede inhoudelijk doel zijn er (nog) geen inhoudelijke ondervragen aan de orde gesteld. CBS en de NEa bekijken gezamenlijk na afronding van het project de optie om in overleg met de relevante ministeries en eventueel andere partijen of er technische en financiële mogelijkheden zijn het huidige onderzoek een vervolg te geven. Een vervolg gericht op uitbreiding van de tijdreeks en op inhoudelijke onderzoeksvragen.

1.3 Producten en rapportage

Het project is in drie fasen opgedeeld met het plan van aanpak, de data-infrastructuur en rapportage. Daarbij zijn met de opdrachtgever een viertal producten afgesproken: Een het plan van aanpak voor de data-infrastructuur en statistiek in een (interne) notitie; De data-infrastructuur waarin de NEa populatie vergunninghouders (ETS-bedrijven) verbonden worden met het Algemeen Bedrijven Register (ABR) en de overige data-koppelingen bij CBS; Dit rapport met de bevindingen van genoemde datakoppeling en tenslotte een rapport met beschrijvende statistiek gereed na fase.

Het plan van aanpak is uitgevoerd in fase twee, hierbij werd de ETS-populatie vastgelegd en de data-infrastructuur met de beoogde data-koppelingen opgezet. De rapportage over dit project vindt uiteindelijk plaats in twee rapporten. Dit eerste rapport behandelt de data koppeling van de ETS-populatie, de benodigde data bronnen intern en extern en de ontwikkelde data-infrastructuur bij CBS voor nu met een beschrijving voor continuering van de infrastructuur in de toekomst. Het tweede rapport toont de eerste resultaten van de datakoppelingen met diverse statistieken aan de hand van beschrijvende statistiek voor de ETS-bedrijven in Nederland.

Dit rapport is als volgt opgebouwd: Hoofdstuk 2 geeft een beschrijving van de populatie vergunninghouders bij de NEa en in de verschillende registers bij CBS. Hoofdstuk 3 vervolgt met een beschrijving van de databronnen en organisatie van de data-infrastructuur en technische koppelingen in het project. Hoofdstuk 4 bespreekt de methode met technische

aspecten en bevindingen van de datakoppelingen aan het ABR. Hoofdstuk 5 toont kort de technische resultaten van de koppeling. Waarna in hoofdstuk 6 het plan voor toekomstig gebruik, beheer en onderhoud van de in dit project ontwikkelde data-infrastructuur bij CBS wordt behandeld. Tenslotte trekken we in hoofdstuk 7, op basis van de bevindingen in het project, een aantal conclusies en doen we een aantal aanbevelingen voor toekomstig gebruik van de data infrastructuur.

2. Populatie ETS bedrijven

Dit hoofdstuk beschrijft de afbakening en de uiteindelijk vastgelegde populatie van ETS-bedrijven in Nederland. Het betreft de bedrijven in Nederland die verplicht deelnemen aan het Europese handelssysteem voor CO₂, het EU ETS.

2.1 Populatie ETS-bedrijven en regels voor deelname EU ETS

Middels het EU ETS 'reguleert' de overheid in Nederland middels een marktconform beleidsinstrument de uitstoot van zo'n 350 vergunningshouders met grofweg 450 vergunningen. De ETS bedrijven veroorzaken gezamenlijk grofweg 45% van de CO₂-emissies in Nederland en moeten hun uitstoot monitoren, bij de NEa rapporteren en vervolgens jaarlijks emissierechten inleveren navent hun gerapporteerde uitstoot. De ETS-bedrijven zijn meestal grote, energie-intensieve bedrijven die vooral te vinden zijn in de (petro-)chemische industrie, metaalsector en de elektriciteitsproductie¹. De Europese richtlijn voor emissiehandel bepaalt welke inrichtingen (installaties) bij bedrijven verplicht zijn deel te nemen aan het EU ETS. De bijlage bij de richtlijn beschrijft de regels hiervoor, maar de precieze interpretatie van die regels behoeft wel specifieke kennis. Het recente EU ETS voor de periode 2013-2020 bepaalt dat het verbranden van brandstof in installaties met een totaal nominaal thermisch ingangsvermogen van meer dan 20 megawatt verplicht tot deelname van die installaties, enkele uitzonderingen daargelaten. Het ingangsvermogen geeft aan hoeveel brandstof maximaal verstoekt zou kunnen worden. De Bijlage-I beschrijft in detail de activiteiten die zich hiervoor kwalificeren en ook de uitzonderingen daarop, zoals installaties voor het verbranden van huishoudelijk en gevaarlijk afval of eenheden die ingezet worden als reserve capaciteit.

2.2 Monitoring ETS-bedrijven populatie door de NEa

De populatie vergunningshouders met circa 450 vergunningen voor luchtmissies die deelnemen aan het EU ETS, staat geregistreerd bij de NEa op basis van hun emissievergunning via een vergunningsnummer met diverse aanvullende gegevens. Veruit het grootste deel van de deze zogenaamde 'operators' in het NEa-register zijn verplicht deel te nemen door overschrijding van de emissies boven de drempelwaarde. De emissies en rechten worden in detail gemonitord door de NEa. Daarnaast zijn ook de zogenaamde handelsbedrijven opgenomen in de registers van de NEa, bedrijven die kunnen handelen in CO₂-emissierechten maar die niet meedoen met hun eigen emissies. Deze laatste categorie, de handelsbedrijven vallen buiten dit onderzoek.

De populatie van ETS-bedrijven, zoals geregistreerd bij de NEa bestaat in het verslagjaar 2013 en 2014 uit respectievelijk 453 en 448 vergunningen (inrichtingen) met unieke vergunningsnummers bij een kleiner aantal vergunninghouders. Bij iedere vergunninghouder, ieder ETS-bedrijf horen een of meerdere vergunningsnummers voor registratie en monitoring door de NEa en bij de emissieregistratie. Aan dit vergunningsnummer is een rekening in het Nederlandse deel van het CO₂-register gekoppeld, waarop het bedrijf zijn emissierechten houdt

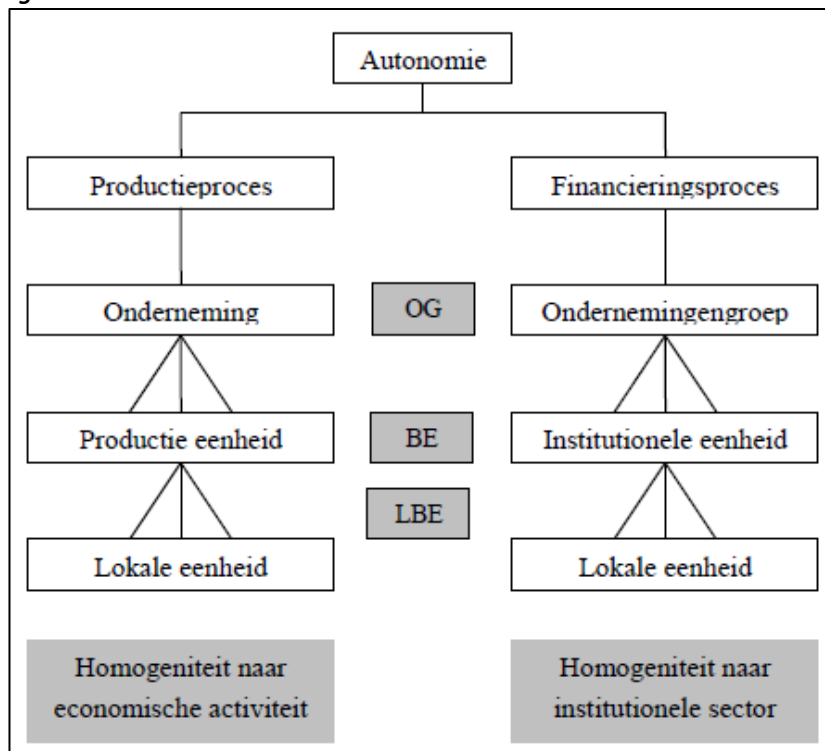
¹ Zie ook: <https://www.emissieautoriteit.nl/onderwerpen/deelnemers-ets>.

en de gegevens van het bedrijf vermeld zijn. Hoewel doorgaans een vergunning, vergunningsnummer en bijbehorende rekening bij de NEa slechts één bedrijf vertegenwoordigt zijn er een aantal uitzonderingen. Er bestaan zogeheten verzamelaccounts die de CO₂-emissies en rechten van een aantal verschillende bedrijven vertegenwoordigt of ‘verzamelt’. Dit verschijnsel wordt vooral waargenomen op een aantal grote bedrijvensites waar bedrijven fysiek gegroepeerd zijn. Deze verzamelaccounts en/of gebundelde (emissie-)gegevens van de ETS-bedrijven beïnvloeden de koppelbaarheid van de NEa vergunningsnummers met de bedrijfsgegevens bij CBS.

2.3 Populatie ETS-bedrijven in het CBS-bedrijvenregister ABR

Het ABR is voor CBS het basisbestand met de gehele Nederlandse bedrijven populatie. De koppeling tussen de ETS bedrijven met 450 keer een vergunningsnummer in de NEa boekhouding en de bedrijfseenheden in het ABR bij CBS vormt de kern van dit onderzoek. Bij het uitvoeren van statistisch onderzoek aan een bedrijvenpopulatie is het van belang dat de statistische eenheden en de onderzoekspopulatie(s) eenduidig zijn afgebakend, zodat helder is waaraan de uitkomsten van het onderzoek zijn gerelateerd. Voor de economische statistieken worden de Statistische Eenheden (SE) apart gedefinieerd en afgebakend.² CBS kent in het ABR drie soorten statistische eenheden de Ondernemingsgroep (OG), de Bedrijfseenheid (BE) en de Lokale Bedrijfseenheid (LBE), zie hiervoor figuur 2.1.

Figuur 2.1. Schematisch overzicht van de actoren in de economie



Bron: CBS; 2012.

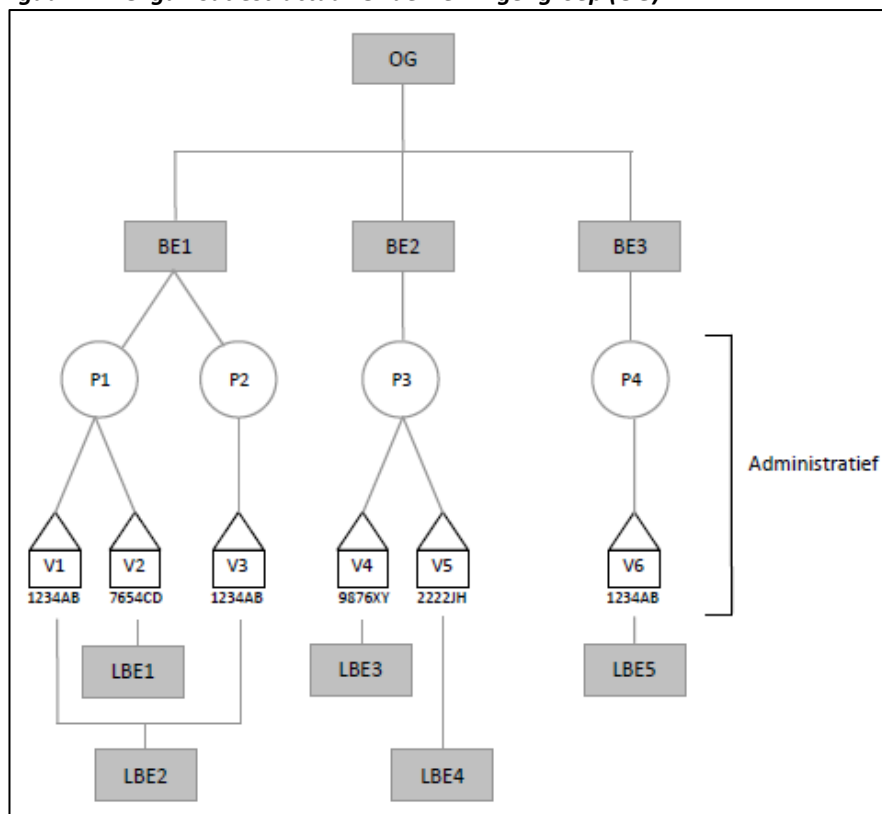
² Een SE in de institutionele statistiek is een actor in de economie, d.w.z. een onderneming. Kenmerkend hiervoor zijn eigenschappen als autonomie, externe gerichtheid, beschrijfbaarheid en continuïteit (CBS, 2012).

