



Centraal Planbureau

CPB-analyse voorstellen Nationaal Groeifonds

Derde beoordelingsronde, 2023

Het CPB heeft 35 voorstellen geanalyseerd op de domeinen innovatie en kennisontwikkeling.

De CPB-analyses vormen input voor het advies van de adviescommissie.

Deze deelpublicatie geeft de analyse weer van het voorstel:

Data-economie

Subcommissie

Veiligheid en digitalisering

Q4.1 Data-economie

Belangrijkste aandachtspunten

- De indieners noemen verschillende Europese initiatieven op het vlak van datadeling, zoals de Data Governance Act en de Data Act. Het wordt uit het voorstel niet duidelijk in hoeverre Europese ontwikkelingen de problemen rond datadelen al (zullen) ondervangen.
- De keuze voor vijf + vijf sectoren wordt door de indieners nauwelijks onderbouwd. Ook geven de indieners niet aan welke selectiecriteria voor de casussen zullen worden gebruikt.
- Een groot deel van de private bijdrage zal *in-kind* worden geleverd en is nog niet toegezegd. De indieners beogen de datadeelinfrastuur na afloop van de NGF-financiering te laten bekostigen door gebruikers, maar dit wordt in het voorstel niet concreet uitgewerkt.
- De legitimiteit van overheidsingrijpen is niet te beoordelen. Datadeling kan zowel positieve als negatieve effecten hebben, die per casus zullen verschillen. Bij negatieve effecten kan gedacht worden aan, bijvoorbeeld, aantasting van de privacy, prijsdiscriminatie en selectie.

Beschrijving voorstel

Het voorstel Data-economie beoogt het delen van data tussen organisaties uit verschillende sectoren te vergemakkelijken. De indieners geven aan dat de afgelopen jaren verschillende initiatieven zijn gestart om datadeling binnen sectoren mogelijk te maken, en stellen dat er een sectoroverstijgend initiatief nodig is om eilandvorming te voorkomen. Door het delen van data tussen sectoren makkelijker te maken, kunnen volgens de indieners nieuwe producten en diensten ontstaan die economische en/of maatschappelijke waarde hebben. Een voorbeeld hiervan is iDeal, dat data-uitwisseling tussen onder andere webwinkels en banken faciliteert. Een ander (nog te ontwikkelen) voorbeeld is het doorgeven van data van de slimme meter aan een hypotheekverstrekker om daarmee groene leningen te faciliteren.

De kern van het voorstel is de realisatie van een generieke datadeelinfrastuur, die bestaat uit afspraken, standaarden, en diensten met betrekking tot het delen van data. Het voorstel bestaat uit vijf programmalijnen (PL): (1) *Data Spaces*, (2) Harmonisatie, (3) *Human Capital*, (4) Community & Communicatie, en (5) Strategie & PMO. De programmalijnen (1) *Data Spaces* en (2) Harmonisatie worden hieronder beschreven:

- In **PL *Data Spaces*** werken de indieners in een aantal concrete casussen (*use cases*) uit hoe data tussen organisaties uit verschillende sectoren (*data spaces*) gedeeld kunnen worden. De ervaringen met de casussen zullen gebruikt worden bij de ontwikkeling van een generieke datadeelinfrastuur.
- In **PL Harmonisatie** onderzoeken de indieners welke afspraken nodig zijn om data binnen een *data space* en tussen *data spaces* te delen. Hierbij willen de indieners zoveel mogelijk gebruikmaken van bestaande standaarden en (Europese) initiatieven. De resulterende generieke datadeelinfrastuur zal worden toegepast in de casussen. Daarnaast zal de kennis worden gedeeld met (markt)partijen en worden (markt)partijen geholpen om hun datadeelinitiatieven te testen.

