

Handelingsperspectief ten aanzien van de nog niet aan de Staat overgedragen mijnbouwgegevens

TNO 2024 R11976 – 22 oktober 2024

Handelingsperspectief ten aanzien van de nog niet aan de Staat overgedragen mijnbouwgegevens

Auteurs	I.C. Kroon, R.A.F. Dalman, V. van der Kraan, J. Dijk en N.H. Zornig
Rubricering rapport	TNO Publiek
Titel	TNO Publiek
Rapporttekst	TNO Publiek
Bijlagen	TNO Publiek
Aantal pagina's	63 (excl. voor- en achterblad)
Aantal bijlagen	6
Opdrachtgever	Ministerie van Economische Zaken en Klimaat
Programmanaam	EZK Werkprogramma 2022
Projectnaam	Borging MB data vertrekkende operators
Projectnummer	060.51941/01.03.05

Alle rechten voorbehouden

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van TNO.

© 2024 TNO

Samenvatting

De Mijnbouwwet verplicht olie- en gasexploitanten om hun gegevens, en daarvan afgeleide interpretaties, te beheren en ze periodiek of op aanvraag aan de Nederlandse Staat te verstrekken. Na (eenmalige) overdracht vervalt de beheerverplichting. Gezien de huidige verwachting, dat veel exploitanten op afzienbare termijn hun commerciële activiteiten in Nederland zullen verkleinen of stoppen, is het met het oog op hergebruik in diverse maatschappelijke vraagstukken, cruciaal om de nog bij de exploitanten in beheer zijnde gegevens tijdig over te dragen en te borgen door de Staat.

In opdracht van het ministerie van Klimaat en Groene Groei (KGG) heeft de Geologische Dienst Nederland (GDN) een verkennend onderzoek uitgevoerd naar het hergebruikspotentieel (baten) en naar de benodigde inspanning (kosten) van het borgen van deze gegevens. Dat hebben wij gedaan door middel van een enquête onder de GDN collega's en interviews met zusterinstituten in andere landen, waarbij we onder meer hebben gevraagd naar alternatieven, zoals het later opnieuw inwinnen van (een deel van de) gegevens. Aansluitend hebben we het handelingsperspectief verder ingekaderd door de kostenschattingen uit te werken voor zes perspectieven, die onderling verschillen in de hoeveelheid en datatype(n) van de te borgen gegevens en interpretaties.

Onze eerste bevinding is, dat de exploitanten momenteel 30-60% (afhankelijk van het datatype) van de unieke historische en actuele gegevens en interpretaties beheren. Ten tweede is het hergebruikspotentieel aanzienlijk, waarbij de voorziene baten per datatype sterk variëren tussen 17 maatschappelijke subthema's (en de respondentengroep aan de fysische data het hoogste hergebruikspotentieel toeschrijft).

Voor wat betreft de alternatieven voor borging van de bestaande gegevens geldt: alle datatypen zijn moeilijk reproduceerbaar en opnieuw inwinnen is kostbaar. Bovendien is verlies van kennis en vaardigheden voor hergebruik tot 2050 reëel. Ook geldt voor alle datatypen, dat borging van deze extra data meer inspanning vraagt dan de huidige inspanning bij GDN voor het borgen van de huidige datasets. We schatten de kosten voor minimale borging van alle gegevens en interpretaties (exclusief opslag van kernmateriaal in één of meer kernhuizen en met uitgifte voor hergebruik uitsluitend op projectbasis) op 40 miljoen euro, waarvan circa 22 miljoen euro voor de seismische velddata op tapes. De kosten voor het volledig integreren van alle gegevens en interpretaties, inclusief seismische velddata op tapes, in de GDN-databank - wat de kans op hergebruik vergroot - bedragen naar schatting 126 miljoen euro. De onzekerheid in de genoemde bedragen bedraagt naar schatting 40% op grond van vooral de onzekerheid in het aantal datasets per datatype.

Op grond van bovenstaande bevelen we aan om gegevensverlies te voorkomen. Dat kan door:

- een databorgende partij het recht te geven om alle gegevens en interpretaties bij vergunningsteruggave op te vragen en proportionele sancties in te stellen bij niet-naleving;
- exploitanten te verplichten een inventarislijst van gegevens en interpretaties te maken en deze periodiek te actualiseren;
- afspraken te maken met de sector over gefaseerde overdracht van fysieke gegevens en interpretaties;
- in de komende jaren voldoende middelen vrij te maken voor de (minimale) borging van gegevens via overdracht, inname en beheer en op zoek te gaan naar co-sponsoring;

- op korte termijn te beslissen over (de wijze van) uitbreiding van opslagcapaciteit voor fysieke data;
- maatschappelijke bewaarcriteria op te stellen voor data (variërend van eeuwig bewaren tot natuurlijke bewaartermijnen).

Om hergebruik bij minimale borging te stimuleren, bevelen we aan om hergebruik, opwerking en uitgifte van data op projectbasis aantrekkelijk te maken. Dat kan onder meer door:

- bij overdracht voorrang te geven aan de top 15 prioritaire gegevens (waarbij de overheid een eigen top 15 kan samenstellen op basis van specifieke maatschappelijke subthema's of gebruikerswensen);
- innovatieve data-uitgifte manieren te ontwikkelen voor opkomende gebruikersgroepen, zoals AI/ML-experts;
- regelmatig in gesprek te gaan met bestaande gebruikersgroepen over hun wensen en drempels.

Mocht in een later stadium blijken, dat hergebruik via ontsluiting op projectbasis onvoldoende van de grond komt, dan kunnen de gegevens en interpretaties alsnog via de GDN-databank worden ontsloten. Bij vrijgave dient de overheid rekening te houden met de zorgen bij exploitanten over de juridische aansprakelijkheid bij hergebruik door derden (dit betreft met name de interpretaties) en over het beantwoorden van (aan hergebruik gerelateerde) vragen over de overgedragen gegevens en interpretaties. Dit laatste punt kan waarschijnlijk deels worden ondervangen door de communicatie via een aangewezen partij te laten verlopen.

Tenslotte bevelen we aan om op korte termijn te starten met gerichte data-inname (via pilots) om de kosten beter te kunnen inschatten en om ervaring op te doen met inname, beheer en vrijgave van data (zowel opschaling als nieuwe datatypen als innovaties), in afstemming met exploitanten en betrokken overheden en uitvoeringsinstanties.

Inhoudsopgave

Samenvatting	3
1 Inleiding	6
1.1 Aanleiding.....	6
1.2 Achtergrond	6
1.3 Probleem- en doelstelling	7
1.4 Afbakening	8
1.5 Aanpak en leeswijzer	8
2 Hergebruik en borgingsinspanning	10
2.1 Maatschappelijke thema's en datatypen.....	10
2.2 Enquêtemethodiek	10
2.3 Representativiteit.....	12
2.4 Duiding van de enquêteresultaten.....	15
3 Handelingsperspectief	20
3.1 GDN ervaring met databeheer.....	20
3.2 Voorbeelden van eerdere keuzes.....	23
3.3 Fysieke en digitale data.....	25
3.4 Kostenindicaties per scenario	27
3.5 Handelingsperspectief	29
4 Analoge werkwijze in vergelijkbare landen	34
4.1 Gespreksresultaten.....	34
4.2 Inzichten.....	36
5 Bevindingen en aanbevelingen	37
5.1 Inzichten.....	37
5.2 Voorkom dat mijnbouwgegevens verloren gaan.....	39
5.3 Stimuleer hergebruik.....	40
5.4 Pilots	40
Referenties	43
Ondertekening.....	47
Bijlagen	
Bijlage A: Maatschappelijke thema's	48
Bijlage B: De enquêtevragen	49
Bijlage C: De enquête-antwoorden	50
Bijlage D: Kostenberekening	53
Bijlage E: Perspectieven	58
Bijlage F: Interviews	62

1 Inleiding

1.1 Aanleiding

De wet (Mijnbouwwet, -besluit & -regeling) regelt de verplichting tot gegevensverstrekking door in Nederland werkzame olie- en gasexploitanten aan de Staat. Dit betreft (juridisch gezien) alle mijnbouwgegevens, die bij hen aanwezig zijn. Uitgangspunt is, dat de exploitanten deze gegevens gedurende de operationele vergunningsfase zelf beheren en ze op aanvraag overdragen aan de Staat. De belangrijkste uitzondering hierop betreft gegevens, die van direct belang zijn voor beleid en toezicht en daarom door de exploitanten periodiek worden aangeleverd aan de Staat. Een andere uitzondering betreft de aanlevering van gegevens en informatie in het kader van vergunningverlening. Na de (eenmalige) overdracht berust de verantwoordelijkheid voor opslag en beheer van deze gegevens bij de Staat als bronhouder.

Dit heeft ertoe geleid, dat de exploitanten momenteel een aanzienlijk deel van de historische en actuele mijnbouwgegevens en de daarvan afgeleide interpretaties beheren. De vraag is, in hoeverre en op welke termijn deze gegevens en interpretaties dienen te worden overgedragen aan en geborgd door de Staat. In de kern is dit een politiek-bestuurlijke afweging, die nu aan de orde is, omdat:

- naar verwachting veel van de huidige exploitanten op relatief korte termijn (tussen nu en 2035) zullen gaan stoppen met hun commerciële activiteiten in Nederland;
- er - afhankelijk van de te maken beleidskeuzes - een aantal jaren voorbereidingstijd nodig zijn om de overdracht, borging en vrijgave te kunnen realiseren en
- de samenleving met het oog op hergebruik baat heeft bij (versnelde) vrijgave van deze mijnbouwgegevens en de daarvan afgeleide interpretaties.

Op verzoek van het ministerie van Klimaat en Groene Groei (KGG) en in het kader van onze wettelijke taak als gedelegeerd databeheerder namens de Staat hebben we verkennend onderzoek uitgevoerd naar dit vraagstuk. Rode draad daarbij is de vraag of de benodigde inspanning (en de daarmee gepaarde kosten) opweegt tegen het hergebruikspotentieel in diverse maatschappelijke vraagstukken (en de daarmee gepaarde baten).

1.2 Achtergrond

Bovenstaand vraagstuk speelt tegen de achtergrond van het klimaatbeleid, de energietransitie, en andere crises. Deze maatschappelijke opgaven worden steeds complexer, zodat belanghebbende partijen meer gegevens en interpretaties nodig hebben dan voorheen. Hierbij kennen we nog niet alle vragen, die in de toekomst gesteld gaan worden. Voorbeelden van een veranderende vraagstelling zijn:

- Beperken van de huidige gas schaarste;
- Ontmanteling/hergebruik van de gasinfrastructuur;
- Opschaling van het winnen van aardwarmte;
- Toenemend gebruik van de matig diepe ondergrond (300-1500 m);
- Innovaties voor de opslag van energie houdende materialen;
- Interacties tussen ondergrond en bovengrond;
- Informatievoorziening van regionale overheden.