Genoemde statistische eenheden moeten zowel waarneembaar als beschrijfbaar zijn en de gevraagde statistische informatie kunnen leveren in hun rol bij het samenstellen van de economische statistieken bij CBS. De functies die hierbij horen naar statistische eenheid:

- de OG voor de beschrijving van het financieringsproces;
- de BE voor de beschrijving van het productieproces;
- de LBE voor de beschrijving van het productieproces op regionaal niveau.

De relatie van de Ondernemingsgroep (OG), via de Bedrijfseenheid (BE), CBS-persoon (P), vestiging (V) en lokale bedrijfseenheid verschilt tussen bedrijven. De CBS-persoon, vormt het aanknopingspunt voor het KvK-nummer.³ Een aantal veelvoorkomende relaties worden getoond in figuur 2.2). Onder een OG kan zowel een als meerdere CBS-personen voorkomen met elk ook een of meerdere BE's (BEID). Een zelfstandige BE kan vervolgens op een óp meerdere vestigingsplaatsen (vestigingen) voorkomen en vestigingen kunnen worden gecombineerd in een LBE.

Figuur 2.2. Organisatiestructuur Ondernemingsgroep (OG)



Bron: CBS; 2012.

Figuur 2.2 illustreert een situatie waarbij een OG bestaat uit 3 BE's. De aan de BE gekoppelde CBS-Personen worden aangeduid met P1 t/m P4 en de gekoppelde CBS-Vestigingen V1 t/m V6. Deze situatie gaat er vanuit dat de postcodes van het vestigingsadres V1, V3 en V6 gelijk zijn (namelijk alle drie 1234AB) en dat de postcodes van V2, V4 en V5 verschillen. Voor BE1 worden

³ CBS-persoon vormen de bouwstenen voor de statistische eenheden in het ABR. Een CBS-persoon is in het algemeen een juridische of administratieve eenheid, een natuurlijk persoon of niet-natuurlijk persoon zoals die in de buitenwereld bestaan, bijvoorbeeld een eenmanszaak, een B.V., etc. Deze CBS-persoon wordt in het systeem gekenmerkt door zijn KvK nummer.

in het ABR in zo'n geval twee LBE's aangemaakt (V1 & V3 en afzonderlijk V2); voor BE2 worden ook twee LBE's aangemaakt (V4 en V5); voor BE3 wordt één LBE aangemaakt (V6). Op dit moment worden in het ABR op het niveau van een LBE, uitsluitend de BEID, de postcode en de registratiedatums als kenmerken vastgelegd. Dus niet de economische informatie in detail. Dat betekent dat we voor economische informatie van de bedrijven dus zijn aangewezen op het hogere niveau van de BE's.

De essentie van het datakoppelingsproject is om de ETS-vergunningen, zoals beschreven in paragraaf 2.2, te koppelen aan de CBS eenheden zoals beschreven in paragraaf 2.3.

3. Databronnen en organisatie

Dit hoofdstuk beschrijft de databronnen en databases en organisatie van de data. De populatie ETS bedrijven in het ABR dient als ruggengraat voor alle datakoppeling in dit project.

3.1 De databronnen en statistieken voor koppeling

Deze paragraaf beschrijft enerzijds de databronnen die nodig zijn om de koppeling tussen de ETS-bedrijfseenheden bij NEa en CBS überhaupt tot stand te kunnen brengen en anderzijds de diverse databronnen die gebruikt worden of mogelijk gebruikt kunnen worden voor beschrijvende statistiek en aanvullend inhoudelijk onderzoek.

3.1.1 NEa bronnen

Hieronder volgt een overzicht van de relevante databronnen (bestanden) van de NEa met data naar verslagjaar en met een opsomming van de relevante variabelen:

- a. Eindejaarstand van de ETS-populatie in de NEa database (2013 en 2014). Hierin zijn de volgende bij de NEa bekende variabelen opgenomen:
 - i. Vergunningsnummer (VNR) ofwel emissievergunningsnummer;
 - ii. Inrichtingsnaam, soms met een aanvullend cijfer in geval van meerdere VNR's per inrichting⁴;
 - iii. Locatie (vestigingsplaats van de inrichting);
 - iv. Postcode (vier cijfers en twee letters van de inrichting);
 - v. Adres (straatnaam met huisnummer van de inrichting);
 - vi. KvK-nummer van de 'drijver' zoals bekend bij de NEa;
 - vii. Jaarlijkse CO₂-emissiegegevens per VNR;
- b. Jaarlijkse gegevens van de toewijzing (allocatie) van CO₂-emissierechten per VNR.

Kort gezegd vormt het KvK-nummer van de 'drijver' (zie ook voetnoot), met de naam, adres, woonplaats ofwel vestigingsplaatsgegevens van de inrichting inclusief postcode, de basis voor de koppeling. Hoewel de jaarlijkse mutaties in de ETS-populatie doorgaans niet bijzonder groot zijn, dient voor elk aanvullend verslagjaar de populatie met de bijbehorende gegevens verkregen en geanalyseerd te worden om wijzigingen in de tijd adequaat te kunnen monitoren. Hierbij kan de 'eventsdatabase' als onderdeel van het ABR behulpzaam zijn. Deze beschrijft mutaties ofwel 'gebeurtenissen' als fusies, splitsingen, overnames bij bedrijven. Voor ieder verslagjaar dient een afgebakende vergunningenpopulatie te worden verkregen van de NEa en vastgesteld ten behoeve van de koppeling aan ABR eenheden per verslagjaar.

⁴ Bij de NEa heeft een inrichting maximaal 1 vergunning. Terwijl ook het begrip 'drijver' wordt gehanteerd, hieronder kunnen meerdere inrichtingen met bijbehorend vergunningsnummer bestaan.

3.1.2 CBS bronnen

De data per ETS-vergunning verkregen van de NEa wordt zo goed mogelijk gekoppeld aan de CBS-microdata van het ETS bedrijf dat bij die vergunning hoort. Daar draait het om in dit project. Bij CBS gaat het om de volgende databronnen en bronstatistieken:

- a. Koppeling ETS – vergunningen aan bedrijfseenheden bij CBS:
 - i. Algemeen Bedrijven Register, dit vormt de ruggengraat voor de datakoppeling;

Via een goede koppeling aan de bedrijfseenheden van het ABR, kunnen de ETS-vergunningen relatief eenvoudig worden verbonden met de andere statistieken, zoals:

- b. Koppeling aan de economische statistieken:
 - i. Productiestatistiek (PS) voor economische gegevens van bedrijven;
 - ii. Statistiek Internationale Handel (SIH) voor bedrijfsgegevens over export en import, dat kan naar goederen IHG en in mindere mate naar diensten, IHD;
 - iii. Statistiek Investeringsgegevens voor investeringsgegevens bedrijven;
 - iv. Arbeidsrekeningen voor arbeidsvolume of arbeidskosten;
 - v. Polis administratie voor (meer gedetailleerde) loonkosten;
 - vi. Zeggenschap van de bedrijven voor een indicatie Nederlands versus buitenlandse zeggenschap per bedrijf;
 - vii. Verkopen voor een selectie van industriële producten naar productgroep (PRODCOM);
- c. Koppeling aan onderzoek en innovatie statistieken:
 - i. Enquête Research en Development (R&D-enquête) voor gegevens over R&D activiteiten van bedrijven;
 - ii. Innovatie enquête (CIS) over innovatie-activiteiten;
- d. Koppeling aan energie- en milieustatistieken alsmede milieurekeningen:
 - i. Milieukosten van Bedrijven (MILKBED) voor milieukosten en overdrachten;
 - ii. Energiestatistieken voor individuele bedrijfsgegevens over energiegebruik;
 - iii. milieustatistieken, voor individuele bedrijfsgegevens CO₂-emissies;
- e. Zie voor een uitgebreide catalogus: <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>.

3.1.3 Overige bronnen

Naast de NEa- en CBS-gegevens van de ETS bedrijvenpopulatie op microniveau, zijn aanvullende inzichten en data van externe bronnen nodig om de datakoppeling van de ETS-vergunningen met de eenheden in het ABR verder te vervolmaken en/of de kwaliteit te controleren en verbeteren.

Hiervoor zijn ook gegevens van buiten CBS en NEa gebruikt om de informatie van de individuele bedrijven zo compleet mogelijk te krijgen. Dit betreft individuele bedrijfsgegevens vanuit de Emissieregistratie, en gegevens verkregen uit de (elektronische) milieujaarverslagen (e-MJV) zoals die worden opgesteld voor grote bedrijven. Van een deelpopulatie van de bedrijven in de Emissieregistratie wordt de luchtmissies gemonitord en dit zijn ook vaak de ETS-bedrijven.

Voor toekomstig onderzoek is een vergelijking van de ETS-populatie met de totale bedrijvenpopulatie van de Nederlandse economie op verschillend terrein (diverse variabelen) mogelijk interessant. Dat kan zowel op macroschaal als op mesoniveau, zodat gegevens over de

ETS-populatie met de rest van een branche of met de gehele economie vergeleken kan worden. Bijvoorbeeld kunnen economische gegevens uit de Nationale Rekeningen (NR) op macro- en mesoniveau voor ETS-bedrijven vergeleken worden met andere niet ETS-bedrijven en totalen per bedrijfstak. Deze NR-gegevens worden samengesteld binnen dezelfde afdeling als de milieurekeningen, en zijn daarom voor de vergelijking in detail beschikbaar. Daarnaast zijn vergelijkingen op andere gebieden mogelijk, zoals investeringen, arbeid en milieu. Dit komt later nog aan de orde in het rapport.

3.2 De organisatie van de data

Deze paragraaf beschrijft de organisatie van de data en de opzet van de data-infrastructuur voor het vastleggen en zo goed mogelijk verbinden van de data van de ETS - bedrijven in Nederland met de verschillende CBS-databases en -bronnen. De data is verzameld voor de verslagjaren 2013 en 2014 én gekoppeld. De data is wel al (deels) ook voor verslagjaren terug in de tijd georganiseerd en zo mogelijk voor verslagjaar 2015, als de data al beschikbaar was. De data voor deze overige jaren is echter nog niet gekoppeld, het streven is om dat in de toekomst wel ook voor elkaar te krijgen. De relevante databronnen van de NEa, CBS en extern zijn in voorgaande paragraaf 3.1.3 benoemd. Voor verdere details zoals met de variabelen, zie eventueel de bijlagen V - IV en voor meer details de bijbehorende rapportage met de beschrijvende statistiek.

Bij de opzet van de ETS-data infrastructuur in de beveiligde omgeving van CBS zijn de volgende elementen van belang:

- a. De NEa-data worden via de beveiligde upload verbinding en beveiligd met een wachtwoord naar CBS gestuurd;
- b. Opslag van de relevante databestanden inclusief meta data vindt steeds plaats in afzonderlijke jaarmappen op het CBS-computernetwerk bij milieurekeningen (Nationale Rekeningen) en bij het Centrum voor Beleidsstatistiek. Voor ieder verslagjaar wordt elk bronbestand van de relevante data opgeslagen onder zo'n jaar map;
- c. Per jaarmap worden de koppelingen getest en waar mogelijk compleet gemaakt;
- d. Het dataformat is vooralsnog in MS-Access. Waar nodig, zoals voor analyse kan ook SPSS worden toegepast en de data daarvoor worden geconverteerd. Dit laatste is gedaan voor de aanvullende rapportage met beschrijvende statistiek en van belang in een mogelijk vervolgtraject met analyse;
- e. Toegang tot deze mappen met microdata is alleen mogelijk voor medewerkers van CBS - milieurekeningen of van medewerkers van het Centrum voor Beleidsstatistiek van CBS;
- f. Analyse op de microdata kan plaats vinden door onderzoekers van CBS;
- g. Voor externe onderzoekers moet eerst worden getoetst of zelfstandig onderzoek of analyse kan worden gedaan op de hier omschreven data met (deels) vertrouwelijke microdata. Daarvoor moet dan eerst toestemming worden verkregen.

Voor details van de meer praktische aspecten van de organisatie van de data, data vastlegging en structurering van de data in de ICT-infrastructuur bij CBS, zie ook bijlage II.

Voor gebruik, instandhouding en onderhoud van de data-infrastructuur die is opgezet, zijn heldere afspraken nodig tussen de NEa, het Ministerie van Infrastructuur en Milieu en CBS. De insteek is dat na afronding van dit project daarover concrete afspraken worden gemaakt.