Deze *quickscan* gaat niet of zijdelings in op de programmalijnen (3) *Human Capital*, (4) Community & Communicatie, en (5) Strategie & PMO. PL (3) *Human Capital* en PL (4) Community & Communicatie betreffen het verspreiden van kennis onder professionals en studenten door middel van onderwijs (*Human Capital*) en het uitbreiden van de coalitie door actief informatie te verspreiden (Community & Communicatie). Beide programmalijnen zijn beperkt wat betreft financiële omvang. De programmalijn Strategie & PMO betreft

onder meer programmamanagement; het CPB heeft op dit vlak geen expertise. Het is wel opvallend dat deze programmalijn relatief omvangrijk is.

Het voorstel is ingediend door TNO, EZK en de Data Sharing Coalition (hierna: DSC). De DSC zal in 2023 samen met de werkgroep datadelen van de AI-coalitie en de Gaia-X hub NL opgaan in de Center of Excellence for Data Sharing & Cloud. Het voorstel vraagt een bedrag uit het Nationaal Groeifonds (NGF) van **62,2 mln euro** om te besteden over een periode van vijf jaar, 2024-2028. De totale benodigde financiering bedraagt 115,9 mln euro. In de volgende tabel zijn de kosten per PL uitgesplitst.⁴⁹

Tabel **Overzicht van programmalijnen, investeringsbedrag en NGF-bijdrage**

Programmalijn	Totale investering (mln euro)	NGF-bijdrage (mln euro)	NGF-bijdrage (%)	Onderdeel CPB quickscan
Data Spaces	36,8	12,3	33	✓
Harmonisatie	39,4	22,0	56	✓
Human Capital	14,8	7,4	50	
Community & Communicatie	6,4	2,3	36	
Strategie & PMO	16,3	16,2	99	
Uitvoerings- en evaluatiekosten	1,8	1,8	100	
Totaal	115,5	62,0		

1. Scan probleemstelling

Vraag	Bevindingen
a. Is duidelijk welk(e) probleem/ problemen het voorstel tracht op te lossen?	<p>Het probleem is op hoofdlijnen duidelijk. De indieners noemen verschillende problemen die het delen van data tussen organisaties belemmeren:</p> <ul style="list-style-type: none"> • Het bereiken van afstemming over het delen van data is complex en vaak (te) duur voor individuele (markt)partijen. Ook ontbreekt het individuele (markt)partijen aan kennis hierover. • Het delen van data binnen en tussen sectoren kan op verschillende manieren, maar (markt)partijen zijn niet in staat te coördineren op een standaardmanier voor het delen van data. • (Markt)partijen willen controle hebben over wat met de data gebeurt na het delen daarvan met anderen. Zij zijn bang dat het delen van data hun concurrentiepositie kan schaden of <i>compliance</i>-risico's met zich meebrengt, zoals een schending van de AVG. <ul style="list-style-type: none"> ○ De indieners noemen kleine marktpartijen specifiek als partijen die vanwege hun onderhandelingspositie alleen data kunnen delen met grote (markt)partijen (<i>BigTechs</i>) onder ongunstige voorwaarden. <p>De indieners noemen verder dat er zowel op Nederlands als Europees</p>

⁴⁹ Door afrondingsverschillen wijken de totalen in de tabel licht af van de genoemde totalen in de tekst.