Om dergelijke vragen te kunnen beantwoorden heeft de samenleving (en de Geologische Dienst Nederland (GDN) als onderdeel van TNO in dienst daarvan) toegang nodig tot de nog niet aan de Staat overgedragen mijnbouwgegevens. De gegevens afkomstig van de gas- en oliewinning zijn momenteel in beheer bij de diverse exploitanten. Deze door de exploitanten verzamelde data en hun interpretaties zijn uniek en vaak kostbaar om opnieuw te verkrijgen (als dat al kan). In het algemeen geldt daarbij, dat hoe groter de diepte hoe minder informatie over de diepe ondergrond beschikbaar is.

Wij verwachten, dat het vertrek van exploitanten (en daarmee het natuurlijke verloop van de overdracht) gedreven door de prijsontwikkeling geclusterd zal gaan gebeuren. Vanuit de winningsplannen en ons jaarlijkse overleg met de in Nederland werkzame olie- en gasexploitanten¹ hebben we een grove inschatting van wie wanneer van plan is zijn vergunning(en) terug te geven aan de Staat². Op grond hiervan weten we, dat het vraagstuk van overdracht snel opgepakt moet worden: als de overdracht niet tijdig geregeld wordt, neemt de kans toe, dat data onbedoeld verloren gaat, terwijl het vergaren van gegevens over en effecten van het werken in de (diepe) ondergrond kostbaar zal blijven. Ter vergelijking, in 2021 is het “Totaalrapport kosten en baten Basisregistratie Ondergrond (BRO)” (Hulsker et al., 2021) opgeleverd met als conclusie, dat investeren in het vergaren en behouden van BRO-data maatschappelijk gezien een goede investering is geweest.

Het meest acuut is de situatie van de NAM. Door de definitieve productiestop van de gaswinning uit het Groningen veld neemt het bedrijfsbelang van de daaraan gerelateerde gegevens snel af. Ten tweede blijft de gasproductie uit de kleine velden krimpen (als gevolg van de maturatie van het Noordzeebekken voor gasexploratie en hooguit wat vertraagd door het versnellingsplan gas³). De NAM heeft bovendien recent gekozen voor een slanke bedrijfsvoering, waarbij de kosten voor databeheer tot het minimum zijn gereduceerd. Als gevolg hiervan zijn gesprekken gaande tussen de KGG, NAM en de Geologische Dienst Nederland (GDN) over versnelde overdracht van fysieke en digitale data.

1.3 Probleem- en doelstelling

De sector brede overdracht van gegevens vormt om meerdere redenen een uitdaging:

- De reikwijdte van de overdracht is niet goed bekend. Er bestaat geen sector brede inventaris van gegevens, die onder de Mijnbouwwet vallen en die nog niet via de GDN aan de Staat zijn overgedragen.
- Ook is niet bekend, welke data wij (in dienst van de samenleving) minimaal nodig hebben voor het vervullen van onze toekomstige taak;
- Het moment van overdracht hangt samen met de prijsontwikkeling en is daarom notoir moeilijk te voorspellen;
- Datasets en de interne structuur daarvan verschillen per exploitant en zijn vaak samengesteld van karakter door eerdere fusies van bedrijven en veranderingen in databeleid (Akinyemi et al., 2020, Chapela, 2023).

Daarmee is ook onbekend, welke kosten en organisatorische aspecten daarbij een rol spelen. Zodoende lopen we als Geologische Dienst Nederland op dit moment achter de feiten aan.

¹ Spirit Energy Nederland B.V.; Dana Netherlands B.V.; Neptune Energy B.V.; Nederlandse Aardolie Maatschappij B.V.; ONE-Dyas B.V.; Petrogas E&P Netherlands B.V.; TAQA Energy B.V.; TotalEnergies EP Nederland B.V.; Kistos plc.; Vermilion Energy Netherlands B.V.; Wintershall Noordzee B.V.; Jetex Petroleum Ltd

² Deze informatie is bedrijfsvertrouwelijk en daarom niet opgenomen in deze rapportage. Het is wel apart gedeeld met KGG.

³ <https://www.tweedekamer.nl/downloads/document?id=2022D31381>

Om de zaken nog ingewikkelder te maken geven de exploitanten zich bij hun datastrategie nauwelijks rekenschap van het hergebruikspotentieel van hun data en informatie voor maatschappelijke vraagstukken; ze richten zich primair op de eigen bedrijfsstrategie (Hanton en Anantharamu, 2023, DeMarco, 2023) en een veilige werkomgeving (Parker et al., 2006). Dit sluit aan bij onze ervaring, dat exploitanten hun overdracht nauwelijks voorbereiden. Om kostenbesparingen te realiseren hebben ze de neiging hun data-experts vóór de daadwerkelijke overdracht te ontslaan, wat resulteert in – vermijdbaar – verlies van expertise (amnesie) over de context, metadata, geschiedenis en gegevenskwaliteit van de betreffende datasets. Ook hierom moeten we tijdig anticiperen.

Het borgen, beheren en vrijgeven van al deze gegevens vergt een investering (met een onbekend rendement). Het is dus essentieel om:

- de borgingsinspanning en de kans op het verloren gaan van data te kunnen inschatten;
- en dit te vergelijken met de potentiële maatschappelijke relevantie van de gegevens bij hergebruik en
- met de impact van het (later) opnieuw zelf inwinnen van deze gegevens.

Dit onderzoek beoogt het handelingsperspectief inzichtelijk te maken voor de opdrachtgever.

1.4 Afbakening

Dit onderzoek is een eerste verkenning, waarbij we ons beperken tot het Nederlands grondgebied conform de Mijnbouwwet, waarbij we ons richten op de in Nederland werkzame olie- en gasexploitanten met winningsvergunningen voor olie en gas (actief, slapend of teruggeven aan de Staat). Ook beperken we ons tot de vraag, welke informatie op hoofdlijnen voor de Nederlandse samenleving zinvol kan zijn om te borgen voor hergebruik, inclusief een eerste schatting van de bijbehorende kosten. Het toetsen van onze bevindingen bij andere belanghebbenden en de eventuele uitvoering laten we buiten beschouwing.

1.5 Aanpak en leeswijzer

Een maatschappelijke kostenbatenanalyse (MKBA) is zinvol voor dit vraagstuk, maar om verschillende redenen is dat nu niet aan de orde:

- Het uitvoeren van een volledige MKBA is tijdrovend, terwijl het probleem acuut is;
- We weten te weinig van de toekomstige, maatschappelijke baten (TNO, 2023) om deze te kunnen kwantificeren.

Daarom kiezen we voor een kwalitatieve aanpak met nadruk op:

- de hergebruikswaarde van de hoofdtypen in huidige en toekomstige vragen;
- de moeite van het opnieuw inwinnen van de hoofdtypen (kosten);
- impact (organisatorisch en qua kosten) van het borgen, beheer en beschikbaar stellen van de hoofdtypen (met alle datasets).

Om deze inzichten te verzamelen en te verwerken tot een handelingsperspectief volgen we de volgende aanpak:

- Een enquête om de GDN experts te bevragen over het hergebruikspotentieel op diverse maatschappelijke thema's en de benodigde borgingsinspanning van de hoofdtypen met hun eventuele vervangbaarheid door hernieuwde data-inwinning (hoofdstuk 2);
- Documenteren van relevante GDN ervaringen ten aanzien van kosten van inwinning, beheer en beschikbaar stellen van diverse mijnbouwgegevens, op basis waarvan we (in

combinatie met de bevindingen uit de enquête) vervolgens voor vier scenario's een kostenindicatie maken, die de basis vormen voor het uitwerken van het handelingsperspectief (hoofdstuk 3);

- Interviews met collega's uit andere landen, met een deels vergelijkbare situatie, om te leren van hun aanpak en inspiratie op te doen (hoofdstuk 4).

Tot slot vatten we in hoofdstuk 5 onze opgedane inzichten over het handelingsperspectief samen en komen we tot een aantal aanbevelingen.

2 Hergebruik en borgingsinspanning

In dit hoofdstuk maken wij een inschatting van het hergebruikpotentieel door publieke of private partijen van dat deel van de mijnbouwgegevens en de interpretaties daarvan, welke in beheer zijn bij de nu in Nederland werkzame olie- en gasexploitanten, aangevuld met een relatieve inschatting van de benodigde borgingsinspanning van de hoofddatatypes plus hun vervangbaarheid door hernieuwde data-inwinning. Deze inschatting betreft het oordeel van de bij GDN aanwezige deskundigen en is getoetst middels een enquête.

We beginnen dit hoofdstuk met een korte toelichting op de maatschappelijke thema's en hoofddatatypes (paragraaf 2.1), die aan de basis liggen van deze eerste verkenning (bij gebrek aan een sectorbrede inventaris van gegevens, die onder de Mijnbouwwet vallen en die nog niet via de GDN aan de Staat zijn overgedragen, zie paragraaf 1.3). Vervolgens bespreken we de enquêtemethodiek (paragraaf 2.2) en de representativiteit van de bevroegde expertgroep (paragraaf 2.3). Tot slot duiden we de enquêteresultaten (paragraaf 2.4).

2.1 Maatschappelijke thema's en datatypes

Onze insteek is het verkennen van de hergebruikpotentie door publieke of private partijen. Daarbij maken we gebruik van zeventien maatschappelijke (en voor de GDN relevante) subthema's. Voorbeelden hiervan zijn geothermie, energieopslag, grondwater en seismiteit. Voor de volledige lijst zie Bijlage A.

Om een eerste inschatting of ruwe prognose te kunnen maken van de borgingsinspanning onderscheiden we zeven hoofddatatypes (Tabel 2.1). Naast digitale gegevens omvatten deze datatypes ook andersoortige gegevens, zoals monsters of papieren data. Naast gegevens over de ondergrond betreft dit gegevens over de mijnbouwactiviteiten, -effecten en schades.

2.2 Enquêtemethodiek

Om tot een prognose van het hergebruikspotentieel en de borgingsinspanning te komen hebben we gekozen voor een enquête, wat (in de afwezigheid van data-gedreven modellen) een efficiënte methode is. Waar de prognose voor de borgingsinspanning (bij goed boekhouden) uiterlijk in 2050 getoetst kan worden voor de dan geborgde gegevens en informatie, geldt dit niet/minder voor de hergebruikpotentie bij diverse maatschappelijke vraagstukken. Immers, de diversiteit van de vraagstukken en het feit, dat de GDN beperkt zicht heeft op wie welke gegevens uit de nationale databank (verder 'GDN databank') waarvoor gebruikt, hinderen inventarisatie achteraf. En daarbij zal zelfs onder optimale omstandigheden niet alle potentie worden gerealiseerd. Dit gegeven vraagt van de respondenten een intrinsieke motivatie om zich de inspanning te getroosten om hun meest accurate inschattingen met ons te delen. Doordat de enquête geen wedstrijdelementen

(zoals beloningen of competitie tussen groepen) bevat, zijn er noch prikkels voor respondenten om hun inspanning voor waarheidsvinding te vergroten, noch om af te wijken van hun eigen waarheidsbevinding (de zogenaamde gaming-the-system response, Karger et al., 2021).