4. Methode datakoppeling

Dit hoofdstuk beschrijft de technische datakoppeling van de populatie ETS-vergunningen aan het ABR en aan de andere bestanden en statistieken bij CBS. Hierbij wordt aandacht besteed aan de processtappen, de koppelingssystematiek met aanvullende bewerkingen.

4.1 Processtappen voor realisatie van de koppeling

Hier worden de processtappen getoond ter realisatie van de data koppeling:

- a. Vergunningenbestand, van de NEa ontvangt CBS een bestand met de vergunning nummers en bijbehorende gegevens, waaronder het Kamer van Koophandel nummer (KvK-nummer);
- b. Door de CBS-afdeling die het ABR beheert wordt bij elk door de NEa bijgeleverde KvK-nummer technisch een OG gekoppeld met alle bijbehorende BEID's.⁵ Dit is een variatie op de aanpak in de pilot waarbij de VNR met bijbehorende ETS emissies direct aan de BEID werd gekoppeld. Deze koppeling wordt voor ieder verslagjaar herhaald omdat het aantal BEID's binnen een OG kan veranderen. Dit bestand wordt in het datasysteem (Access) gezet. Hierbij wordt de procedure en het bestandsformaat vastgelegd;
- c. Aan het bestand met ABR gegevens worden per BEID geautomatiseerd extra gegevens gekoppeld, waaronder de Productie Statistiek gegevens (PS-gegevens). Dit betreft de gegevens op basis van de PS-vragenlijst per BEID gekoppeld in het systeem. Belangrijk hierbij zijn ook extra gegevens volgens de concepten van de Nationale Rekeningen (NR), zoals de Productiewaarde, Inkoopwaarde en Toegevoegde waarde;
- d. Handmatig analyseren van productie-eenheden op specifieke productielocaties, dat is vooral van betekenis bij de categorie vergunningsnummers met een lettertoevoeging, die mogelijk wel een unieke KvK (en of BEID) hebben i.p.v. een gecombineerde. Dat verdient aandacht in het proces. De relevante bedrijfsnamen op die locaties worden bij gezocht en er wordt getracht de ontbrekende KvK-nummers en waar mogelijk ook de BEID's toe te voegen of zo nodig te corrigeren. Hierbij is onder andere van belang duidelijkheid te verkrijgen over de samenhang met mogelijke andere vestigingen en of andere productielocaties onder de gevonden BEID's. Hierbij is het streven om de economische prestaties correct te alloceren;
- e. Nu kunnen de andere CBS gegevens en statistieken op BEID niveau, op microniveau worden gekoppeld. In eerste instantie wordt de PS omzet en de CO₂-emissie gekoppeld om ook bij stap vii te gebruiken als leidraad bij de handmatige verbetering van de koppelingen.
- f. Hierbij is het nodig plausibiliteitschecks uit te voeren en resultaten te interpreteren. In geval van bijzonderheden nagaan wat de bijzonderheid verklaart en waar nodig de koppeling aanpassen.
- g. Tenslotte kunnen de andere statistieken worden gekoppeld, dat kan met beide gebruikte ICT-tools, zowel met Access als met SPSS (of indien nodig andere tools) voor analyse.

De volgende paragraaf gaat meer in detail in op de precieze aanpak van het koppelingsproces en de keuzes die worden gemaakt bij de koppelingen.

⁵ Voor toekomstige datakoppeling kan deze processtap mogelijk zelfstandig door de onderzoeksafdeling worden uitgevoerd op een complete afslag van het ABR.

4.2 Koppelingssystematiek ETS-populatie aan het ABR

De koppeling van de ETS-bedrijven populatie aan het ABR vormt de kern van het datakoppelingsproject. De koppeling van de ETS-bedrijven start vanuit de NEA-vergunningsnummers en de bijhorende gegevens met KvK-nummer en adresgegevens en koppelt aan de bedrijfseenheden in het Algemeen Bedrijven Register (ABR) bij CBS (zie paragraaf 3.2).

De **kernvraag van de koppeling** luidt: zijn de individuele bedrijfseenheden in beide registratie systemen te identificeren en af te bakenen, zodanig dat ze beiden dezelfde activiteit beschrijven en aldus verantwoord aan elkaar kunnen worden gekoppeld.

In de pilotstudie (CBS, 2015a), werden de NEa vergunningen rechtstreeks gekoppeld aan de BEID's in het ABR. Dus niet indirect via de overkoepelende OG zoals in dit onderzoek. Deze koppeling met het ABR werd toen tot stand gebracht rechtstreeks (en alleen) via de koppelvariabele KvK-nummer van de 'drijver' verkregen bij de NEa. Bij een drijver kan zowel een als meerdere vergunningen voorkomen. De NEa heeft per verguningsnummer in hun administratie een KvK nummer opgenomen die is bijgeleverd aan CBS. Die bijgeleverde KvK vormt nu eveneens het startpunt van het koppelingsproces, nu eerst naar de OG.

Een BEID in het ABR als feitelijke transactor in het economische productieproces werd in de pilot gehanteerd als de sleutelvariabele voor alle koppelingen. Deze variabele is nodig om vanuit het ABR de andere databestanden te kunnen koppelen en de NEa-bestanden te verrijken met de eerder omschreven gegevens vanuit de verschillende CBS-statistieken. Uitgaande van de 450 emissievergunningen in Nederland, levert die koppeling van de bijgeleverde KvK's aan het ABR meer dan 550 hits aan BEID's op. Een beperkt deel van de 450 ETS-vergunningen kan gekoppeld worden aan meer dan een BEID. Daartegenover staat dat een substantieel deel van de BEID's aan meerdere VNR's is te koppelen. Dit uiteenlopende koppelresultaat is reeds te zien in de samenvattende tabel uit de pilot studie uitgevoerd in 2015 (CBS, 2015a; Bijlage II). Geregeld kunnen meerdere VNR's worden toegekend aan één economische eenheid (BEID). Daarnaast is voor een weliswaar beperkt deel van de VNR's de toekenning van een en hetzelfde KvK-nummer niet toereikend. Daarom moeten voor deze gevallen ontbrekende KvK-nummers *handmatig* worden bij gezocht. Die procedure doen we in dit onderzoek anders en we koppelen de geleverde KvK-nummers aan de OG's in het ABR. Dat levert een grotere set aan gekoppelde BEID's op. Die kunnen we afhankelijk van het type vervolgonderzoek meer of minder stringent uitselecteren. Waarbij meer of minder van gerelateerde economische bedrijvigheid bovenop de primaire activiteit met grote emissies, zoals bijvoorbeeld groothandel, detailhandel, concerdiensten en holdingactiviteiten worden meegenomen. De mate van uitselecteren bepaald of we spreken van de '*ruime benadering*' of ingeval van scherpe selectie tot aan een BEID per VNR de '*enge benadering*'.

De **kern van de koppelingsproblematiek** is: de ETS-vergunningen zijn deels wel, maar deels ook niet een-op-een koppelbaar aan de CBS-eenheden.

In dit project verloopt de koppeling anders dan bij de pilot, namelijk via de route van de OG, via de zogenaamde '*ruime benadering*'. Dit mede om dat conceptueel en data-technisch te testen bij de diverse benodigde processtappen en voor het effect op het koppelresultaat. Daarmee

laten we met het oog op vervolgonderzoek ook de keuze open over welke onderliggende BEID's worden meegenomen. Deze keuze is te maken met de scenario's die hierna worden toegelicht. Ook met deze route vragen de grootschalige productiesites en/of voor verzamelaccounts in Nederland extra aandacht. Dit gaat om de situatie op bepaalde grootschalige bedrijvensites waar emissies van diverse bedrijven en/of inrichtingen zijn verzameld en ondergebracht in een gespecialiseerde organisatie die de zaken omtrent CO₂-emissiehandel regelt. De onderliggende economische bedrijvigheid moet ook hier zo goed mogelijk worden geïdentificeerd en de economische omvang vastgesteld. Daartoe dienen ook deze onderliggende economische activiteiten het correcte KvK-nummer toegekend te krijgen en vervolgens aan de juiste OG te worden gekoppeld.

Een voorbeeld

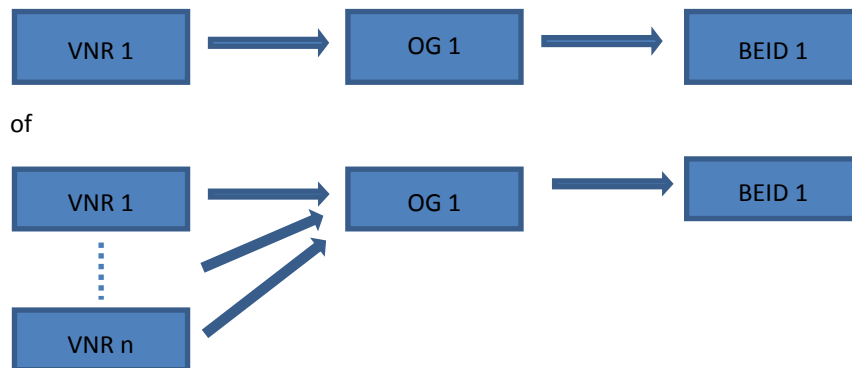
Voor de bedrijvensite Chemelot bestaat een serie aan VNR's, die worden allen gedekt door een bedrijf namelijk de 'Chemelot Site Permit'. Dat leverde vanzelfsprekend ook slechts een KvK, een OG en een BEID op van beperkte economische omvang. Omdat de onderliggende economische bedrijvigheid op de site veel groter is, waren aanvullend zoekwerk en bewerkingen nodig. De bedrijven actief op de site werden erbij gezocht met bijbehorende KvK's en BEID's. Vervolgens moesten de VNR's, met de bijbehorende ETS-emissies zo goed mogelijk worden gealloceerd en verbonden met de relevante OG's en BEIDs. Daarvoor werden de beschikbare gegevens, ook van de productielocaties, emissieverslagen, adresgegevens gebruikt om tot een zo goed mogelijke allocatie te komen.

In fase 1 van het project zijn ter verkenning de gegevens over verslagjaar 2013 gebruikt om na te gaan of koppeling tussen de NEa-vergunningen (VNR's) en CBS-informatie op bedrijfsniveau mogelijk is. Koppeling van de VNR met de zogenaamde Ondernemingsgroep (OG) in het ABR blijkt in alle gevallen mogelijk. Met andere woorden de ETS-bedrijven zoals bekend bij de NEa zijn volledig traceerbaar in het ABR. Echter voor een correcte afbakening is een aanzienlijke hoeveelheid tijd en inspanning vereist. De belangrijkste vraag hierbij luidt: *is de emissievergunning zoals afgebakend in de NEa administratie, te identificeren en overeenkomstig en eenduidig af te bakenen in het ABR?* De koppeling van de ETS-vergunningen met de onder de OG voorkomende bedrijfseenheid of -eenheden is in een deel van de gevallen gecompliceerd. Dat is niet voor de gehele populatie eenduidig en daarvoor zoeken we een aanpak die vraagt om *'anders om te gaan met de data'* dan doorgaans plaatsvindt. Voorgaande betekent dat we in dit project het koppelproces starten door een verbinding te leggen van de ETS-vergunningen met de OG's in het ABR in plaats van direct aan de BEID's. Later in het koppelproces zijn alternatieven, 'scenario's' te kiezen om ook de koppeling aan de individuele bedrijfseenheden, de BEIDs mogelijk te maken. Die worden hierna besproken.

Figuur 2.2 in paragraaf 2.3 toont de mogelijke relaties die tussen een OG en BEID voorkomen in het ABR. De figuur laat zien dat per OG meerdere BEID's kunnen voorkomen. En dat per BEID een situatie van één zogenaamde CBS-persoon, één vestigingsadres en één Lokale Bedrijfseenheid (LBE) kan bestaan. Maar ook dat dat evengoed anders kan zijn, met verschil in de afbakening ten opzichte van de ETS-vergunningen. De match van de VNR-KvK's met de OG blijkt er dus steeds te zijn. De match met de BEID, het laagste niveau waarop de benodigde economische informatie beschikbaar is kent voor een deel van de onderzoekspopulatie een match. Het overige deel toonde niet de gewenste een-op-een match en toonde uiteenlopende combinaties.