	<p>niveau diverse datadeelinitiatieven zijn of worden ontwikkeld. Het verschil tussen deze initiatieven en dit voorstel is echter dat dit voorstel zich richt op data delen <i>tussen</i> sectoren en dat de andere initiatieven veelal gericht zijn op data delen <i>binnen</i> sectoren. De indieners willen met dit voorstel voorkomen dat er verschillende sectorale oplossingen voor het delen van data ontstaan die het delen van data tussen sectoren praktisch onmogelijk maken.</p> <p>Bij deze probleemstelling maken we de volgende kanttekeningen:</p> <ul style="list-style-type: none"> • Het wordt uit het voorstel niet duidelijk in welke mate er op Europees niveau al ontwikkelingen zijn richting harmonisatie van standaarden voor het intersectoraal delen van data. De indieners noemen de Data Governance Act, die databemiddelingsdiensten reguleert, en de aankomende Data Act, die interoperabiliteit moet verhogen en de waarde van data eerlijk moet verdelen. Deze wetgeving zou een deel van de genoemde problemen al kunnen ondervangen. • De indieners noemen het risico van ongewenste machtsconcentratie door schaal- en netwerkeffecten als een probleem. De indieners lijken hier te doelen op een situatie waar een beperkt aantal partijen (<i>BigTechs</i>) data verzamelt en beheert. Het wordt in het voorstel echter niet onderbouwd wat de risico's zijn en hoe realistisch dit scenario is, mede met het oog op huidige en toekomstige Europese wetgeving en het toezicht door de Europese en nationale mededingingsautoriteiten.
<p>b. Is aannemelijk dat de initiatieven de problemen verhelpen (gegeven de bij ons beschikbare kennis over het ecosysteem)?</p>	<p>Overall beeld</p> <p>Het voorstel zou bij kunnen dragen aan het realiseren van een generieke datadeelinfrastuctuur die het intersectoraal delen van data vergemakkelijkt. PL1 en PL2 bieden gelegenheid om ervaring op te doen met intersectoraal datadelen en dit vast te leggen in een generieke datadeelinfrastuctuur. De andere programmalijnen dragen bij aan de verdere verspreiding van deze kennis.</p> <p>De planuitwerking roept echter wel een aantal vragen op. De indieners willen starten met intersectorale casussen in vijf sectoren en willen na een jaar hier nog vijf sectoren aan toevoegen. De initiële vijf sectoren zijn gekozen, omdat ze volgens de indieners voorlopen in intrasectoraal datadelen en passen in het topsectorenbeleid.</p> <ul style="list-style-type: none"> • Het is onduidelijk waarom voor dit aantal sectoren is gekozen en of het wenselijk is op korte termijn al tien sectoren te betrekken. Een beperkter aantal zou meer focus kunnen aanbrengen. • De indieners lichten daarnaast niet toe in hoeverre de gekozen sectoren de sectoren zijn waar het delen van data (de meeste) waarde toe kan voegen of het meest bij kan dragen aan het ontwikkelen van een generiek afsprakenstelsel. <ul style="list-style-type: none"> ○ Een van de initiële sectoren is gezondheid. Deze is vanwege de AVG (gezondheidsdata zijn bijzondere persoonsgegevens) niet representatief en mogelijk lastiger om in te passen in een generiek afsprakenstelsel.

	<ul style="list-style-type: none"> Het is ook niet duidelijk op basis van welke criteria de casussen gekozen zullen worden. Deze keuze is van belang, omdat de casussen als input dienen voor de ontwikkeling van een generiek afsprakenstelsel. <p>Bij het voorstel maken we verder de volgende kanttekeningen:</p> <ul style="list-style-type: none"> De indieners beogen na het einde van de NGF-financiering de datadeelinfrastuctuur te vermarkten op een soortgelijke manier als iDEAL, door het vragen van periodieke <i>fees</i> en/of <i>fees</i> per datadeeltransactie. Dit wordt in het voorstel echter niet verder uitgewerkt. Risico's zijn onder andere dat de datadeelinfrastuctuur nog niet van voldoende kwaliteit is na afloop van de NGF-financiering, of dat niet voldoende partijen bereid zijn om te betalen voor de diensten. Een aanzienlijk deel van de private bijdrage zal <i>in-kind</i>, in de vorm van uren, worden geleverd door deelnemers. Een groot deel van de deelnemers heeft deze uren nog niet toegezegd. Daarnaast zijn de kwantiteit en de kwaliteit van de toegezegde uren lastig te controleren. Naast de DSC zijn er verschillende andere Nederlandse initiatieven op het vlak van intersectoraal datadelen. De indieners geven aan dat de DSC in 2023 samen met de werkgroep Datadelen van de AI-coalitie en de Gaia-X hub NL op zal gaan in het Center of Excellence for Data Sharing & Cloud. Het is niet duidelijk of er naast deze partijen nog andere intersectorale datadeelinitiatieven in Nederland zijn. Ook is niet duidelijk of de NGF-subsidie alleen ten goede komt aan de DSC, of ook aan het Center of Excellence.
<p>c. Zijn er, gegeven onze beschikbare kennis, andere projecten die het probleem verhelpen?</p>	<p>Zoals bij de probleemstelling genoemd, zijn er verschillende initiatieven op Europees niveau. De indieners geven aan hierbij te willen aansluiten en actief te willen bijdragen aan internationale standaarden en ontwerpprincipes. Maar, zoals de indieners ook zelf aangeven, er is nog veel in ontwikkeling, zoals wetgeving op EU-niveau. Het risico bestaat daarom dat het onderzoek naar een Nederlandse generieke datadeelinfrastuctuur ingehaald wordt door deze ontwikkelingen. Een alternatief zou kunnen zijn deze ontwikkelingen, met name wetgeving, af te wachten. Het CPB kan niet beoordelen wat de voor- of nadelen zijn van een meer afwachtende houding.</p>