Als respondenten fouten maken of als hun antwoorden door ons verkeerd zijn geduid, is het lastig/ondoenlijk om de foutenbron te achterhalen. Daarom hebben we de enquête zo opgezet, dat de antwoorden kunnen worden gesplitst in het meest gegeven antwoord met een representatieve bandbreedte ($N \geq 45$ respondenten) en de mening van de experts, die we op basis van zelfrapportage hebben gekoppeld aan specifieke datatypes. Hiermee verkrijgen we een beeld in hoeverre de 'groepswijsheid' verschilt van de 'onderbouwde gissingen' van de vakspecialisten, waarbij significante afwijkingen kunnen wijzen op verminderde betrouwbaarheid. In dat geval is aanvullende, empirische analyse (bijvoorbeeld in de vorm van pilots) noodzakelijk.

Op voorhand weten we overigens niet welke van de twee het meest betrouwbare resultaat geeft; bij studies naar (korte termijn) prognoses blijken groepen van betrokken amateurs zowel groepen van experts als (statistische) modelberekeningen te kunnen overtreffen (Metaculus, 2020; Farrow et al., 2017, Tetlock, 2005). Wel is duidelijk, dat de inschattingen van expertgroepen voor dit soort prognoses van waarde kunnen zijn – en vrijwel altijd beter dan die van individuele experts (McAndrew en Reich, 2020).

Tabel 2.1: Hoofddatatypes.

Datatype	Beschrijving
Biologische data	Monsters en interpretaties van ondergrondse materialen, die via (micro-) fossielen informatie bevatten over statiegrafie, afzettingsmilieu en/of paleoklimaat
Fysische data	Seismische- en putgegevens en de afgeleiden hiervan, die betrekking hebben op de statische en dynamische eigenschappen van de ondergrond
Chemische data	Gegevens van ondergrondse materialen (doorgaans afkomstig van kernen of monsters), die betrekking hebben op de samenstelling van en chemische reacties binnen ondergrondmaterialen
Operationele data	Ondergrond- en oppervlaktegegevens met betrekking tot mijnbouwactiviteiten, zoals de bouw, productie, onderhoud en ontmanteling van mijnbouwinfrastructuur
Mijnbouweffectendata	Meetplan- en andere gegevens om het mogelijk optreden van mijnbouweffecten mee vast te kunnen stellen
Schadedata	Gegevens over schade aan gebouwen of infrastructuur in (mogelijk) verband met mijnbouw
Geïnterpreteerde data	Analyse- en onderzoeksresultaten uit studies over de levenscyclus van mijnbouw (van exploratie tot ontmanteling)

De enquête bestaat uit vier delen (Bijlage B). In het eerste deel wordt respondenten gevraagd op basis van zelfrapportage een indicatie te geven van hun vakkennis en ervaring. Delen twee en drie zijn gewijd aan het schatten van respectievelijk de benodigde inspanningen op het gebied van databeheer en het hergebruikpotentieel bij diverse maatschappelijke vraagstukken binnen een geobjectiveerd en voor de respondenten herkenbaar

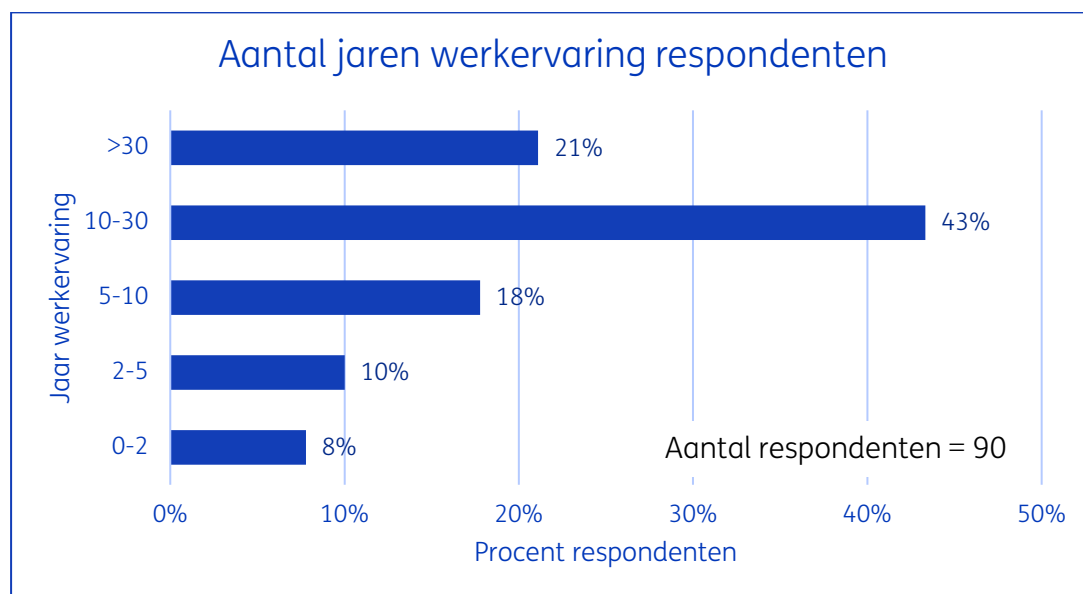
referentiekader ('meetlat'). Beide variabelen hebben we onafhankelijk van elkaar uitgevraagd, zodat eventuele afhankelijkheden niet kunnen leiden tot dubbelzinnigheden in de antwoorden (zie bijvoorbeeld Karger et al., 2021). Bij de borgingsinspanning kunnen respondenten ervanuit gaan, dat de gegevens altijd zullen worden hergebruikt. Omgekeerd kunnen respondenten bij het schatten van het hergebruikpotentieel ervanuit gaan, dat de borgingsinspanning is uitgevoerd. Deel vier biedt respondenten de mogelijkheid om hun inzichten en suggesties over het vraagstuk in hun eigen woorden te delen met het projectteam. In totaal bestaat de enquête uit zesenzeventig vragen. Op basis van een pilot vooraf door het projectteam is de gemiddelde invultijd vastgesteld op ongeveer een uur. De invultijd per respondent is niet bijgehouden en de enquête-analyse is niet vooraf formeel vastgelegd.

2.3 Representativiteit

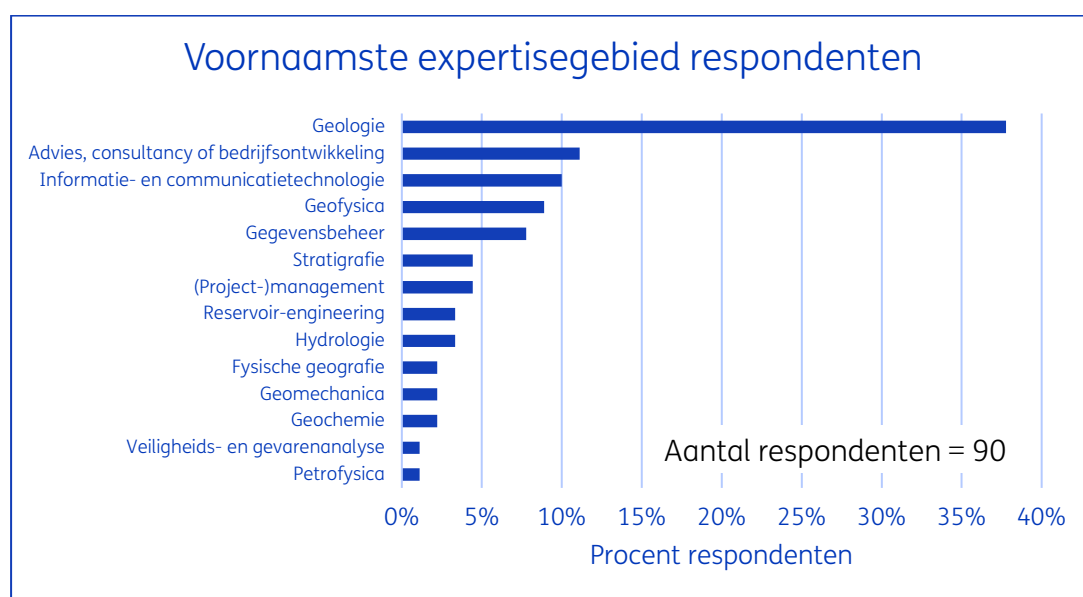
De enquête is uitgezet over de volledige Geologische Dienst (circa 290 personen). De respondenten zijn eerst per mail en later persoonlijk benaderd om de enquête online in te vullen. De totale doorlooptijd van het uitzetten van de enquête was ongeveer een maand. In totaal hebben 90 collega's de enquête volledig ingevuld (~31% responsgraad). Onvolledig ingevulde enquêtes zijn niet toegelaten tot de analysefase en ook niet apart geregistreerd.

De intrinsieke motivatie van de respondentengroep schatten we gemiddeld in op medium tot hoog op basis van de benodigde inspanning om de enquête helemaal in te vullen (een uur om zesenzeventig vragen te beantwoorden). Vervolgens zijn we de robuustheid hiervan nagegaan door de analyse te herhalen en daarbij de respondenten (N=4) uit te sluiten, die onvoldoende gebruik hebben gemaakt van de meetlat (75% of meer van de datatypen bij het hergebruikpotentieel #must have). De resultaten blijken robuust voor deze correctie.