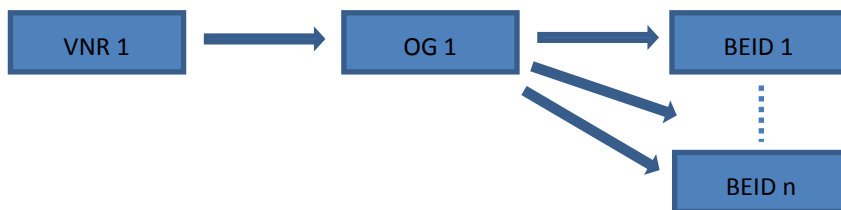
We tonen hierna de mogelijke **koppelrelaties** startend vanuit de VNR en het bijgeleverde KvK-nummer, vervolgens via koppeling aan OG naar de BEID's. Iedere koppelrelatie werkt slechts voor een bepaald gedeelte van de populatie, gezamenlijk dekken de vier koppelrelaties de gehele ETS-populatie.

Koppelrelatie 1: Eén OG heeft **1 BEID** en er zijn **een of meer vergunningen (VNR)** aan toegekend. In dit geval is de totale emissie uit de vergunning(en) te koppelen aan gegevens van deze unieke BEID.



Koppelrelatie 2: Eén OG heeft **meerdere BEID's** en er is **één vergunning** aan toegekend. In dit geval moet worden **beslist** welke BEID's worden meegeteld:

- Optie 1: alle of;
- Optie 2: alleen die tot dezelfde 'productie bedrijfstak (SBI)' horen of;
- Optie 3: BEID op dezelfde locatie van de vergunning of;
- Optie 4: een andere handmatige keuze.

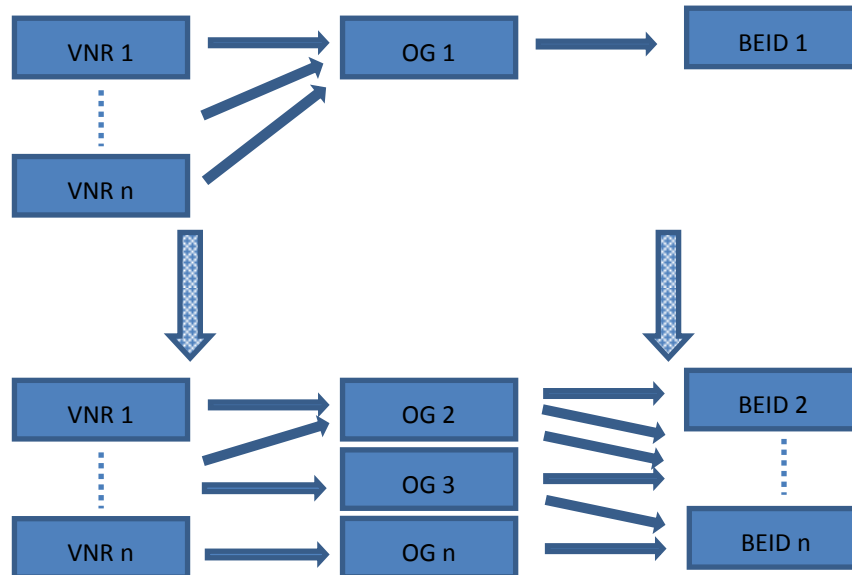


Koppelrelatie 3: Eén OG heeft **meerdere BEID's**, én er zijn **meerdere vergunningen** toegekend. In dit geval moet wederom worden beslist welke BEID's worden meegeteld zoals in koppelrelatie 2, maar kan bovendien worden gekeken of sommige vergunningen betrekking hebben op slechts een selectie van de gekoppelde BEID's. Dit vraagt aandacht en inzicht in de eenheden aan zowel de vergunningen kant als aan de kant van de BEID's.



Koppelrelatie 4: Bedrijven sites, onder andere Chemelot, Dow Benelux B.V., Shell Nederland Chemie B.V. (vestiging Moerdijk) en Yara Sluiskil. Hierbij kan sprake zijn van een **specifieke situatie** waarbij een reeks aan vergunningen worden geregistreerd onder één bedrijf met één

KvK. Hierbij is niet vanzelfsprekend dat per bedrijf gevestigd op een bedrijvensite ook een afzonderlijke VNR is aangevraagd. Er moet in dergelijke situatie handmatig worden uitgezocht welke BEID's actief zijn in een bepaalde activiteit op zo'n site, alsook op de andere relevante sites en uiteindelijk mogelijk welke vergunning daarbij hoort.



Op het moment dat bij een van de vier voornoemde koppelrelaties **een keuze** voorligt van welke BEID's mogen meetellen, spreken we van een **scenario**. Een scenario dat we kunnen laten afhangen van het type vervolgonderzoek. Hiertoe vertalen we deze keuzen naar een zestal scenario's:

Scenario A: De volledige OG opnemen met alle BEID's die daaraan verbonden zijn, deze procedure is eenvoudig want kan *geautomatiseerd* worden uitgevoerd. Deze meest uitgebreide selectie resulteert vanzelfsprekend in hoge uitkomsten voor de variabelen die later worden gekoppeld aan zo een activiteit.

Scenario B: Opnemen van een selectie van BEID's bij iedere OG die tot een bepaalde *afgebakende bedrijfstak* of *bedrijfstakgroep* (SBI 2-digit of 3-digit) behoren. Dit zijn de BEID's die vooral de fysieke productie voor hun rekeningen nemen en het overgrote deel van de bijbehorende emissies veroorzaken. Het betreft hier meestal de hoofdactiviteit van de OG. Alle BEIDs onder de OG die tot een andere SBI behoren vallen af. Vaak terugkerende voorbeelden van specifieke economische activiteit hierbij zijn, de groothandel en/of detailhandel onderdeel, het transportdeel van het bedrijf, de holdings en/of de conserndiensten. Hier kan ook worden gewerkt met een range aan bedrijfstakken, die vallen binnen een bepaalde logische bandbreedte, een *range van SBI's* (bv hetzelfde SBI Letterniveau). Ook deze procedure kan geautomatiseerd worden uitgevoerd.

Scenario C: Vergelijkbaar met scenario A. In eerste instantie een *geautomatiseerde* koppeling op OG, maar daarop worden vervolgens *handmatig correcties* toegepast, waarbij BEIDs worden uitgegeselecteerd in de range SBI's die de industriële activiteiten vertegenwoordigen. Dit kan eenvoudig door te sorteren op volgorde van SBI en de eerste, laagste mee te nemen. Dit levert doorgaans snel een plausibel resultaat voor de industrie.

Scenario D: Vergelijkbaar met scenario A. In eerste instantie een *geautomatiseerde* koppeling resulterend in maximale omvangcijfers omdat de complete OG wordt meegenomen. Hierop worden vervolgens *handmatig correcties* toegepast waarbij alle activiteiten die niet tot de kernactiviteit met aanzienlijke emissies behoren worden uitgegeselecteerd. Hierbij worden met name de SBI's in servicesectoren, zoals bij de handel en conserndiensten uitgegeselecteerd.

Scenario E: Geen automatische startkoppeling, maar alles handmatig koppelen op naam, adres, postcode, (schoorsteen) locatie, etcetera. Of eventueel alleen handkoppeling na eerst geautomatiseerde koppeling op basis van KvK uit te voeren. Dit scenario is erg arbeidsintensief en ligt daarmee niet voor de hand. Bovendien is het uiteindelijke resultaat niet noodzakelijkerwijs beter. Dit omdat bij de koppeling met het ABR en de vertaling naar de economische statistiek, zoals met controlevariabelen als productiewaarde of omzet uit de productiestatistieken, dat al veel houvast geeft voor de handmatige koppeling. Dat reduceert de kans op willekeurige allocatie.

Scenario F: Een van de voorgaande ontwikkelde scenario's met combinatie geautomatiseerd en handmatig gecorrigeerd, waarbij extra nadruk wordt gelegd op tijdsovergangen tussen de jaren. Grote overgangen in de tijd worden hier aanvullend geanalyseerd en waar relevant gecorrigeerd, zodanig dat dat recht doet aan de gewijzigde bedrijfssituatie voor wat betreft gebeurtenissen in de bedrijfsorganisatie ('events') zoals overname, fusie, opsplitsing of beëindiging.

Toelichting op de Scenario's

Deze set aan scenario's laat zien dat er verschillende mogelijkheden zijn om de koppeling uit te voeren. Bij vervolgonderzoek moet steeds worden gekeken welk scenario de beste gegevens oplevert, en zal zo nodig een alternatief scenario kunnen worden overwogen afhankelijk van de wensen van de gebruiker en de eisen die aan het onderzoek worden gesteld.

Bij de handmatige verbetering van de koppeling kan gebruik worden gemaakt van twee hulpvariabelen, omzet en CO₂ emissie, die in de koppeldatabase zijn opgenomen. BEIDs met activiteiten die voor CO₂-emissie niet relevant zijn kunnen worden uitgeselecteerd. Voorbeelden hiervan zijn handelsmaatschappijen, winkels/afdeling verkoop, personeelsverenigingen, etc. Door deze focus op de relevantie van bedrijfsonderdelen kan de relatie tussen economische data en emissiedata nauwkeuriger worden bepaald.

Voor bedrijven die uiteenlopende producten maken kan worden gekeken naar hoe de bedrijfsonderdelen gerubriceerd konden worden: bijvoorbeeld een bedrijf X dat zowel glas als blik maakt, deels op verschillende locaties, maar dezelfde naam draagt. Dit illustreert de keuze die moet worden gemaakt om delen van een bedrijf toe te rekenen aan een specifiek product en activiteit. Soms moeten keuzes worden gemaakt op basis van veronderstellingen. Resultaat is een bestand waarbij alle VNR's (al of niet gegroepeerd) zijn gekoppeld aan één of meerdere unieke BEIDs, zodat eenduidig beschrijfbaar economische eenheden ontstaan. Dit is het laagste statistisch beschrijfbaar niveau dat met de huidige koppeling mogelijk is (Zie ook paragraaf 2.6).

Toetsing van de scenario's

Voor de statistische eenheden behorende bij de 450 VNR's is het resultaat van de grotendeels geautomatiseerde bewerking uit Scenario B (op 2 digit SBI) getoetst. De koppelingen zijn handmatig gecontroleerd en er zijn correcties op gepleegd. Dit is gedaan voor ongeveer een derde van de VNR's, te beginnen bij degenen met de grootste CO₂-emissies. Deze selectie van statistische eenheden dekte ruim 95% van de CO₂-emissies in het Nederlandse deel van het EU ETS. Dit betrof zowel een kwaliteitscheck als een logicatetest. Voor deze handmatige koppelingen zijn deels openbare informatiebronnen gebruikt, zoals (detail)informatie van de KvK, Google en Google Maps, en achterliggende data. Anderzijds zijn meer gedetailleerde interne gegevens gebruikt waaronder die van het ABR. En in sommige gevallen gegevens uit de deels vertrouwelijke milieujarverslagen (e-MJV). Dit proces was bijzonder arbeidsintensief, maar heeft sterk bijgedragen aan de betrouwbaarheid van de informatie en inzicht opgeleverd in de sterkte van de automatische koppeling. De handmatige correcties op de koppelingen werden met goed resultaat toegepast.

Echter, afhankelijk van een toekomstige onderzoeksvragen kunnen in voorkomende gevallen ook andere keuzes wenselijk zijn. Voor inzicht in de problematiek rond het koppelen van VNR's aan statistische eenheden wordt in Bijlage III een hypothetisch voorbeeld gegeven.

4.3 Mogelijkheden voor verdere verbetering en uitbreiding van de koppeling

Bij de uitvoering van dit onderzoek bleken er nog mogelijkheden te zijn voor verbeteringen van de koppelingen.

Het beter en meer gebruik maken van emissies per schoorsteen uit het eMJV kan misschien helpen om de handmatige allocatie met ABR aan de hand van de vestigingslocatie en bijbehorende postcode verder te verbeteren. Mogelijk kan worden onderzocht of dat beter kan worden verbonden.

Het verder terugleggen van de koppelingen in de tijd kan in een vervolgonderzoek invulling aan worden gegeven. Dit biedt de mogelijkheid om bedrijven meer in de tijd te volgen. Per bron verschillen de mogelijkheden daarvoor. Idealiter leggen we hierbij terug tot voordat het EU ETS van start ging in 2005. Dat lijkt echter erg optimistisch. Terug tot aan 2010 en misschien 2008 (start 1^e commitmentperiode) biedt betere mogelijkheden.