2. Scan legitimiteit

Vraag	Bevindingen
<p>a. Zijn er economische redenen die overheidsingrijpen legitimeren?</p>	<p>Overall beeld Er zijn verschillende vormen van marktfalen die overheidsingrijpen kunnen legitimeren:</p>

- **Transactiekosten vanwege de complexiteit van het delen van data en het ontbreken van kennis.** Het bereiken van afstemming over het delen van data, met name tussen organisaties uit verschillende sectoren, is complex en duur. Ook ontbreekt het (markt)partijen aan kennis hierover. Dit kan worden gezien als (hoge) transactiekosten die het delen van data belemmeren.
- **Semi-collectief goed in de vorm van een generieke datadeelinfrastructuur.** Een datadeelinfrastructuur bestaat uit afspraken, standaarden, en diensten met betrekking tot het delen van data. Dit kan worden gezien als een (semi-)collectief goed, omdat het gebruik van een datadeelinfrastructuur door een partij het gebruik door een ander niet in de weg hoeft te staan. Wel is het mogelijk anderen (deels) uit te sluiten van het gebruik van deze generieke datadeelinfrastructuur door daarvoor een prijs te vragen.
- **Externe effecten in de vorm van kennis-*spillovers*.** Een deel van het voorstel betreft het uitwerken van concrete casussen voor het delen van data tussen organisaties uit verschillende sectoren. Dit kan worden gezien als een vorm van (fundamenteel of toegepast) onderzoek, waarvan doorgaans verondersteld wordt dat dit positieve externe effecten heeft. Anderen kunnen hierop immers voortbouwen zonder dat zij de kosten hiervoor dragen.

Naast de genoemde redenen voor overheidsingrijpen zou datadeling gepaard kunnen gaan met maatschappelijke effecten. Deze kunnen echter zowel positief als negatief zijn, waardoor het oplossen van de bovengenoemde marktfalen ook juist tot maatschappelijke kosten kan leiden (zie ook Lipsey en Lancaster, 1956). Wij geven een aantal voorbeelden van mogelijke maatschappelijke effecten; dit is geen uitputtende lijst.

- Het delen van data kan positieve effecten hebben als dit bijvoorbeeld leidt tot het leveren van producten die beter aansluiten op de behoeften van de consument, meer inzicht in milieuproblemen biedt of de veiligheid bevordert.
- Het delen van data kan echter ook betekenen dat producenten beter weten welke consumenten bereid zijn meer te betalen; zij kunnen deze informatie gebruiken om aan die consumenten een hogere prijs te vragen (prijsdiscriminatie). In economische zin kan dit een overdracht betekenen van consumenten naar producenten.
- Het delen van data kan ook betekenen dat consumenten die veel kosten uitgesloten worden door producenten. Dit speelt onder meer in de zorg en bij verzekeringen.
- Het delen van data door een consument kan ook informatie onthullen over andere consumenten die deze informatie helemaal niet willen delen (privacy). Dit zijn negatieve externe effecten die individuele consumenten niet meewegen bij de beslissing om data te delen, met als gevolg dat te veel data gedeeld wordt (zie bijvoorbeeld El-Dardiry et al., 2021 of Acemoglu et al., 2022).