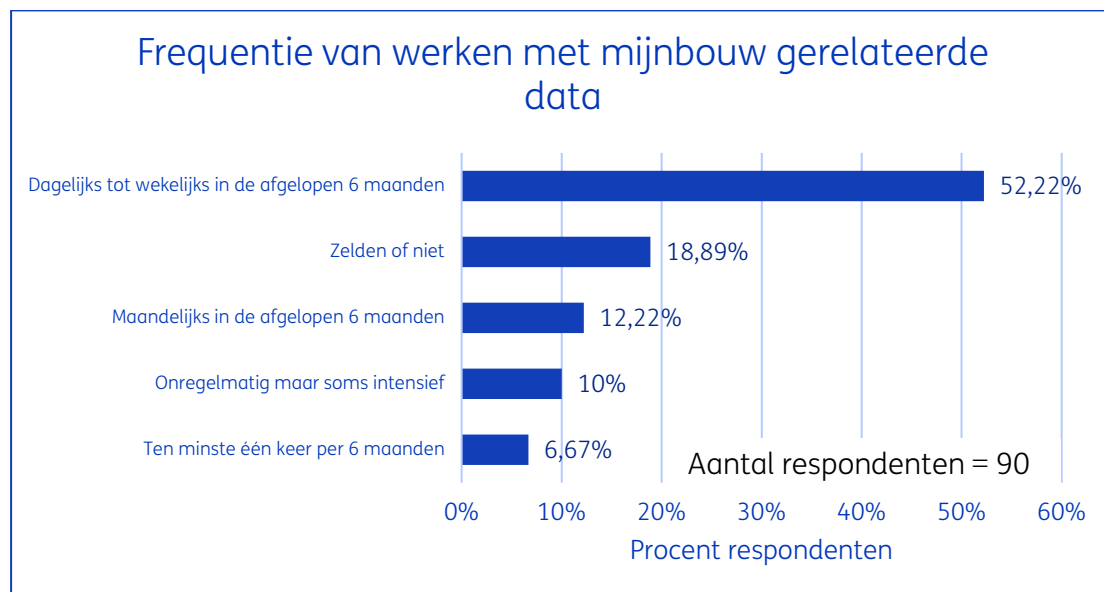
Een andere indruk van de representativiteit en diversiteit van de respondentengroep verkrijgen we op basis van drie variabelen: het aantal jaren werkervaring, de voornaamste expertisegebieden en de frequentie van werken met mijnbouwgegevens of interpretaties. De groep respondenten heeft professionele ervaring: het merendeel (64%) heeft meer dan 10 jaar professionele werkervaring, circa een vijfde (21%) geeft aan meer dan 30 ervaring te hebben (Figuur 2.1). Tegelijk is de respondentengroep op dit aspect ook voldoende divers (de groep tot 5 jaar ervaring betreft krap een vijfde (18%)).



Figuur 2.1: Het aantal jaren professionele werkervaring van de respondenten.



Figuur 2.2: Het expertisegebied waarvan de respondenten hebben aangegeven de meeste ervaring mee te hebben.



Figuur 2.3: De frequentie waarmee respondenten in de dagelijkse werkzaamheden met mijnbouw gerelateerde data werken.

Naast voldoende professionele ervaring biedt de respondentengroep diversiteit in expertisegebieden (Figuur 2.2). Het zwaartepunt ligt bij de geologie (~38%) en andere aardwetenschappelijke expertisegebieden, zoals geofysica en reservoir engineering. Ongeveer een derde van de respondenten rapporteert als voornaamste expertise een andersoortig expertisegebied in het aardwetenschappelijke domein, zoals advisering, IT en (project-)management.

Gezien het vraagstuk is de frequentie waarmee de respondenten met mijnbouwdata werken van belang (zie Figuur 2.3). Hierbij geeft een meerderheid van de respondenten (~52%) aan, dat ze dagelijks tot wekelijks met data hebben gewerkt gedurende de afgelopen zes maanden. De drie type datawerkzaamheden, waar de respondenten op dit vlak de meeste ervaring mee hebben, zijn dataverwerking, analyse en interpretatie (~44%), datavisualisatie (~16%) en datamanagement (~11%). De respondentengroep heeft gemiddeld genomen dus voldoende affiniteit met data om het belang van dit vraagstuk voor henzelf of de Nederlandse samenleving op waarde te kunnen schatten (wat de intrinsieke motivatie om de enquête zorgvuldig en naar waarheid in te vullen kan verhogen).

Voor een eerste verkenning van de borgingsinspanning en hergebruikpotentieel op diverse maatschappelijke thema's biedt de respondentengroep voldoende omvang, ervaring en diversiteit, waarbij we terdege beseffen dat voor het voorliggende vraagstuk de respondentengroep niet volledig representatief kan zijn. Immers, veel en divers georiënteerde organisaties (inclusief nog op te richten bedrijven) en eindgebruikers kunnen straks behoefte hebben aan de data.

2.4 Duiding van de enquêteresultaten

Op basis van de enquêteresultaten (Bijlage C) zijn vier hoofdbevindingen opgesteld. Deze worden hieronder puntsgewijs toegelicht en geduid.

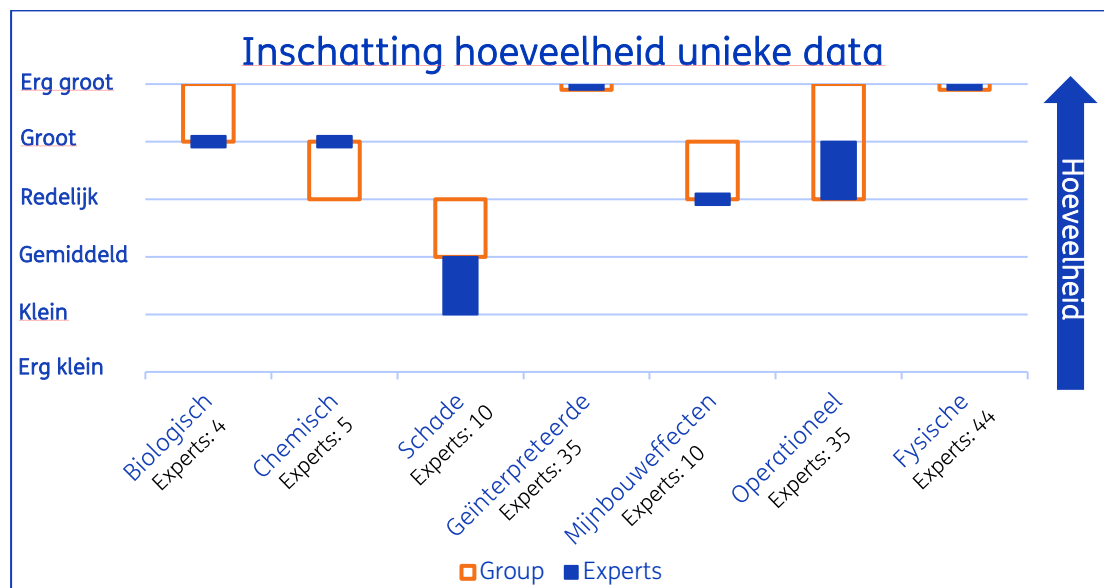
1. De totale hoeveelheid unieke data wordt als groot ingeschat. Een groot deel daarvan is momenteel in beheer bij de mijnbouwbedrijven.

De respondenten hebben hun eerste inschatting gegeven van de totale hoeveelheid unieke data per datatype. Deze inschatting omvat zowel het deel, dat bij de mijnbouwbedrijven in beheer is, als wat bij andere gegevensopslagplaatsen (waaronder de GDN) in beheer is. De totale hoeveelheid unieke data wordt overwegend als groot tot zeer groot ingeschat (Figuur 2.4). Hierbij springen geïnterpreteerde en fysische data eruit als grote datasets (met een smalle onzekerheidsbandbreedte). Nazorgdossiers (schadeafhandelingen en mijnbouweffecten) worden ingeschat als een kleine tot gemiddelde hoeveelheid. In het algemeen geven de experts een vergelijkbaar beeld met een (nog) smallere bandbreedte. Alleen voor de schadedata geven de experts aan, dat de hoeveelheid unieke data significant lager is.

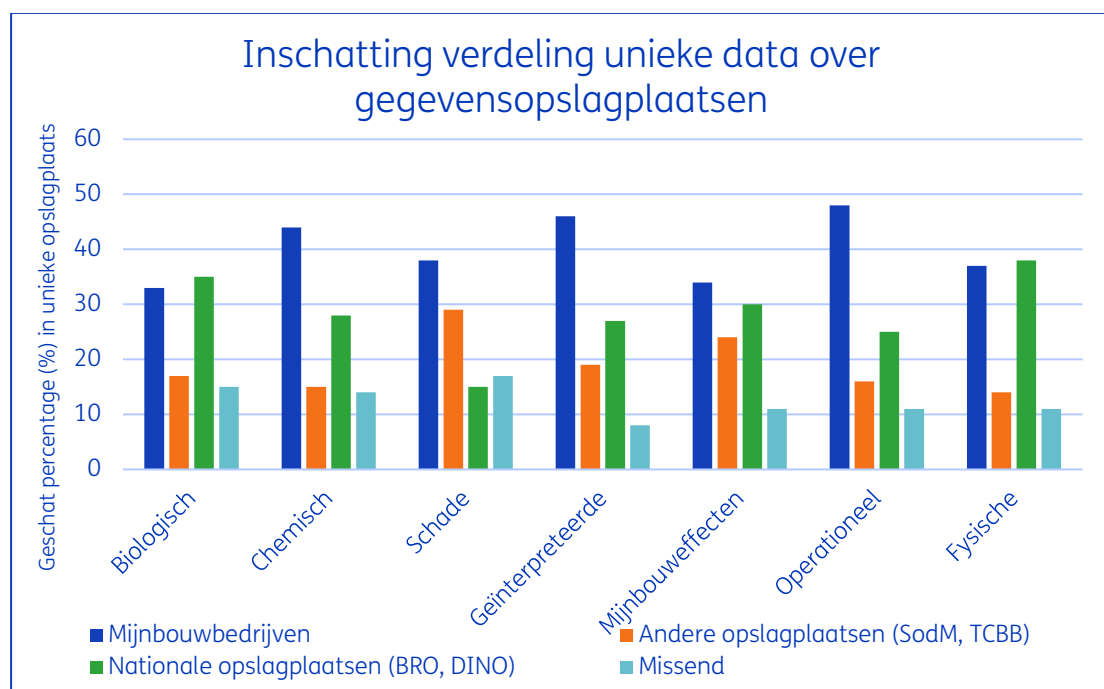
De respondenten zijn ook bevraagd op de locatie waar de unieke gegevens op dit moment geborgd worden. Specifiek is de respondenten gevraagd een procentuele inschatting te maken van de verdeling van de data over vier gegevensopslagplaatsen, waarbij het totaal op 100% moet uitkomen:

- Nationale databanken, waar geowetenschappelijke gegevens over de diepe en ondiepe ondergrond van Nederland worden verzameld en beheerd op één centrale plek (BRO en DINO);
- Gecombineerde databanken van de exploitanten, die momenteel in Nederland actief zijn;
- Andere dataopslagplaatsen in Nederland (o.a. SodM, EBN, TCBB, universiteiten);
- Missende en ontbrekende data.

Uit de resultaten hiervan (Figuur 2.5) blijkt, dat een significant gedeelte van de data reeds bij GDN, SodM of de TCBB in beheer is. Tegelijk is nog een groot deel (~30% - ~60%) bij de exploitanten in beheer. We zien dat met name chemische, geïnterpreteerde en operationele data voor bijna het merendeel nog bij de mijnbouwbedrijven aanwezig is. Dit ligt in de lijn der verwachting: deze data zijn vaak bedrijfsgevoelig (denk aan interne studies naar nieuwe ontwikkelmogelijkheden en de operationele prestaties van een bedrijf) en beperkt van waarde voor de gestelde beleids- en toezichtdoeleinden.