De monitoring van zogenaamde gebeurtenissen ('events') in het ABR, geeft veel en nuttig inzicht in situaties van veranderingen bij bedrijven. Dat behoeft passende aandacht in geval van vervolgproject met uitbreiding van de tijdreeks. Het gaat om veranderende identiteit bij bedrijven bijvoorbeeld als gevolg van fusie, overname, opsplitsing, tijdelijke sluiting, reorganisatie).

Deze veranderingen kunnen voor een groot deel worden opgespoord met de events database, maar ook door trends te analyseren, trends van de variabelen in de gekoppelde statistieken en cijfers.

Het combineren en confronteren van data van de gekoppelde statistieken en databronnen die fysieke gegevens bevatten, biedt de mogelijkheid om de consistentie tussen de verschillende gekoppelde bronnen en data van bepaalde variabelen te toetsen. Dit als kwaliteitscheck.

5. Technisch resultaat van de koppeling

In dit hoofdstuk gaat het vooral over de vraag in welke mate het lukt om de ETS-vergunningen te koppelen aan het ABR. Besproken wordt wat daarbij een rol speelt en er wordt een overzicht gegeven van de koppelingsratio's met de andere statistieken.

5.1 Methode en koppelingsresultaat

5.1.1 Koppeling met het ABR

De werking van het ABR gegeven de relatie met de gekozen populatie is beschreven in paragraaf 2.3. In de paragrafen 4.1 en 4.2 is de systematiek voor koppeling van de ETS-populatie, in termen van vergunningen aan het ABR beschreven en ook het viertal koppelrelaties dat daarvoor nodig is. Vervolgens de scenario's die daarbij als keuze-instrument bij het gebruik van de data-infrastructuur kan worden ingezet. Bij de koppeling van de VNR's met het ABR, in dit geval via de OG-route voor de 2013 en 2014 data, resulteerde dit in de volgende koppelingen:

- a. Voor 109 vergunningen bleek bij elke unieke vergunning één unieke BEID uit het ABR te koppelen. In 95 gevallen had één unieke BEID meerdere vergunningen. Voor koppelrelatie 1 (a of b) zijn dus $109 + 95 = 204$ combinaties uniek, waarbij de unieke BEID één of meerdere vergunningen heeft. De hoofdactiviteit (SBI) van de BEID komt in deze gevallen ook overeen met die van de OG;
- b. De rest $465 - 109 - 95 = 261$ bestaat dus uit BEID'S, waarbij één (koppelrelatie 2) of meer vergunningen (koppelrelatie 3) bij meerdere BEID'S kunnen horen. Hierbij zitten ook enkele speciale gevallen zoals een enkele grote bedrijvensite, die tot koppelrelatie 4 behoort.

Door de aanpak via de ondernemingsgroep (OG) en de mogelijkheid om middels scenario's een keuze te maken van op te nemen BEID's is dit cijfer minder precies geworden. De precieze cijfermatige uitkomsten, ook na koppeling met de diverse statistieken, hangt nu mede af van de keuzen die tijdens het koppeltraject wordt gemaakt. Dat betreft de keuze tussen de '*ruime benadering*' of de '*enge benadering*' of zelfs een keuze daartussen en daarnaast om de scenariokeuze.

De ETS-vergunningen koppelen voor 100 procent aan het ABR via de OG-route, die wordt aangeduid als de '*ruime benadering*', is dus volledig voor 2013. Dit resultaat betekent dat in geval er behoefte is aan de complete set aan BEID's onder de gekoppelde OG, die BEID's dan allen koppelen. Zodra men scherper selecteert naar de exacte BEID met de ETS – emissies, wordt dit aangeduid als de '*enge benadering*'. Hierbij loopt dit percentage niet vanzelfsprekend terug maar komt de overeenkomst van de afbakeningen tussen enerzijds de VNR-eenheid en anderzijds de BEID of BEID's mogelijk in het geding en moet daar extra tijd en aandacht aan de kwaliteit van de koppeling worden besteed. Er is aanvullende expertise en inzicht van de individuele bedrijven nodig om de bepalende BEID, in sommige gevallen BEID's, voor de ETS emissies te selecteren, voornamelijk in de SBI categorieën 6 (delfstoffen) – 43 ('bouw').

Bij het pilotproject dat voorafging aan dit project, leverde de koppeling aan BEID's een afgebakend cijfer voor het 'succes van de koppeling', dit geeft echter geen volledig inzicht. Er kan daarbij evengoed incorrect of incompleet gekoppeld zijn (geen correcte match van de

geselecteerde BEID met de ETS emissies). Van beiden incorrect en incompleet zijn voorbeelden gevonden in dit project. Bij de koppeling via de OG-route is er weliswaar minder rechtstreeks een 'koppelsucces' of koppelpercentage te bepalen dat volledig inzicht geeft. Er is hierbij nu echter meer zekerheid dat de juiste BEID (met ETS emissies) wel onder de gevonden OG hangt. Dat dankzij zowel de anders uitgevoerde geautomatiseerde procedure als de handmatige checks en aanvullingen daarop als uitgevoerd in dit project. Het 'formele koppelsuccespercentage' kan bij de OG-route kan enigszins beïnvloed worden door het gekozen scenario. De OG-route geeft een kleinere kans om ergens in het proces de BEID's met emissies over het hoofd te zien. Dat is een positief gevolg van de gekozen aanpak met de koppeling initieel aan de OG. Vervolgens moet dan wel de correcte selectie plaats vinden van de reeks aan BEID's onder de gevonden OG. Daarvoor is dan aanvullend kennis en inzicht vereist bij het handmatige deel van het proces. Hierdoor neemt de kwaliteit toe en blijft een kleine kans over voor incorrecte allocatie naar BEID.

5.2 Koppelresultaten

Het onderzoek levert uiteindelijk twee koppelresultaten: De VNR-ABR koppeling en daarnaast de ABR - overige statistieken koppeling.

Voor de VNR-ABR koppeling volgt hier een samenvatting. De koppeling van de ETS-vergunningen aan het ABR via de OG-route, volgens de '*ruime benadering*' is voor 2013 en 2014 100 procent. Het vervolg van de koppeling (voor sommige onderzoeken) naar de juiste BEID met de ETS emissies de '*enge benadering*' kan deels geautomatiseerd. Echter voor een deel is aanvullend werk en expertise nodig voor checks te controle en soms correctie. Dit bepaalt uiteindelijk het succes van de koppeling van de VNR aan de juiste BEID met de ETS emissies.

De koppeling per BEID kan nu kwalitatief beter zijn bij de OG-route ten opzichte van de BEID route omdat er minder snel BEID's niet worden geïdentificeerd. Daartoe is dan wel een vereiste dat de selectie geautomatiseerd en deels handmatig zo goed mogelijk plaatsvindt. Dat blijft een vereiste voor een kwalitatief goed koppeling. De beoordeling van het koppelsucces ligt vooral in de correcte overeenkomstige afbakeningen tussen VNR en de eenheid in het ABR, zowel de OG als BEID en dient ook kwalitatief beoordeeld te worden. Het koppelpercentage biedt hier niet het volledig inzicht in.

Het tweede koppelresultaat van de ABR-overige statistieken koppeling is afhankelijk van de eerste koppeling en daarbij ook van het gekozen scenario. Hoe meer eenheden (en dat is doorgaans meer kleine eenheden), hoe lager het koppelpercentage. Voor het absolute aantal gekoppelde eenheden maakt het weinig verschil, de grote bedrijven koppelen frequenter aan de diverse statistieken. Een groot deel van de ETS bedrijven blijkt een TOP 2000 bedrijf, waarbij de OG behoort tot de geselecteerde groep van 2000 grootste bedrijven in Nederland. Bedrijven worden hiervoor doorgaans geselecteerd op basis van de werknemersaantallen en bedrijfstak. De figuur geeft het aantal BEID's weer, waarbij het hoge totaal cijfer laat zien dat dit het resultaat is van de koppeling naar OG, kortom van de '*ruime benadering*' en dus nog met alle onderliggende BEID's. Waarbij ook de BEID's met een kleine omvang minder vaak voorkomen in de verschillende statistieken. Ondanks dat zijn er redelijke koppelpercentages. Een scherpe selectie met primair een focus op de doorgaans grote BEID's mét de ETS emissies en bijvoorbeeld in de emissie-intensieve industrieën zal aanzienlijk hogere percentages opleveren. Voor een uitgebreidere verhandeling en analyse van resultaten verwijzen we naar het belerende rapport met de beschrijvende statistiek.

Figuur 5.1. Koppeling ETS-bedrijven aan andere CBS-statistieken ('ruime benadering')

	aantal BEID	Koppel %
Totaal	872	100
Polisadministratie met hun economische activiteit en grootteklasse	788	90
Arbeidsrekeningen	732	84
Productie Statistieken	514	59
Statistiek verkopen industriële producten naar productgroep (PRODCOM)	238	27
Nationale investeringen in (im)materiële vaste activa bij niet-financiële bedrijven	684	78
Internationale Handel in Goederen	803	92
Research Technological Development (RTD)	493	57
Community Innovation Survey	330	38
Energie statistieken	268	31
Milieukosten van bedrijven	245	28

Figuur 5.1 toont volgens de 'ruime benadering' gekozen populatie het absolute aantal BEID'S dat gekoppeld kon worden aan de diverse CBS-statistieken en het resulterende koppelpercentage ten opzichte van de ETS – populatie gemeten in BEID's. Dit levert hier in termen van koppelingspercentages een behoudend resultaat op, door het relatief grote aantal gekoppelde BEID's. Naarmate men scherper zou selecteren bij het aantal BEID's dat is meer richting de 'enge benadering' en bijvoorbeeld handel en holdings- en conerndiensten zou uitsluiten dan neemt het aantal gekoppelde BEID's niet of slechts in beperkte mate af. Er zijn dan vooral minder (economisch) kleine eenheden, die vaak niet koppelen en dus geen invloed hebben op het aantal. Grote eenheden koppelen doorgaans beter waarmee naar verwachting de koppelpercentages zullen toenemen. De energie- en milieukostenstatistieken tonen een relatief laag koppelresultaat. Bij deze statistieken ligt de nadruk doorgaans evengoed op grote bedrijven, maar is er vaker sprake van een steekproef met beperkte omvang. Een deel van de bedrijven wordt dan niet waargenomen.

Overigens toont figuur 5.1 het koppelpercentage aan de eenheden in de statistiek. De koppeling per variabele kan nog apart worden gemeten en beoordeeld evenals de scores per bedrijfstak. Zie hiervoor het rapport met de beschrijvende statistiek.

Bij de tellingen en analyse komen sommige bedrijven vaker voor, hierbij moet vanzelfsprekend worden gewaakt voor dubbeltelling. Dit geldt overigens niet voor de gegevens getoond in de tabel.

6. Gebruik, beheer en onderhoud data-infrastructuur

Dit hoofdstuk beschrijft het gebruik, beheer en onderhoud van de data-infrastructuur en de software die de koppeling tussen de vergunningsnummers en de CBS-data tot stand brengt.

6.1 Data-infrastructuur

De data wordt opgeslagen op de CBS-servers. Het betreft hier de basisinfrastructuur binnen de beveiligde omgeving van CBS. Hiervan wordt regelmatig een back-up gemaakt.

Ten behoeve van de datakoppeling en vervolgonderzoek is de volgende mappenstructuur aangemaakt:

Figuur 6.1. Mappenstructuur gekoppelde ETS-bedrijven

Map	Inhoud
...\Onderzoek_NEA_CBS\	Hoofdmap
...\Onderzoek_NEA_CBS\DATA\NEA	Data afkomstig van de NEA
...\Onderzoek_NEA_CBS\DATA\CBS	Data afkomstig van CBS
...\Onderzoek_NEA_CBS\DATA\CBS\ABR	Integratiekaders ABR per jaar
...\Onderzoek_NEA_CBS\Koppeling_NEA_CBS	Software Access en Excel aanmaak Basis-Koppelbestand tussen vergunningsnummers en CBS-bedrijfseenheden.
...\Onderzoek_NEA_CBS\Koppeling_ABR_CBS	Software SPSS voor koppelen CBS gegevens aan het Basiskoppelbestand.
...\Onderzoek_NEA_CBS\Analyse\...	Map voor verdere analyseprojecten
...\Onderzoek_NEA_CBS\Documentatie	Documentatie en rapportage basiskoppeling

Daarnaast wordt binnen CBS steeds meer data aangeboden via het Data Service Center (DSC) waar CBS-data wordt opgeslagen in een standaardformaat voorzien van metadata en documentatie. Binnen het DSC wordt ook de update van deze bestanden gewaarborgd. Het streven is om hier zoveel mogelijk gebruik van te maken omdat het beheer en onderhoud dan ook goed is geregeld volgens een gestandaardiseerde aanpak.