PL1

De legitimiteit van overheidsingrijpen bij deze programmalijn is in zijn algemeenheid niet te beoordelen.

- De activiteiten in deze programmalijn zijn gericht op het uitwerken van concrete casussen voor datadelen. Omdat bij het uitwerken rekening wordt gehouden met de mogelijkheid dezelfde oplossingen te (her)gebruiken voor andere casussen, kan sprake zijn van kennis-*spillovers* die enige mate van overheidsingrijpen zouden kunnen legitimeren.
- Zoals hiervoor beschreven, kunnen er zowel positieve als negatieve maatschappelijke effecten optreden. Dit zal per casus verschillen en is in het algemeen niet te beoordelen. Er zal per casus een weging moeten plaatsvinden van de wenselijkheid van datadeling.
- Het uitwerken van concrete casussen kan tot private baten leiden. Een private cofinanciering is daarom passend. PL 1 vraagt 33% subsidie uit het NGF. Het CPB kan niet inschatten of dit inderdaad een passende weergave is van de private baten.

PL2

De legitimiteit van overheidsingrijpen bij deze programmalijn is in zijn algemeenheid niet te beoordelen.

- De activiteiten in deze programmalijn zijn gericht op het ontwikkelen van een semi-collectief goed in de vorm van een generieke datadeelinfrastructuur. Dit kan enige mate van overheidsingrijpen legitimeren.
- Een generieke datadeelinfrastructuur kan negatieve maatschappelijke effecten hebben, bijvoorbeeld als deze drempels voor datadelen wegneemt en leidt tot excessief datadelen en aantasting van de privacy (zoals hiervoor beschreven).
- De ontwikkeling van een generieke datadeelinfrastructuur zou positieve maatschappelijke effecten mogelijk kunnen maken. Ook zouden negatieve effecten deels kunnen worden voorkomen als deze infrastructuur de risico's van datadeling (onder andere op het vlak van privacy) goed ondervangt en daarmee een beter alternatief vormt dan bijvoorbeeld initiatieven van individuele bedrijven.
- De private baten vormen een aandachtspunt. Het is mogelijk een prijs te vragen voor de datadeelinfrastructuur, zoals de indieners ook voorstellen. Wanneer deze toegangsprijs alleen de onderhoudskosten van de infrastructuur dekt, lijken de private baten zeer beperkt en kan overheidsingrijpen nodig zijn om de infrastructuur tot stand te brengen. De mogelijkheid om een prijs te vragen voor gebruik zou echter ook kunnen betekenen dat marktpartijen een prikkel hebben om standaardoplossingen voor het delen van data te ontwikkelen. De indieners lijken hier op te doelen wanneer zij het risico van *BigTechs* benoemen. Het CPB kan niet beoordelen in hoeverre private oplossingen mogelijk of wenselijk zijn.

Literatuur

Acemoglu, D., A. Makhdoumi, A. Malekian en A. Ozdaglar, 2022, Too much data: Prices and inefficiencies in data markets, *American Economic Journal: Microeconomics*, vol. 14, nr. 4, pag. 218-256.

El-Dardiry, R., B. Overvest, M. Dinkova en R. Albers, 2021, *Brave New Data. Databeleid In Een Imperfecte Wereld*, CPB Policy Brief, Den Haag: Centraal Planbureau.

Lipsey, R.G. en K. Lancaster(1956), The general theory of second best, *The Review of Economic Studies*, vol. 24, nr. 1,, pag. 11-32.