Figuur 2.4: De eerste inschatting van de totale hoeveelheid unieke data per datatype door de respondenten. Hierin is onderscheid gemaakt tussen de antwoorden van experts (op basis van zelfrapportage) en de bandbreedte van minimaal 50% van de respondenten ($N \geq 45$) rondom het meest gegeven antwoord.



Figuur 2.5: Inschatting van de verdeling van unieke data over verschillende gegevensopslagplaatsen. De inschatting betreft het mediane antwoord van de respondenten ($N = 90$).

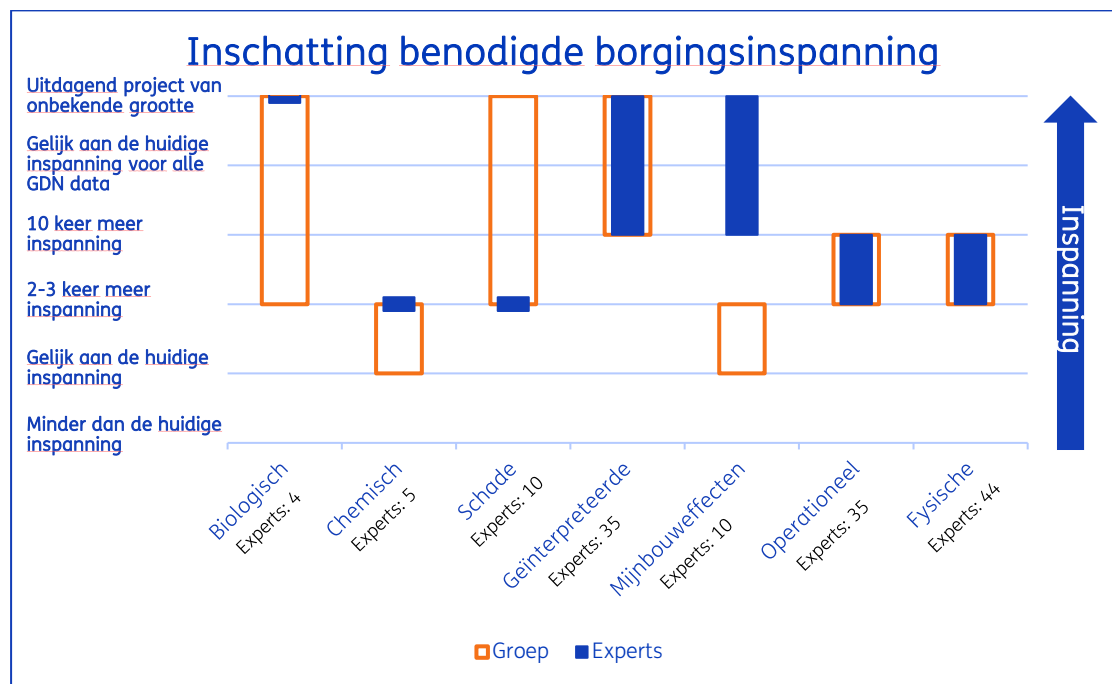
2. Alle datatypes zijn moeilijk reproduceerbaar. En als opnieuw verzamelen wel kan, dan tegen hoge kosten.

We hebben voor alle datatypes gevraagd wat hun beeld is aangaande drie datarisico's: verlies van bepaalde kennis en vaardigheden, reproduceerbaarheid van de data/interpretaties en de kosten van hernieuwde inwinning. Het (toekomstige) verlies van bepaalde kennis en vaardigheden rondom gaswinning kan zowel de nazorg van de gaswinning als het hergebruik van de data en interpretaties bemoeilijken (zoals blijkt uit onderzoek in andere sectoren, bijvoorbeeld Manning et al., 2021, Liakat en Hassan, 2021, Galan, 2023, Leal-Rodríguez et al., 2015, Rashid et al., 2019). Daarnaast is de reproduceerbaarheid, ofwel de mate waarin de mijnbouwinformatie vervangen kan worden door hernieuwde inwinning en interpretatie, niet vanzelfsprekend vanwege (1) het dynamische karakter van een deel van de data (zo kan bijvoorbeeld de initiële reservoirdruk na winning niet meer worden vastgesteld) en (2) de toenemende druk op het ruimtegebruik op land (IenW en EZK, 2018) en op zee (IenW, LNV en EZK, 2020). De kosten van de eventuele, hernieuwde inwinning (indien mogelijk) moeten te zijner tijd worden afgewogen tegen de gebruiksbaten. Deze kosten (mankracht en materialen) kunnen - in vergelijking met de (verwachte) kosten voor de borgingsinspanning van de bestaande data/interpretaties - naar de toekomst in principe zowel afnemen (door innovaties, vooral gedreven door opkomende energietechnologieën, zie bijvoorbeeld Santhakumar et al., 2023) als toenemen (door schaarste of door verlies aan kennis en vaardigheden, zie referenties hierboven).

Uit de enquêteresultaten komt een duidelijk beeld naar voren: opnieuw inwinnen kan slechts beperkt en tegen hoge kosten, terwijl het verlies van de benodigde kennis en vaardigheden voor hergebruik in de periode tot 2050 reëel is. Dit beeld is gelijk over alle datatypes, wat de noodzaak van het (tijdig) borgen van (zo veel mogelijk) data en interpretaties onderstreept. Hierbij merken we op, dat er verschillende niveaus zitten tussen weggooien, niet verzamelen, volledige verzameling, volledige borging met uitgifte en maatwerk voor optimaal hergebruik in het publieke domein.

3. Voor alle datatypes geldt dat een grotere inspanning nodig is voor het borgen en afdoende ontsluiten voor hergebruik dan wat nu nodig is voor de huidige mijnbouwdata, die GDN beheert.

Voor het maken van een inschatting voor de borgingsinspanning is de respondenten gevraagd een schatting te maken van de jaarlijkse inspanning voor databorging tot het jaar 2050 in het scenario, dat de GDN alle unieke data van een specifieke datatype (conform Figuur 2.6) van de exploitanten ontvangt. De meetlat voor de respondenten om de borgingsinspanning aan te refereren is de huidige GDN borgingsinspanning. De respondenten verwachten voor alle datatypes, dat een grotere inspanning nodig is dan de huidige inspanning (Figuur 2.6). Dit geldt ook voor datatypes, die GDN nu al deels beheert. De minste inspanning wordt verwacht voor chemische en operationele data (gelijk aan tot 2 à 3 keer de huidige inspanning). Specifiek voor het datatype mijnbouweffecten schatten de experts de inspanning (veel) hoger in dan de groep. Dit is een punt van aandacht op moment, dat borging van dit datatype aan de orde komt.

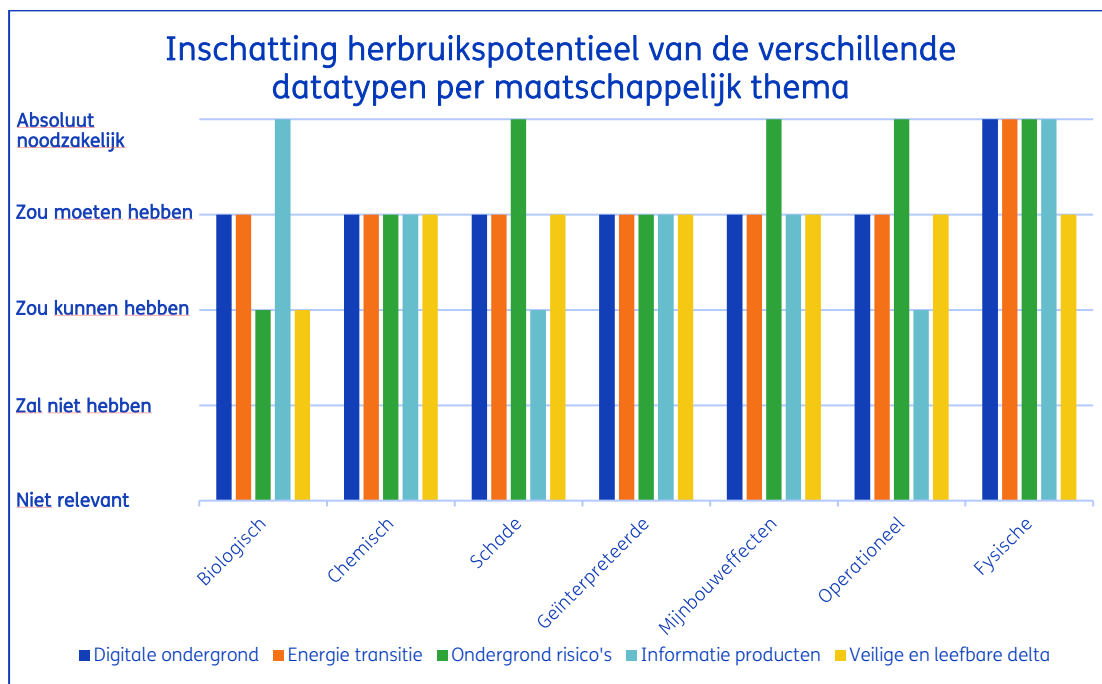


Figuur 2.6: De eerste inschatting van de borgingskosten per datatype tot het jaar 2050, waarbij de huidige (jaarlijkse) gegevensbeheerinspanning van de GDN voor hetzelfde datatype als referentie worden gebruikt. Hierin is onderscheid gemaakt tussen de antwoorden van experts (op basis van zelfrapportage) en de bandbreedte van minimaal 50% van de respondenten ($N \geq 45$) rondom het meest gegeven antwoord.

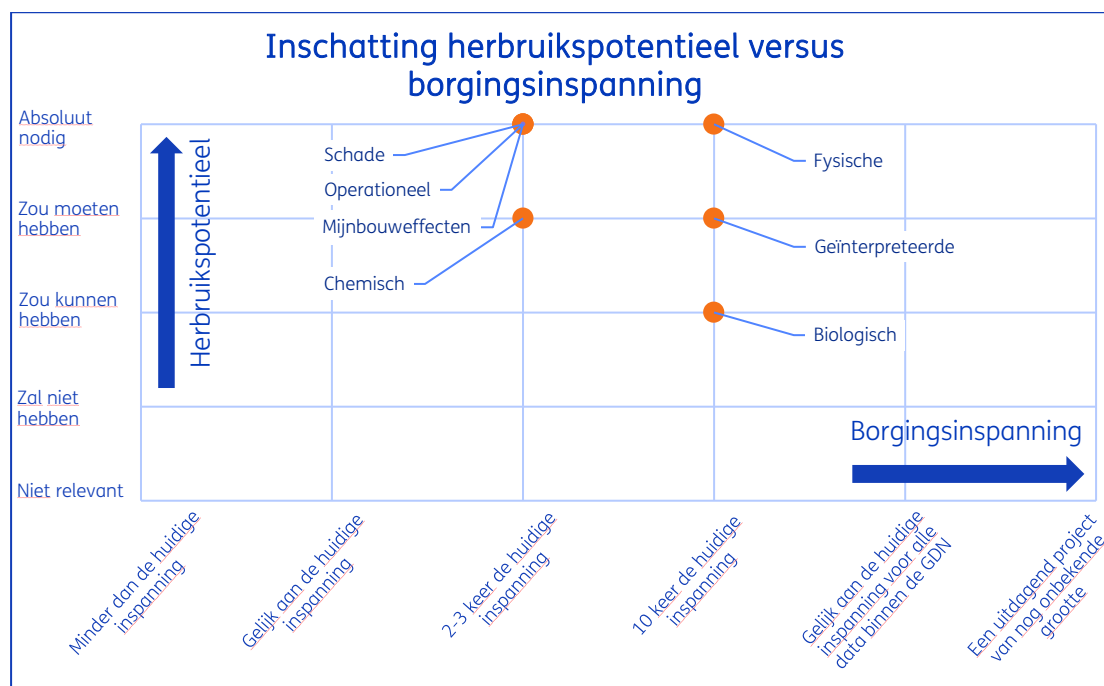
4. Het prioriteren van de inname of uitgifte kan primair op basis van de tamelijk fijnmazige relatie tussen het hergebruikspotentieel per datatype (# must haves) per maatschappelijke subthema.