Overigens komen via Remote Access steeds mee microbestanden beschikbaar bij CBS. Deze data is intern ook beschikbaar voor analyses binnen CBS.

6.2 Gebruik

Toekomstige analyses worden bij CBS uitgevoerd vanwege de gevoelige en vertrouwelijk beschikbaar gestelde informatie op bedrijfsniveau waarop de analyses moeten worden uitgevoerd. Onthulling van deze gegevens is niet toegestaan. Hierbij dient het basis koppelbestand zoals in hoofdstuk 4 beschreven steeds het uitgangspunt, waarbij voor elk onderzoek moet worden nagegaan hoe sommige koppelingen in detail gemaakt moeten worden. Dit is afhankelijk van het doel en de vragen in het onderzoek.

Voor het gebruik van het Basiskoppelbestand en andere CBS-microdata zijn er de volgende mogelijkheden:⁶

- Aan het CBS wordt een opdracht gegeven voor een statistisch onderzoek op maat. Het onderzoek wordt uitgevoerd door CBS;
- Het onderzoek wordt uitgevoerd in een samenwerkingsverband tussen CBS en bijvoorbeeld NEA.

6.3 Beheer en onderhoud

Het beheer van de data en software wordt uitgevoerd door CBS binnen het team milieurekeningen van de divisie Nationale Rekeningen.

6.3.1 Beheer en onderhoud koppelbestand

Jaarlijks zal een nieuw koppelbestand worden gemaakt, gebaseerd op de laatste set met vergunningsnummers en met bijbehorend KVK-nummer te verkrijgen van de NEa en het laatste integratiebestand van het ABR bij CBS. Hieraan worden, indien al beschikbaar, ook de omzetgegevens uit de productiestatistiek van het CBS (PS) en de emissies afkomstig uit het NEa-bestand met vergunningsnummers gekoppeld. De laatste twee variabelen worden gebruikt bij de koppelingsacties en dienen als keuze-criterium voor de juiste koppeling. Dit koppelbestand wordt toegevoegd aan de tijdreeks van eerder gemaakte basiskoppelbestanden, zodat verschillen over de jaren heen kunnen worden bekeken. Hiermee wordt een koppeling over meerdere jaren verkregen tussen de vergunningsnummers van de NEa en de bedrijfseenheden uit het ABR van CBS.

6.3.2 Beheer en onderhoud software

Voor de koppeling van de NEa-vergunningnummers aan de CBS-bedrijfseenheden wordt gebruik gemaakt van een Access database waarin met behulp van query's een eerste koppeling tot stand wordt gebracht. Daarna wordt handmatig de koppeling gecheckt en verrijkt, waarna deze wordt toegevoegd aan de tijdreeks. Het onderhoud bestaat uit het bijwerken van de benodigde data ten behoeve van het koppelproces en eventueel het converteren van de software naar andere platformen indien dit in de toekomst nodig zou blijken.

Voor de koppeling van de reeks aan het basiskoppelbestand wordt gebruik gemaakt van SPSS. Het voordeel is dat bij vervolganalyses ook statistische functies en de mogelijkheid voor het maken van rapporten ter beschikking staan als standaard tools in SPSS. De benodigde scripts worden onderhouden door het Centrum voor Beleidsstatistiek (CvB) van CBS.

6.3.3 Beheer en onderhoud CBS-data

CBS-data zijn afkomstig van meerdere statistieken en afzonderlijke waarnemingsprocessen binnen CBS. Bij elk proces wordt gezorgd voor validatie en opslag van de gegevens. Steeds meer zal hiervoor gebruik worden gemaakt van het Data Service Center (DSC), waar zowel langdurige dataopslag als de beschikbaarheid van een goede beschrijving wordt gegarandeerd. Voor

⁶ Voor meer informatie over de mogelijkheden zie <https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata>.

vervolgonderzoek zal waar mogelijk gebruik worden gemaakt van data uit het DSC. Daarnaast bezit het CvB veel microdatabestanden in eigen beheer, waarvan gebruik kan worden gemaakt.

7. Conclusies en aanbevelingen

7.1 Conclusies

De onderzoeksvraag

Vanuit de gedachte dat het interessant kan zijn om data van NEa te combineren met CBS-data, hebben de NEa en CBS een pilot uitgevoerd om de koppelmogelijkheden te verkennen. De pilot toonde aan dat de koppeling mogelijk zou zijn maar dat vervolgonderzoek nodig was om de data-infrastructuur daadwerkelijk op te zetten en na te gaan hoe de koppeling zo goed mogelijk kan worden gerealiseerd met het oog op toekomstig onderzoek.

In samenspraak en met steun van het Ministerie van IenM hebben de NEa en CBS dit vervolgonderzoek naar koppelmogelijkheden uitgevoerd. Onder dit door het Ministerie van IenM bekostigde 'Anders omgaan met data project' is onderzocht wat aan koppelingsrelaties kan worden gerealiseerd.

De eerste belangrijke onderzoeksvraag, of een werkende data-infrastructuur kan worden opgezet, is gerealiseerd. De NEa-data kan worden gekoppeld aan de BEID's in het ABR. Vervolgens zijn data van andere statistieken gekoppeld voor 2013 en 2014 zoveel als mogelijk was.

De tweede belangrijke onderzoeksvraag was hoe de koppelingssystematiek verder kon worden ontwikkeld met het oog op de mogelijkheden voor toekomstige analyse en voortgaande verbetering daarvan ten opzichte van de pilotresultaten. Ook die vraag kan positief worden beantwoord. We kennen beter de mogelijkheden en beperkingen van het systeem voor datakoppeling en hebben mede vanwege het nieuwe kader van 'Anders omgaan met data' ook voor een afwijkende aanpak gekozen. Anders ten opzichte van de aanpak bij de pilot en ook anders dan de aanpak die doorgaans wordt gehanteerd bij CBS voor koppeling van microdata.

De wens van NEa, Ministerie IenM en CBS om jaarlijks nieuwe informatie aan de data-infrastructuur toe te voegen en zo een reeks terug in de tijd te maken en verder in de toekomst op te zetten, stuit ook niet op technische bezwaren.

Databeveiliging en geheimhouding

De noodzaak om ten behoeve van het project een procedure te vinden voor veilige uitwisseling van data, rekening houdend met de geheimhoudingsregels van beide betrokken organisaties kon eveneens passend worden ingevuld. De data van de NEa kon via de beveiligde dataverbinding worden verkregen bij de NEa, ondanks dat uiteindelijk geen beveiligde NEa data is gebruikt. Ook bij data van enkele andere organisaties en organisatieonderdelen bij CBS wordt met een beveiligde dataverbinding gewerkt. Er waren daarbij geen grote beperkingen om binnen de beveiligde omgeving van CBS te werken met koppeling van alle beoogde databronnen. Wel vergde het soms wat extra tijd en aandacht om langs een vorm van bescherming te komen.

Koppelingsmogelijkheden

De kern van het onderzoek, de koppelingsmogelijkheden van de NEa-data aan de CBS-data, meer specifiek aan het Algemeen Bedrijfsregister bij CBS, blijken goed. Alle bedrijfseenheden met emissievergunningen in de NEa-boekhouding kunnen worden verbonden met de CBS-data. Dit ten minste op het niveau van de ondernemingsgroep (OG). Alle eenheden blijken dus

identificeerbaar. Gezien de omvang van de ETS-bedrijven mocht dat worden verwacht, maar het is prettig als dat ook daadwerkelijk lukt.

De tweede beoordeling wat betreft de koppelingsmogelijkheid gaat over de vraag of de koppeling kan worden uitgevoerd tussen de bedrijfsdata van de NEa en CBS, zodanig dat die resulteert in dezelfde afbakening van ieder specifiek bedrijf in de populatie. Daarbij waren de bevindingen wat genuanceerder. Een deel van de NEa-vergunningshouders (VNR's) kan naadloos worden gekoppeld aan een ETS-vergunning, dit betreft doorgaans helder afgebakende bedrijven op een of een enkele productielocatie en bijvoorbeeld niet opererend onder een grootschalige en complexe holdingstructuur. Voor een ander deel is meer maatwerk vereist.

Hoe tenslotte resultaten van maatwerk eruit kunnen zien, hangt af van de onderzoeksvraag bij potentieel vervolgonderzoek. Als alle BEIDs onder de OG's die worden gekoppeld dat wil zeggen volgens de *'ruime benadering'*, onderdeel mogen uitmaken van de data analyse dan krijgt men hogere waarden bij de resultaten dan wanneer een koppeling zo beperkt mogelijk op BEID-niveau wordt gerealiseerd volgens de *'enge benadering'*.

De koppelingsmogelijkheden overstijgen eerdere verwachtingen bij de start van het project.

Database

Bij CBS staat nu een database van het ABR voor 2013 en 2014 met daaraan gekoppeld het bestand met de populatie VNR's van NEa en aanvullende informatie. Dit is tot stand gebracht door middel van een serie aan query's, zodat deels geautomatiseerd en deels handmatig de VNR's op OG- en op BEID-niveau worden gekoppeld. Vervolgens is aan deze basisselectie van ETS-eenheden in het ABR als ruggengraat in het koppelingssysteem, een serie aan bronstatistieken met microdata verder gekoppeld. In voorkomende gevallen moesten voor het handmatige deel van de koppelingen keuzes worden gemaakt op basis van informatie en kennis van de bedrijven. In de database is gemarkeerd waar dit het geval was en welke motivatie ten grondslag lag aan een keuze. Dit kan een keuze betreffen aangaande de locatie van de (fysieke) productie van een bedrijf of de aard van de productie op een locatie. In het geval dat er nieuwe keuzes worden gemaakt op basis van gewijzigde veronderstellingen, werkt dit door in de uitkomsten. Dit maakt dat de database voor verschillende doeleinden bruikbaar zal zijn en voor verschillende typen onderzoeksvragen kan worden ingezet.

Beheer en onderhoud data-infrastructuur

Voor behoud van de ontwikkelde data-infrastructuur in Access en deels in SPSS is beheer en onderhoud nodig op de opgezette mappenstructuur op het CBS-netwerk. Voor uitbreiding van de data-infrastructuur en de daarvoor benodigde databronnen en de structurele levering daarvan door de verschillende CBS-afdelingen is het nodig dat er vervolgsafspraken worden gemaakt. Dit betreft afspraken met de relevante stakeholders buiten CBS zoals met de NEa en met de ministeries IenM en EZ over de inrichting en over hoe die uitbreiding eruit gaat zien, onder andere welke dimensies die moet hebben. De timing van de benodigde leveringen dienen in geval van vervolg ook nader te worden vastgelegd.

Technisch resultaat van de koppeling

Het tweede belangrijke resultaat van het project is dat er na de geautomatiseerde en handmatige koppeling een infrastructuur ligt die de NEa-vergunningnummers koppelt aan ABR en andere statistieken, zowel op OG (*'ruime benadering'*), als op BEID niveau (*'enge benadering'*). Daarnaast zijn selecties mogelijk naar gelang de breedte van de gewenste informatie over betrokken economische activiteiten (bedrijven). Er is keuze mogelijk in aantallen in een range van minimaal 550 en maximaal 900 gekoppelde BEID's, niet een enorm verschil. De grofweg 900 BEID's in het ABR van de *'ruime benadering'* dekken naar verwachting

geheel of praktisch geheel de ETS-vergunningen. Van het verschil van 350 BEID's die uitgeselecteerd moeten worden, spreekt een groot deel voor zich. Dit omdat deze in een voor de omvangrijke ETS emissies uitzonderlijke SBI zitten en/of omdat de fysieke productie met veel CO₂-emissies eenvoudigweg niet kunnen plaatsvinden op bepaalde locaties. Dit levert daarom een behoorlijke zekerheid over de uitgeselecteerde en uiteindelijk over de overblijvende BEID's in de 'enge benadering'.

Anders omgaan met data

Dit project heeft naast de data-infrastructuur opzet de mogelijkheid geboden *anders met micro-datakoppeling om te gaan* dan doorgaans gebruikelijk is. Dit door een afwijkende koppelroutine te hanteren, namelijk door de koppeling in eerste instantie uit te voeren op OG-niveau ('ruime benadering'). Met het project kunnen aan alle VNR-eenheden met hun gekoppelde KVK, een corresponderende OG worden toegekend en daarmee ook de relevante onderliggende BEID's worden verbonden. Dat dankzij de ontwikkelde data-infrastructuur met verbinding aan het ABR. Dit betekent dat alle ETS-bedrijven, kortom alle VNR's identificeerbaar zijn gebleken in het ABR. Dat maakt dat het niet nodig was om bijvoorbeeld gebruik te maken van de semantiek van bedrijven voor deze koppeling. Alleen was het nodig om tijdens het project een passende route te ontwikkelen voor het uitselecteren van onbedoelde BEID's. Dat is gedaan door enerzijds een aantal keuzemogelijkheden te introduceren bij de processtappen middels de zogenaamde scenario's, anderzijds door handmatig checks en correcties uit te voeren op de in eerste instantie geautomatiseerd gekoppelde data.

Vanuit het perspectief van het programma 'Anders omgaan met data' door het Ministerie van IenM lijkt er reden om de aanpak om met een frisse en alternatieve aanpak naar (bestaande) data te kijken vaker te hanteren. Terwijl bij de experts bij betrokken organisaties al langer de wens bestond om dit project te doen, is dat niet eerder gebeurd. Voor een dergelijke, meer creatieve omgang met data is vaak geen ruimte. Een les uit dit 'Anders omgaan met data' project is dat wanneer expliciet tijd en capaciteit wordt besteed om met zowel een creatieve als deskundige blik naar nieuwe combinaties van bestaande datasets te kijken, nieuwe informatie wordt gegenereerd. Informatie die bruikbaar zijn voor verdiepende analyses en uiteindelijk voor beleid en uitvoering.

Vervolg op project

Het project heeft een aantal producten en inzichten opgeleverd die reden geven om samen met belanghebbende en geïnteresseerde organisaties te bekijken of het koppelproduct kan worden uitgebreid in scope en waar mogelijk met verdergaande aandacht voor kwaliteit van de koppelingen.

Een conclusie is ook dat bij ieder mogelijk vervolgproject een deel van de benodigde capaciteit besteed moet worden aan de verdere verbetering van kwaliteit van de koppeling. De mate en vorm waarin is dan afhankelijk van de vragen in het vervolgproject. Dat is in elk geval nodig bij de analyse van overgangen in de tijd en de opvallende verschijnselen die daarbij kunnen optreden. Een aanvullend deel van de populatie VNR's handmatig beoordelen en waar nodig corrigeren op de toegekende BEID, gemiddeld met geringere ETS-emissies dan de eerder handmatig beoordeelde bedrijven past daar in.

7.2 Aanbevelingen

Voor vervolg op dit onderzoeksproject benoemen we hier een aantal aanbevelingen. Dit betreft verbeteringen en/of mogelijke uitbreidingen van de datakoppeling en voor gebruik van de gerealiseerde koppeling.

1. Microbenadering

Dit onderzoeksproject richt zich op koppeling van de data van de ETS-bedrijven op microniveau (dus op bedrijfsniveau). Op dit niveau is informatie beschikbaar over bijvoorbeeld bedrijfstak en grootteklasse, wat de mogelijkheid biedt om informatie te aggregeren naar meso-niveau. Daarmee is aggregatie mogelijk naar ieder gewenst aggregatieniveau, bijvoorbeeld naar gewenste bedrijfstakken volgens de standaard bedrijfstakindeling, de SBI2008, bijvoorbeeld op het gewenste SBI 4-digit niveau. Op die manier kunnen met deze gegevens doorkijkjes worden gemaakt van de ETS-sector, zodat bijvoorbeeld het zwaartepunt aan ETS-emissies bij bepaalde bedrijfstakken inzichtelijk gemaakt kunnen worden. Met daarbij de mogelijkheid om aanvullende inzichten te verkrijgen in bedrijfsomvang, regio, energiegebruik of zeggenschap.

2. Tijdreeks, uitbreiding scope met oudere en nieuwere jaren

In deze studie van NEA-CBS is de koppeling gerealiseerd met de ETS-bedrijven voor verslagjaren 2013 en 2014. De aanbeveling is om een uitgebreidere tijdreeks te maken voor het recentste jaar, in dit geval 2015, in statistisch jargon ook wel T-1 genoemd en daarnaast verder terug in de tijd, in elk geval terug naar 2012 vanwege het einde van de tweede fase van het EU ETS (de Kyoto commitment periode) die liep van 2008 t/m 2012. Zo mogelijk terug in de tijd tot en met het begin van de tweede fase in 2008.

Dit biedt meer mogelijkheden voor het monitoren van de ETS-bedrijvendynamiek en voor het meten en analyseren van de economische en milieuprestaties in relatie met het EU ETS.

3. Gebruik van 'eventsdatabase'

Voor het nagaan van de dynamiek in de populatie wordt aanbevolen gebruik te maken van de 'events-database' binnen het ABR. De monitoring van 'gebeurtenissen' (events) in het ABR geeft veel en nuttig inzicht ingeval van veranderende situaties bij bedrijven in de tijd. Dat behoeft passende aandacht in geval van een vervolgproject met uitbreiding van de tijdreeks. Het gaat daarbij om veranderingen bij bedrijven als gevolg van bijvoorbeeld fusie, overname, opsplitsing, tijdelijke sluiting.

4. Samenstellen van controlegroep voor vergelijkend onderzoek met non-ETS

Een volgende uitbreiding op het onderzoek zou kunnen zijn het samenstellen van een zogeheten controlegroep bedrijven. Een controlegroep bestaat uit bedrijven die zoveel mogelijk lijken op de ETS-populatie van circa 350 ETS-bedrijven. Gelijkenis wordt nagestreefd in de soort economische activiteit (SBI), in de omvang (grootteklasse), economische omvang, enzovoort. Deze zouden in de data-infrastructuur verwerkt kunnen worden om analyse mogelijk te maken tussen ETS-bedrijven en bedrijven die daarop lijken, maar die niet deelnemen aan het EU ETS. De selectie van de controlegroep is zeker niet vanzelfsprekend. Uit buitenlands onderzoek bestaan wel voorbeelden van hoe dergelijk onderzoek werd uitgevoerd en hoe de controlegroep tot stand kwam. In de matchingsvariabelen kan worden geprioriteerd.

5. Kwaliteit en onzekerheid van de resultaten

We hebben in dit project en deze rapportage gekozen voor de werkgelegenheidscijfers als steunvariabele. Dit is met het gegenereerde datamodel later eenvoudig te herhalen met andere

economische steunvariabelen, zoals productiewaarde en toegevoegde waarde. Hiermee kan later het onderzoek worden uitgebreid en getoetst op het effect van de steunvariabele op de CO₂-emissies en inzicht worden verschaft in de bandbreedte van resultaten.

Een aanbeveling is om in een mogelijk vervolgproject passend capaciteit te besteden aan consolidering en zo mogelijk verdere verbetering van de kwaliteit van de koppeling en daarbij aan de gekozen afbakeningen van de bedrijven. Het accent daarbij hangt af van het vervolgproject. Een belangrijke aanbeveling is om voor 2014 en voor mogelijk nog een sleuteljaar een extra deel van de populatie VNR's handmatig te beoordelen, de geautomatiseerde toekenning te checken en waar nodig te corrigeren. Dat dan ook voor de bedrijven met gemiddeld geringere ETS-emissies.

6. Verbinden met zeggenschapsinformatie / eigendomsinformatie

Het verbinden met zeggenschapsinformatie op bedrijfsniveau bij vervolgonderzoek biedt bijvoorbeeld een mogelijkheid voor het analyseren van het aandeel buitenlandse zeggenschap van de Nederlandse ETS-bedrijven en waar die zeggenschap geografische vooral is gesitueerd. Dit biedt mogelijkheden voor analyse, bijvoorbeeld in termen van geopolitieke consequenties.

Bronnen

AER, 2014. Ralf Martin, Mirabelle Muûls, Laure B. de Preux, and Ulrich J. Wagner, 2014. Industry Compensation under Relocation Risk: A Firm-Level Analysis of the EU Emissions Trading Scheme. *American Economic Review* 2014, 104(8): 2482–2508.

CBS, 2016a. CBS-NEa datakoppeling ETS - Notitie fase 1. 15215 Datakoppeling ETS-bedrijven. Cor Graveland, Albert Pieters, Kathleen Geertjes en Esther de Heij. 13 april 2016 (interne notitie). CBS, Heerlen/Den Haag.

CBS, 2016b. Datakoppeling ETS-bedrijven - Beschrijvende Statistiek. Cor Graveland, Kathleen Geertjes en Albert Pieters. CBS, Heerlen/Den Haag.

CBS, 2015a. Notitie NEa-CBS koppeling, 29 januari 2015 (vertrouwelijk). CBS, Heerlen/Den Haag.

CBS, 2015b. Memo Voorstel CBS: Meten prestaties ETS bedrijven, milieutechnisch en economisch. In samenwerking met NEa. 20 februari 2015. CBS, Heerlen/Den Haag.

CBS, 2012. Statistische eenheden in de institutionele statistieken. Rapport. CBS, Heerlen/Den Haag.

Maarten van Rossum, Cor Graveland, Sjoerd Schenau, Bram Edens, 2013. Ch.4. 'Sustainability: according to the quantifying CO2-emissions control-criterion'. *Internationalisation Monitor 2013*. Statistics Netherlands, Heerlen/The Hague.

Petrick, S., Wagner, U. (2014). *The Impact of Carbon Trading on Industry: Evidence from German Manufacturing Firms*. Kiel Working Paper, 1912, Kiel Institute for the World Economy, Kiel, 52 pp.

Statistics Netherlands, 2014. *Environmental accounts of the Netherlands 2013*, Heerlen/The Hague.

Bijlagen

Bijlage I: Samenvatting pilot koppeling NEa- en CBS-eenheden 2015

	Naam variabele	Aantal
Aantal records (Na koppeling ABR)	-	552
Aantal bedrijven (ETS-vergunningen)	VNR	457
Aantal KvK-nummers	KvK	274
Aantal unieke bedrijfseenheden	BE_ID	287
KvK * BEID	KvK * BE_ID	312

De tabel toont dat de NEa 274 KvK-nummers had aangeleverd. Bij deze 274 KvK-nummers werden 287 verschillende, unieke, beids gevonden. Daarbij bleek reeds dat sommige KvK-nummers gekoppeld konden worden aan meerdere beids.

Deze relatie werkt echter ook de andere kant op. Eén unieke beid kan namelijk ook koppelen aan meerdere KvK-nummers. Daarom staat in bovenstaande tabel ook een kruising tussen KvK en beid. De kruising toont in hoeverre er een één op één relatie is tussen KvK-nummer enerzijds en beid anderzijds. De kruising leverde een uitkomst van 312 op, wat betekent dat er bij maximaal 35 bedrijfseenheden (312 minus 287) een probleem is bij de match van KvK-nummer en beid.

De 287 unieke beids zijn gekoppeld aan verschillende statistieken. Per statistiek kan dus maximaal 287 keer gekoppeld worden. Een aantal van de CBS-data is echter gegenereerd door middel van een enquête en is dus steekproefsgewijs. Dit betekent dat sommige records niet kunnen worden gekoppeld aan één of meerdere statistieken. Van de 287 unieke beids zijn er 3 (1%) die aan geen enkele statistiek kunnen worden gekoppeld. Aan de andere kant zijn er 118 beids (41%) die aan alle statistieken gekoppeld kunnen worden.

Bijlage II: Praktische aspecten van data-organisatie en data-structurering

NEA Mappenstructuur: Organisatie per verslagjaar

Er is voor gekozen de data, bestanden en databasekoppeling afzonderlijk per verslagjaar te organiseren. Hiervoor is de volgende mappenstructuur aangemaakt op de CBS-servers:

/NEA_Productie/JJJJ/Koppelbestanden/	map met o.a. 'Basis_Koppelbestand_NEA_CBS_JJJJ'
/NEA_Productie/JJJJ/Werkbestanden/	map met aan het basis_koppelbestand gekoppelde microdata
/NEA_Productie/JJJJ/Analyse/	map voor analyses
/NEA_Productie/JJJJ/Productiestatistiek/	mappen met microdata
/NEA_Productie/JJJJ/IntHandel/	mappen met microdata
....	