Bij de bepaling van de waarde van de data/interpretaties voor hergebruik hebben de respondenten hun inschatting gegeven per datatype en per maatschappelijk thema (Figuur 2.7). Hieruit blijkt dat de respondentengroep de meeste waarde hecht aan het borgen van fysische data. Voor de andere datatypes is de waarde nagenoeg gelijk, maar verschilt per maatschappelijk subthema.

Het prioriteren van de inname of uitgifte kan op basis van de relatie tussen het hergebruikspotentieel per datatype (# must haves) per maatschappelijke subthema. We hebben als alternatief gekeken naar de combinatie per datatype van het hergebruikspotentieel en de borgingsinspanning (Figuur 2.8). Het onderscheidend vermogen hiervan beoordelen we als ontoereikend voor een zinvolle prioritering.



Figuur 2.7: Inschatting van hergebruikspotentieel van de verschillende datatypen onderscheiden naar maatschappelijk thema. De inschatting betreft het mediane antwoord van de respondenten (N = 90).



Figuur 2.8: Inschatting hergebruikspotentieel van de verschillende datatypen afgezet tegen de verwachte borgingsinspanning. De inschattingen zijn de mediane waarden van de respondenten (N = 90) over alle maatschappelijke thema's.

3 Handelingsperspectief

Het systematisch innemen, beheren en vrijgeven van gegevens en informatie doet GDN meerdere decennia en sinds 2003 op basis van wettelijke taken. Zodoende is een significant gedeelte van de data reeds bij GDN of bij een aan de Staat gelieerde uitvoeringsdienst in beheer (Figuur 2.5). Tegelijk beheren de exploitanten zelf ook nog een aanzienlijk deel (~30% - ~60%) van de gegevens en informatie. De overdracht van zoveel en andersoortige gegevens in relatief korte tijd (de komende 5 tot 15 jaar) is een uitdaging. In dit hoofdstuk gebruiken we onze ervaring met relevante projecten en processen om meer inzicht te krijgen in de verwachte kosten, de organisatorische aspecten, die een rol kunnen spelen bij de benodigde opschaling en de prioritering.

In paragraaf 3.1 geven we een overzicht van de huidige activiteiten. Het laat zien welke data we momenteel borgen en op basis van welk type afspraken dat gebeurt. Dit vormt een inleiding op paragraaf 3.2, waarin we aan de hand van voorbeelden laten zien welke keuzes hierbij komen kijken. In paragraaf 3.3 bespreken we de belangrijkste factoren, die de kosten van en de impact op de GDN organisatie van het borgen van (alle) mijnbouwgegevens bepalen, waarbij we onderscheid te maken tussen fysieke en digitale data. In paragraaf 3.4 werken we voor vier scenario's de kosten voor het borgen van de digitale data nader uit. Deze scenario's vormen op hun beurt de bouwstenen voor het handelingsperspectief. Dat werken we in paragraaf 3.5 uit in zes perspectieven.

3.1 GDN ervaring met databeheer

De wettelijke gegevensverstrekkingverplichting vormt de grondslag voor de datalevering aan GDN als gedelegeerd databeheerder namens de Staat. Om invulling te kunnen geven aan deze taak heeft de GDN in de loop der jaren detailafspraken gemaakt met dataleveranciers over de wijze van aanlevering. Voorbeelden hiervan zijn:

- *Afspraak (toen RGD) ⇔ NAM*: Over opslag NAM kernmateriaal in het NAM kernhuis, bij gebrek aan capaciteit in GDN kernhuis. Na verstrijken confidentialiteitstermijn is deze data in het NAM kernhuis toegankelijk voor het publiek;
- *Afspraak GDN ⇔ NAM*: Over het vervroegd beschikbaar stellen van confidentiële bewerkingen van seismiek ten behoeve van de veiligheid rondom boorlocaties, waarbij de boorgatdata direct beschikbaar komt voor de NAM (in plaats van na de confidentialiteitsperiode van 5 jaar);
- *Afspraak EBN ⇔ GDN*: Over levering van alle, voor EBN bestemde, digitale logdata aan GDN. EBN kan vervolgens alle logdata (openbaar en confidentieel) bij GDN opvragen;
- *NOGEPA/ElementNL ⇔ GDN*: Handvest met overzicht van de (verplichte) periodieke aanlevering van data, aangevuld met onderlinge afspraken over datalevering.

Mijnbouwgegevens, die door andere aan de Staat gerelateerde organisaties worden ingenomen (bijvoorbeeld het SodM), worden bij die organisaties opgeslagen en beheerd (Figuur 2.5).

Zoals eerder aangegeven onderscheiden we fysieke en digitale dataprocessen, want het innemen van fysieke data is bewerkelijker (metadata moet gegenereerd worden) en ook de opslag en vindbaarheid vergen intensievere handelingen dan voor digitale data. De fysieke

datastroom omvat onder meer kernmateriaal en monsters en wordt opgeslagen in het GDN kernhuis. Verreweg de meeste data wordt tegenwoordig digitaal aangeleverd. Dit betreft onder meer rapportages, geofysische data, boorgatdata, maandelijkse productiecijfers en de jaarlijkse aanlevering van prognoses met aanvullende gegevens. Hieronder volgt daarvan een korte beschrijving.

Geofysische data

Dit betreft meestal data verkregen uit geofysische verkenningsonderzoeken naar de aanwezigheid van delfstoffen of aardwarmte, waarbij seismische data of seismiek het meest voorkomend is. Voor nieuw geschoten seismische datasets gaat het opvragen gestructureerd; GDN vraagt elk jaar in januari bij SodM een lijst op van de gemelde acquisities en neemt vervolgens contact op met de betreffende exploitant. Hierdoor is onze collectie van eerste bewerkingen van nieuwe seismische data nagenoeg compleet.

Elke nieuw aangeleverde seismische dataset bestaat gewoonlijk uit de ruwe velddata en het bewerkte eindproduct (tussenproducten blijven in beheer bij de exploitant). De informatie wordt vrijgegeven na 5 jaar en na 10 jaar voor niet-exclusieve seismische data sets (multi-client of spec-surveys). Een aparte categorie binnen de seismiek vormen de herbewerkingen van de originele datasets (reprocessings). Deze activiteiten worden gemeld bij de jaarlijkse gespreksronde met de exploitanten. Deze meldingsroute is minder gestructureerd en daardoor kwetsbaarder qua compleetheid.

Voor de overige seismische velddata is vastgelegd (artikel 11 van de Mijnbouwregeling), dat ze op aanvraag worden geleverd door de exploitant. Het innemen, controleren en opslaan van velddata is arbeidsintensief. GDN heeft op dit moment voor een klein percentage (~5%) van alle beschikbare seismische data sets de velddata opgevraagd en in beheer genomen. Momenteel is er niet voldoende capaciteit om naast het verwerken van de directe aanvragen voor velldatasets ook proactief alle velldatasets op te vragen. Dit vormt een risico bij abrupt vertrek van exploitanten uit Nederland.

Boorgatdata

Bij boorgatdata gaat het om data verkregen bij het aanleggen, repareren, verlaten of monitoren van boorgaten. Dit omvat een breed scala aan verschillende datatypen (Tabel 3.1), waaronder ook fysieke data, zoals boorkernen, boorwandkernen en gesteentemonsters.

De wettelijke leveringstermijnen worden bij dit datatype doorgaans goed nageleefd, zeker voor routinematig verkregen data (zoals end of well reports, composite well logs en MWD/wireline metingen). Maatwerkdata (zoals bepaalde onderzoeksrapporten en specialistische boorgatmetingen/logs) worden in de praktijk minder betrouwbaar aangeleverd, zodat de GDN deze data geregeld actief moet opvragen bij de desbetreffende exploitant.

Tabel 3.1: Overzicht van boorgatdata.

Boorgat	Profiel
	Resultaat boorgatmetingen
	Resultaat metingen aan monsters
	Resultaat productie- & injectietesten
Monsters	Gesteentemonsters

Productiedata

Productie- en opslagdata van voor 2003 zijn incidenteel (van een specifieke exploitant, een specifieke dataleverancier en/of voor een bepaalde periode) in de GDN databank opgenomen en worden niet publiek vrijgegeven. Sinds 2003 worden op maandelijkse basis de productie- en opslagcijfers ingenomen en beheerd. Cijfers behorend bij de objecten voorkomens, licenties, mijnbouwwerken en putten worden geregistreerd. Deze datareeksen zijn compleet. Sinds begin 2022 geschiedt aanlevering van productiedata via een speciaal hiervoor ingericht online innameportaal. Hierdoor zijn de dataleveringen sindsdien beter gestructureerd, tijdiger en eenvoudiger controleerbaar op kwaliteit.

De exploitanten voeren op regelmatige basis correcties uit op reeds aangeleverde data. Deze correcties worden veelal gerapporteerd en doorgevoerd in de GDN databank, maar het komt incidenteel voor, dat deze communicatie uitblijft.

Prognoses en aanvullende gegevens

Onder art 113 van de mijnbouwregeling dient elke vergunninghouder van een koolwaterstoffenvergunning (zowel opsporing als winning) jaarlijks een overzicht aan te leveren van de volgende gegevens (voor zover van toepassing):

- Projectmatige prognose gas, olie of condensaatproductie in aangetoonde voorkomens;
- Overzicht putten en putstatus;
- Gemeten drukgegevens (dus niet reservoirdruk per sé, wel alle metingen);
- Prospectieve structuren (niet aangetoond gasvoorkomens).