Bestanden met microdata voor analyse bevatten in elk geval een BEID als koppelvariabele, zodat koppeling met het 'Basis_Koppelbestand_NEA_CBS' mogelijk is. In de map werkbestanden worden enkele bestanden geplaatst waarin microdata zijn gekoppeld aan het Basis_Koppelbestand. Te denken valt bijvoorbeeld aan een werkbestand met daaraan gekoppeld een aantal basisgegevens per BEID zoals productiewaarde, emissie en handelsomzet.

Ruggengraat: VNR – KvK – OG – BEID

De eerste en belangrijkste stap is het koppelen van de ETS-Operator bedrijven aan het Algemeen Bedrijvenregister (ABR) van CBS. Dit vormt de ruggengraat van het koppelsysteem. Hiervoor wordt gebruik gemaakt van het vergunningsnummer en bijbehorende nummer van de Kamer van Koophandel uit de gegevens van de NEA en het KvK-nummer, OG-nummer en BEID uit het ABR van CBS. De bestanden worden verzameld, verwerkt en opgeslagen in Microsoft Access.

Hiervoor zijn de volgende afspraken gemaakt:

- Van de NEA ontvangt CBS jaarlijks een bestand met Vergunningsnummers en het bijbehorende KvK-nummer zoals dat bekend is bij de NEA.
- Door de sector ERB van CBS wordt jaarlijks een ABR-Integratiekader geleverd met daarin in elk geval per OG alle BEIDs en per BEID alle bekende KvK nummers. Daarnaast bevat het bestand algemene kenmerken van elke BEID en OG zoals SBI, Grootteklasse en NAW gegevens.
- Na koppeling van deze twee bestanden (Deels automatisch en deels handmatig, zie paragraaf 4.1 en 4.2 wordt dit 'Basis_Koppelbestand_NEA_CBS' opgeslagen.

De datastructuur van het 'Basis_Koppelbestand_NEA_CBS' bevat (minimaal) de volgende variabelen:

- Verslagjaar
- NEA_VNR
- NEA_VNR_KvK
- ABR_OG
- ABR_OG_SBI4digit
- ABR_BEID
- ABR_BEID_SBI4digit

- h. ABR_KvK
- i. NEA-CBS_Analyse_Eenheid = OG+Volgnummer

Variabele i: Een **analyse eenheid** is een cluster van één of meer bedrijven gekoppeld aan één of meer emissievergunningen (en geregistreerd als ETS – operatorbedrijf). Dit is een combinatie van VNR's en BEIDs die samen één beschrijfbare eenheid vormen en voor analysedoeleinden altijd als (kleinste) eenheid moet worden behandeld. Zie paragraaf 4.3 voor de Koppelstrategieën en paragraaf 4.2 voor achtergrondinformatie over het koppelen. Aangezien een elke VNR altijd uniek aan één OG kan worden gekoppeld is besloten het ABR_ OG + een volgnummer te gebruiken als unieke identificatie van de analyse eenheid.

Het '**Basis_Koppelbestand_NEA_CBS**' wordt opgeslagen als Microsoft ACCESS bestand onder de naam '**Basis_Koppelbestand_NEA_CBS_JJJJ.accdb**' waarbij JJJJ het verslagjaar is. Een werkbestand kan daarnaast als '.csv' bestand voor algemeen gebruik, of als '.sav' bestand voor SPSS worden opgeslagen.

Ten behoeve van vervolgonderzoek kunnen gegevens aan dit basisbestand worden gekoppeld op basis van het NEA_ VNR nummer als de gegevens van bijvoorbeeld de NEA afkomstig zijn, of aan het ABR_BEID nummer als de gegevens van CBS afkomstig zijn.

CBS-Gegevens

Productiestatistiek (PS)

PS bestanden worden jaarlijks door ... in geleverd en in de NEA-mappenstructuur geplaatst. Hierin staan per BEID alle gegevens die zijn ingevuld in de PS-Vragenlijst.

LET OP: Dit zijn gaafgemaakte gegevens zoals die gebruikt worden bij de productiestatistiek. Nadat ze op geaggregeerd niveau zijn ingezet in de Nationale Rekeningen worden deze cijfers nog ingepast. Ook worden de cijfers uit de PS aangepast aan het ingezetene begrip van de Nationale rekeningen (o.a. bijschatting zwartwerk, illegaal, correctie omzet Nederlanders in het buitenland (+) en omzet Buitenlanders in Nederland (-/-). Deze in- en aangepaste gegevens zijn niet op BEID niveau beschikbaar. Enkel productie, verbruik en toegevoegde waarde zijn wel op BEID niveau beschikbaar gecorrigeerd voor omzet Nederlanders in het buitenland (+) en omzet Buitenlanders in Nederland (-/-), echter zonder schatting voor zwartwerk en illegaal. Hiermee dient rekening te worden gehouden bij analyses waarbij cijfers van de Nationale Rekeningen worden vergeleken met cijfers uit de PS die zijn gekoppeld aan de VNR-bedrijven.

Internationale Handel (IH)

Internationale Handel bestand met IH gegevens en IH-code <-> BEID in SPSS.

Bijlage III: Praktische illustratie koppelproblematiek

Voorbeeld

Om inzicht te geven in de problematiek rond het koppelen van vergunningsnummers aan statistische eenheden wordt hier een voorbeeld gegeven.

Vergunningsnummer V001 is in gebruik bij een bedrijf met KvK-nummer K001. In het ABR van CBS is daarbij de ondernemingsgroep (OG) gezocht en vervolgens alle statistische eenheden die onder deze OG vallen. We hebben nu een compleet beeld van alle mogelijk actoren:

vergunning	KvK	OG	SBI OG	BEID	SBI BEID	Naam	Omzet mln €
V001	K001	OG1	2010	B001	3000	Bottelaar inpak BV,	400
	K001	OG1	2010	B002	2020	Bottelaar glas BV	200
	K002	OG1	2010	B003	2010	Bottelaar blik BV	600
	K003	OG1	2010	B004	9000	Bottelaar Admin BV	50
	K004	OG1	2010	B005	8000	Bottelaar Handelsmij NV	150

Het bedrijf haalt zijn grootste omzet uit het verwerken van blik wat de SBI-indeling bepaalt van de OG (2010). In het voorbeeld hierboven zien we dat als we naar de geautomatiseerde koppeling op 2 digit SBI kijken, de BEID's B002 en B003 zouden worden meegenomen voor verdere analyse. Nader onderzoek op internet via Google leert ons dat Bottelaar glas BV zelf glazen flessen blaast. Bottelaar Blik BV koopt blik in en stanst daar frisdrankblikjes uit. Bottelaar inpak BV is een verzendbedrijf dat de vervaardigde flessen en blikjes gevuld met aangeleverde vloeistoffen vult en verzendt. Bottelaar admin BV doet enkel de administratie van het bedrijf. Tenslotte is er nog een eenheid die handelsactiviteiten heeft.

Kijken we naar de gevonden gegevens over dit bedrijf, dan is het aannemelijk dat alleen de glasblazerij van Bottelaar Glas BV uitstoot van emissies zal genereren. We kiezen er daarom voor alleen deze statistische eenheid mee te nemen in een vervolganalyse.

Dit eenvoudige voorbeeld laat goed zien met welke koppelingsproblematiek bij vervolganalyse rekening moet worden gehouden.

Bijlage IV: Mogelijke onderzoeksvragen vervolgonderzoek

Voor vervolgonderzoek met gekoppelde data op microniveau bestaan diverse mogelijkheden. We benoemen hier een aantal mogelijke onderzoeksvragen. Hierbij is onderscheid gemaakt naar onderzoek dat met de bestaande koppeling kan worden uitgevoerd en onderzoek waarvoor aanvullend werk moet worden verricht.

Met de data van dit datakoppelproject:

- a. Wat is het aandeel van de ETS-bedrijven in Nederland in de Nederlandse economie gemeten in productiewaarde of toegevoegde waarde?
- b. Wat is de CO₂-emissieintensiteit van de ETS bedrijven micro en naar bedrijfstak (in kg / €) en welke bedrijfstakken of branches (volgens standaard bedrijfsindeling SBI) zijn meest emissie-intensief?
- c. Reduceren de als 'exposed' to international competition' geïdentificeerde ETS-bedrijven hun emissies even sterk als de andere ETS-bedrijven?
- d. Voor welk deel vallen de Nederlandse ETS-bedrijven, hun (Nederlandse) activiteiten en resulterende CO₂-emissies onder buitenlandse zeggenschap, staat daarmee onder invloed van buitenlandse beslissers ('ultimate control')?
- e. Hoe en in welke mate is het energiegebruik van de ETS bedrijven te analyseren?
- f. Wat is de economische omvang van de ondernemingsgroepen die samenhangen met de ETS-bedrijfseenheden?

Uitbreiding van de reeds gekoppelde data met tijdreeks data voor de ETS bedrijven:

- g. Bij welke ETS-bedrijven, ingedeeld naar typen (SBI, grootteklasse, locatie, etc.) vindt de (interne) emissiereductie met name plaats? Dit kan bij behoefte ook op een voorselectie van de ETS-bedrijven, bij voorbeeld op een selectie van de grootste bedrijven of in een selectie op SBI;
- h. Vindt de (interne) emissiereductie bij ETS-bedrijven met name plaats bij de ETS in de emissie-intensieve bedrijfstakken, aangezien zij de grootste stimulans hebben om te reduceren? Dit bijvoorbeeld toepassen op de ETS-bedrijven in specifieke bedrijfstakken;
- i. Hoe wordt de emissiereductie bij de ETS-bedrijven vooral gerealiseerd, door een veranderende brandstofmix, een beter gebruik van restwarmte, door verminderde emissie-intensieve activiteiten, of door een combinatie?
- j. Onderzoek naar verschillen tussen de aan ETS deelnemende bedrijfssectoren. Presteren bepaalde sectoren (bedrijfstakken) beter voor wat betreft de reductie van de uitstoot? Komt dat doordat ETS-bedrijven bepaalde productie activiteiten verplaatsen naar het buitenland? Of zijn er ook andere oorzaken aanwijsbaar met de cijfers?
- k. Wat is de invloed van buitenlandse beslissers ('ultimate control') op de zeggenschapsverdeling van de Nederlandse ETS-bedrijven, hun (Nederlandse) activiteiten en resulterende CO₂-emissies? En vindt daarin een verschuiving plaats?

Gekoppelde ETS-bedrijven en tijdreeks data vergelijken met bedrijven in controlegroep:

- l. Presteren ETS-bedrijven economische vergelijkbaar, beter of slechter dan vergelijkbare niet ETS deelnemende bedrijven (getypeerd naar SBI, productiewaarde, toegevoegde waarde, grootteklasse, locatie)?
- m. Presteren ETS-bedrijven milieutechnisch beter dan niet ETS-bedrijven? En hoe verhoudt zich de interne CO₂ emissiereductie met de vergeleken niet ETS-bedrijven, is die groter?

- n. Innoveren ETS-bedrijven sterk in technologieën specifiek gericht op het reduceren van CO₂-emissies? Of door toepassing van energiebesparende technologie, een combinatie of nog anders?
- o. Hoe kan de ontwikkeling in CO₂-emissie en CO₂-emissieintensiteit over de tijd, worden verklaard? Is daarbij onderscheid mogelijk naar de autonome ontwikkeling en de beleid beïnvloede ontwikkeling? Deze onderzoeksvraag eventueel in samenwerking met een beleidsgerichte partij gezien de beleidsgerichte aspecten;

De gekoppelde data van de ETS-bedrijven aangevuld en gecombineerd met extra data:

- p. Wat is de invloed van de Nederlandse ETS-bedrijven in termen van zeggenschap / eigendom op de activiteiten en resulterende CO₂-emissies van buitenlandse bedrijven?

Bijlage V: Project bijdragen

Project begeleiding:

Kroon, S.H. (Bas) – NEa

Mulder-Bakker, M.H.C. (Margo) - NEa

Bout, M.A. (Rien), MinlenM – DGRW (IenM programma 'Anders omgaan met data')

Holtrop, D.E. (Dennis), MinlenM – DGMI

Project bijdragen:

Remco Kaashoek

Esther de Heij-Vieveen