De aanlevering verloopt meestal redelijk vlot, al leiden vragen over de gegevens geregeld tot enige vertraging bij de inname. De prognosegegevens van bestaande velden worden na 10 jaar vrijgegeven (conform mbr. 116).

Inzichten

De GDN neemt op regelmatige basis al veel mijnbouwgegevens en -informatie in. Vanuit die taak zijn de volgende inzichten van belang om mee te nemen bij eventuele opschaling:

- Duidelijkheid vooraf over de betrouwbaarheid van nieuwe datatypen (en de termijnen daarvoor) is een vereiste.
- Data, die (kunnen) worden ingenomen via een online innameportaal, zijn beter gestructureerd, tijdiger en eenvoudiger te controleren op kwaliteit. Dit geldt in het bijzonder voor de inname van herbewerkingen of correcties van data.
- Het innemen, controleren en opslaan van velddata (tapes) is arbeidsintensief, niet gemakkelijk op te schalen (zeker niet onder tijdsdruk) en vormt daarmee een risico voor borging bij abrupt vertrek van exploitanten.

3.2 Voorbeelden van eerdere keuzes

Hieronder bespreken we relevante voorbeelden van eerder gemaakte keuzes ten aanzien van de impact op de organisatie en kosten van het innemen, beheer en vrijgeven bij eventuele opschaling.

Maatschappelijke kostenbatenanalyse

Bij de invoering van de BRO is tweemaal een maatschappelijke kosten-baten analyse (MKBA) uitgevoerd. In beide gevallen blijken de baten beduidend hoger dan de kosten, met het positieve saldo tot gevolg. De baten betreffen voornamelijk afname in projectkosten van ruimtelijke projecten, waarbij het lastig blijkt om in te schatten waar deze baten precies gaan neerkomen (markt vs. overheid) (Hulsker et al., 2021). Voor de huidige vraagstelling geldt dat de baten zullen komen uit een veel breder scala aan toekomstige maatschappelijke toepassingen met een onzeker toekomstbeeld (bijvoorbeeld door de onzekerheid in transitiepaden voor de energietransitie). Vandaar onze keuze voor een enquête onder de GDN experts (zie hoofdstuk 2) en niet direct via een MKBA (paragraaf 1.5). Indien nodig, kan in een later stadium alsnog een MKBA worden uitgevoerd. Een ander verschil zijn de datastromen. In de BRO betreft dit primair de aanlevering van huidige en toekomstige gegevens door diverse bronhouders, terwijl onze vraagstelling bestaande datasets betreft, die afkomstig zijn van de gasexploitanten in Nederland.

Standaarden

GDN sluit aan bij standaarden voor data en dataverwerking (Lance et al., 2011, Wilkinson et al., 2016, KNAW et al., 2018, Chapela, 2023). Deze standaarden vragen onder meer van ons, dat we streven naar behoud van originele data. Daarnaast houdt de GDN waar mogelijk rekening met industrie-initiatieven (OSDU⁴, Lewis et al., 2023) en draagt onze organisatie bij aan de ontwikkeling van internationale standaarden (Beardsmore et al., 2021). Zo zijn we vanwege de invoering van het internationale PRMS systeem (om te komen tot internationaal vergelijkbare reserveschattingen van beursgenoteerde bedrijven, zie Etherington en Ritter, 2007) overgegaan van prognoses op veldbasis naar projectmatige gasprognoses (EZX, 2022).

Nazorg steenkolenwinning Limburg

Een voorbeeld van onvoorzien hergebruik van mijnbouwgegevens en -informatie is de nazorg steenkolenwinning Limburg met het oog op de na-ijlende effecten (Esteves Martins et al., 2021). In dit kader zijn in het afgelopen decennium duizenden originele, handgetekende mijnkaarten gescand en voorzien van georeferentie, zodat de huidige generatie experts en bestuurders bekend blijft met de daarop getoonde informatie. Daarbij was een punt van zorg, dat de legenda van de mijnkaarten op enig moment vrijwel alleen nog door gepensioneerde experts kon worden gelezen. Dergelijke teloorgang van kennis en ervaring komt vaker voor, ook in andere sectoren (Manning et al., 2021, Rashid et al., 2019, Klammer en Gueldenberg, 2019) en vormt een risico voor onvoorzien hergebruik. Omgekeerd vormt hergebruik de beste bescherming tegen dergelijke amnesie (zie ook paragraaf 2.4).

Geothermie-exploratie

Een ander voorbeeld van het gebruik van bestaande data komt uit de geothermie-exploratie. Voor deze toepassing hebben exploitanten intensief gebruik gemaakt van de eerder verzamelde gegevens uit de gasexploratie. Dit betreft in het bijzonder seismiek en boorinformatie, waaronder de logdata en verschillende typen boormonsters. Daardoor blijft

⁴ open subsurface data universe, www.osduforum.org

de inwinning van nieuwe gegevens (bijvoorbeeld via het SCAN project (Békési et al, 2020) en het TKI project WarmingUp) grotendeels voorbehouden aan de witte vlekken op de kaart: daar, waar de gasexploratie om geologische of economische redenen niet van de grond is gekomen. Wel is voor de nieuwe toepassing gewoonlijk herbewerking van de originele data nodig, bijvoorbeeld omdat de eerdere bewerkingen geoptimaliseerd zijn voor een ander dieptebereik.

Opwerking van informatie voor hergebruik

Een belangrijke en arbeidsintensieve stap is opwerking van bestaande datasets voor nieuwe, specifieke toepassingen. Dit is maatwerk en wordt in principe door de (eind)gebruikers gedaan. Een voorbeeld is het dataportaal van de Groninger Bodembeweging (het Gasbevingen Portaal⁵), waarbij de gebruiker specifieke gegevens uit de DINO databank heeft gecombineerd met gegevens uit andere databanken (onder meer de KNMI bevingen databank⁶) om de eigen achterban te voorzien van actuele informatie over de seismiteit door gaswinning in Groningen. In meer complexe situaties kan de GDN dit soort initiatieven ondersteunen. Een recent voorbeeld daarvan is de publicatie van een publiek softwareprogramma, waarmee een seismische risicoanalyse kan worden uitgevoerd⁷ (de modelketen Groningen).

Veranderende opslagkosten seismiek

In 2003 is bij een aanpassing van de Mijnbouwwet de keuze gemaakt om van de seismiek alleen de ruwe metingen en de eerste bewerking (digitaal) te borgen bij GDN. Voor het borgen van de datatapes en de tussenresultaten waren de opslagkosten destijds te hoog (zie hoofdstuk 2). Door de gedaalde kosten voor geheugenopslag kunnen belangstellenden sinds negen jaar de informatie op de datatapes aanvragen bij de GDN (Meekes, 2015). Dan vraagt de GDN de betreffende tape op bij de exploitant en neemt deze in beheer. Momenteel hebben we de data van ongeveer 5% van de datatapes in beheer, de overige 95% wordt door de exploitanten beheerd (zie ook paragraaf 3.1). Het borgen van al deze datatapes is een punt van aandacht. Hiervoor kan uitstel gunstig zijn vanwege verder dalende kosten voor geheugenopslag (Wet van Moore, zie ook Helm, 2022), maar zeker is dit niet (vanwege de eveneens exponentiele toename in volume van de geschoten seismiek, Chapela, 2023).

Inzichten

GDN is aangesloten bij diverse standaarden voor data en dataverwerking, die onder meer van ons vragen dat we streven naar behoud van originele data.

Bovenstaande voorbeelden laten de waarde zien van de in het verleden vergaarde gegevens, ook voor onvoorziene toepassingen. De tot op heden gemaakte keuzes zijn het gevolg van ontwikkelingen binnen én buiten de mijnbouwsector, al is daarbij vooralsnog geen sprake van een integrale afweging tussen de kosten enerzijds en het brede maatschappelijke belang (conform Bijlage A) anderzijds.

De voorbeelden zijn daarnaast illustratief voor een ander belangrijk aspect: gegevens en informatie afkomstig uit de ene toepassing hebben zonder nadere opwerking gewoonlijk beperkt nut voor andere of nieuwe toepassingen. Het opwerken naar aangepaste of nieuwe informatieproducten (Van der Meulen et al., 2013, Maljers et al., 2015) stimuleert het beoogde hergebruik en gaat het verloren gaan van kennis tegen. Dergelijke opwerking kan

⁵ <https://bevinggevoeld.nl/gasbevingen>

⁶ <https://www.knmi.nl/nederland-nu/seismologie/aardbevingen>

⁷ <https://www.nlog.nl/en/source-code-groningen-model-chain>

beperkt (zoals het opwerken van de deviatiegegevens bij boorgaten) of meer uitgebreid zijn (andere stratigrafie bij geologische modelinterpretaties, Jacobsen et al., 2020).

3.3 Fysieke en digitale data

In deze paragraaf behandelen we voor de fysieke en digitale dataprocessen (inname, ontwikkeling, beheer, uitgifte) de factoren, die bepalen hoe de kosten voor het borgen van de mijnbouwgegevens en informatie eruit kunnen gaan zien.

Fysieke data

Onder de fysieke data verstaan we onder meer kernen, slijpplaatjes, olie-, gas en watermonsters, gesteentemonsters en boorgruis, lakprofielen, fysieke tapes (seismiek) en papieren data, die nog niet is gescand of gedigitaliseerd. Dit betreft vooral de chemische, biologische en fysische datatypes. Hoe hoog de kosten voor het innemen, borgen en beheren van deze fysieke data uitvallen hangt vooral af van een drietal te maken keuzes:

- Opslaglocatie: Centrale locatie versus verspreid over meerdere locaties;
- Borging: Alles bewaren versus alleen unieke hoogkwalitatieve data;
- Vrijgave: Fysiek versus digitaal vrijgeven.

Opslaglocatie

De keuze ligt voor om alle fysieke data op één plek te verzamelen, of dit verspreid over meerdere locaties te borgen. Uitbreiding, laat staan grootschalige uitbreiding, van de opslagcapaciteit op de huidige locatie van het GDN kernhuis in Zeist is niet mogelijk (Van Adrichem et al., 2023). Daardoor brengt de keuze om alle fysieke data op één plek te verzamelen (in een nieuw en aanzienlijk groter kernhuis) hoge kosten met zich mee. Als borging op meerdere locaties een optie is, kan er bijvoorbeeld gekeken worden of de huidige opslagen van de exploitanten (zoals het kernhuis van de NAM in Assen) deels behouden kunnen blijven. Dit reduceert de inspanning om het materiaal te transporteren en te kunnen borgen aanzienlijk en daarmee ook de hieraan verbonden kosten. In dat geval moet wel extra worden geïnvesteerd in een digitaal systeem voor inname en overdracht van materiaal.

Aanvullend moet er ook een keuze gemaakt worden in welke mate andere dan de huidige, fysieke datatypes opgeslagen gaan worden als onderdeel van het kernhuis. De hoeveelheid fysieke data anders dan de kernen, tapes en papieren data (zoals de gas- en vloeistofmonsters en de biostratigrafische / palynologische preparaten) is naar verwachting een fractie van de omvang van het volume van de kernen. Echter, deze andere fysieke data stelt deels aanvullende eisen aan de opslag (Van Adrichem et al., 2023), denk aan klimaatcontrole of extra veiligheidseisen bij opslag van brandbaar of explosief materiaal.

Tenslotte merken we op, dat in het GDN kernhuis ook ondergrondmateriaal wordt opgeslagen, dat niet onder de Mijnbouwwet valt (want afkomstig uit een ander dieptebereik). De opslagcapaciteit van het GDN kernhuis is nu al ontoereikend (zie paragraaf 3.1). Het is onwenselijk als opschaling van de mijnbouwgegevens zou leiden tot verdringing of zelfs de teloorgang van dat materiaal. Temeer, omdat een groot deel van het hergebruikspotentieel van de mijnbouwgegevens en informatie juist te herleiden is naar datzelfde dieptebereik (Figuur 2.7). Zo worden de laatste paar jaar enkele 10.000'en monsters per jaar ingenomen bij de aanleg van windmolenparken op zee en bij kritieke delen van diverse boringen op land. Ook de boringen op Nederlands grondgebied in kader van het Einstein Telescoop-project vallen hier onder.

Borging: keuze om alles op te slaan versus een selectie

Daar, waar fysieke data al aan ons is geleverd, weten we dat er vaak (zeker bij de kernen) vergelijkbare data bij de exploitant in beheer is gebleven. Bij de keuze om fysieke data in te

nemen, ontstaat er derhalve de mogelijkheid om unieke data te scheiden van niet-unieke data en daarbij een onderscheid te maken tussen kwalitatief goede en kwalitatief minder goede data. Hierbij moet besloten worden wat er met de niet-unieke, minder kwalitatieve data gebeurt: wordt slechts de unieke, kwalitatief goede data geborgd en de rest weggegooid (lage kosten), of wordt alles geborgd (hoge kosten)?

Vrijgave

Het kernhuismateriaal is momenteel op afspraak in te zien en te gebruiken voor onderzoek. Naast dit vrijgavemodel (a) bestaan ook andere mogelijkheden: de fysieke gegevens standaard met hoge kwaliteit digitaliseren en zo de gegevens digitaal vrijgeven (b), of (c) dit alleen op projectbasis/aanvraag doen. De digitale ontsluiting (dat wil zeggen beschikbaarheid van een digitale representatie van het materiaal in de vorm van foto's, scans etc.) van materiaal in het GDN kernhuis is momenteel gering. Investeren in hoogwaardige, digitale ontsluiting (opties b of c) kan er toe leiden dat onderzoek grotendeels (op afstand) uitgevoerd kan worden op de digitale representatie van het fysieke materiaal. Dan kan voor de borging van het fysieke materiaal op een goedkoper alternatief worden teruggegrepen.

Inzichten

Het maken van een kosteninschatting voor inname, beheer en vrijgave van fysieke data is vooral afhankelijk van de drie door KGG te maken keuzes (één of meerdere opslaglocaties, wat te doen met de niet-unieke, minder kwalitatieve data en de rol van digitalisatie bij vrijgave). Deze keuzes zijn behalve op de kosten ook relevant voor het tijdig kunnen organiseren van (onder meer) de benodigde opslagcapaciteit. Merk op: de verwachte voorbereidingstijd verschilt per exploitant, afhankelijk van wanneer zij het einde van de operationele fase bereiken (zie paragraaf 1.2 en 3.1).

Anders dan veel andere fysieke datasets kunnen de seismische datatapes en papieren dossiers gedigitaliseerd worden. De GDN heeft daar geruime ervaring mee. Na digitalisatie is het een onderdeel van digitale data.

Digitale data

Onder digitale data verstaan we de data, die digitaal aan ons aangeleverd wordt en ook zodanig kan worden opgeslagen en andersoortige data, als die na digitalisatie bij ons in digitale vorm beschikbaar komt. De hoogte van de kosten voor het innemen, borgen en beheren van digitale data hangt onder meer af van:

- Datatype: De benodigde inspanning per datatype;
- Datamodel: Wel/niet onderdeel van het bestaande datamodel;
- Vrijgave: Wel/niet opslaan in de GDN databank.

Datatype

De relatieve borgingsinspanning verschilt per datatype (Figuur 2.6). Dit punt werken we verder uit in de volgende paragraaf.

Datamodel

Als het datatype al volledig is opgenomen in het bestaande datamodel van de GDN databank, dan is het relatief weinig werk om nieuwe data van dit datatype in te nemen en toe te voegen aan de databank. Normaal gesproken is ook in de uitgifte dan al voorzien (via NLOG.nl), dus ook op dat vlak is de benodigde inspanning relatief klein. Als het datatype geen onderdeel van het bestaande datamodel uitmaakt én voor het realiseren van de hergebruiksbaten opname in de GDN databank wenselijk is, zal de borgingsinspanning drastisch toenemen; het datatype moet dan uitvoerig worden geanalyseerd en het datamodel aangepast, waarbij relaties en onderlinge afhankelijkheden met andere datatypen/objecten worden gelegd (en er moet een nieuw uitgiftesysteem worden opgezet).

Vrijgave

Als vuistregel bij hergebruik geldt van oudsher:

- Hoe beter geïntegreerd in de GDN databank, hoe duurzamer de data is geborgd, hoe gebruiksvriendelijker de data is ontsloten en hoe beter de gebruiker van de data in zijn wensen bediend kan worden;
- Hoe beter geïntegreerd in de GDN databank, hoe groter de borgingsinspanning en hoe hoger de kosten.

Deze vuistregel gaat zeker op voor gebruikers, die weinig vertrouwd zijn met AI/ML. Tegelijk groeit de gebruikersgroep, die hier wel ervaring mee heeft. Daarom is er een ontwikkeling gaande (onder andere via OSDU), waarbij bestaande, ongestructureerde data vaker rechtstreeks flexibel doorzoekbaar wordt aangeboden (en waar nodig automatisch gecategoriseerd). Dit biedt op termijn mogelijkheden om de kosten terug te brengen zonder afbreuk aan de hergebruiksbaten - met als randvoorwaarde de ontwikkeling van andere werkprocessen en een andere IT-infrastructuur (schaalbaar, cloud-based). Zo ver is het nog niet en tot die tijd blijft voor hergebruik een optimalisatieslag in de verhouding baten/kosten, inclusief prioritering, nodig: welke gebruikers worden (het eerst) bediend via de integratie van specifieke datatypen/sets in de GDN databank?

Inzichten

We moeten niet alleen goed kijken naar het opschalen van de huidige werkprocessen (met volledige integratie in de GDN databank als gebruikersideaal), maar tegelijk ook zoeken naar andere, innovatieve werkprocessen en modi van uitgifte om aan te kunnen sluiten bij de veranderende gebruikersgroepen, -wensen en -mogelijkheden. Zolang deze innovaties nog geen feit/in ontwikkeling zijn, volgen we twee sporen. In het eerste spoor richten we ons op (een zekere mate van) integratie in de GDN databank (zie paragraaf 3.4). In het tweede spoor willen we via één of meer pilots ervaring opdoen met het rechtstreeks aanbieden van bestaande, ongestructureerde data en met automatische categorisering aan specifieke gebruikersgroepen (zie paragraaf 5.4).

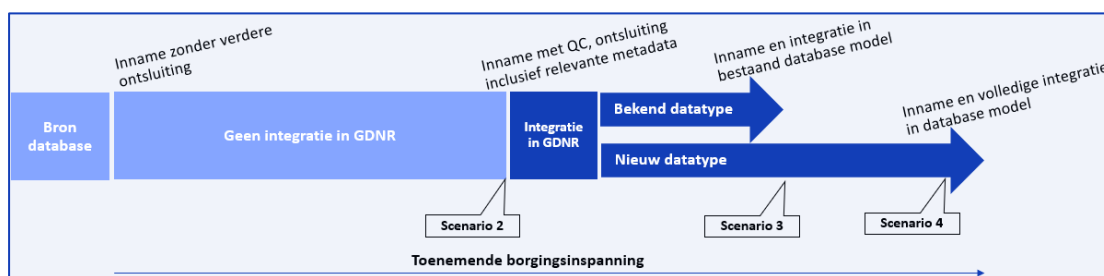
3.4 Kostenindicaties per scenario

Om een globale inschatting van de borgingsinspanning en bijbehorende kosten te kunnen geven voor de digitale data werken we voor de fysieke en digitale data vier scenario's uit.

Scenario's

De kosten hangen af van hoe de data geborgd en ontsloten gaat worden (paragraaf 3.3) en daarin onderscheiden we vier scenario's. Als de dataset is opgenomen in de GDN databank en (dus) deel uitmaakt van het bestaande datamodel, dan valt de dataset onder scenario 1. De volgende scenario's (2-4) definiëren we op basis van een toenemende mate van integratie in de GDN databank (Figuur 3.1):

- Scenario 1: Inname, opslag en vrijgave data binnen het bestaande datamodel;
- Scenario 2: Inname, opslag en vrijgave data zonder integratie;
- Scenario 3: Inname, opslag en vrijgave data met beperkte integratie;
- Scenario 4: Inname, opslag en vrijgave data met volledige integratie.



Figuur 3.1: Schema van de relatieve borgingsinspanning tot uitgifte. Scenario 1 betreft datatypes, waarvoor binnen de GDN-databank al een datamodel bestaat ('bekend datatype'). Via tekstblokken is zichtbaar, waar de scenario's 2, 3 en 4 in het schema passen.

In de uitwerking van de scenario's maken we zoals gezegd onderscheid tussen de fysieke en digitale data. Voor wat betreft de digitale data is scenario 1 van toepassing op de zogenaamde pre-2003 data, die vanaf de mijnbouwwetherizing van 2003 verplicht jaarlijks aangeleverd moeten worden door de exploitanten (zie paragraaf 3.1) en waarvan de data uit de periode daarvoor nog incompleet is. De overige digitale data maken onderdeel uit van scenario 2, 3 en 4. Voor de fysieke data vallen twee datasets onder scenario 1. De eerste dataset betreft de data verkregen bij het aanleggen van boorgaten (in het bijzonder kernmateriaal). De kosten daarvoor hangen vooral af van de door KGG te maken keuzes over uitbreiding van de bestaande opslagcapaciteit (één of meerdere opslaglocaties, wat te doen met de niet-unieke, minder kwalitatieve data en de rol van digitalisatie bij vrijgave, zie paragraaf 3.3). Daarom is het op dit moment weinig zinvol om voor deze dataset een nadere kostenberekening uit te voeren en laten we deze categorie buiten beschouwing in de kostenberekening (Bijlage D). De tweede dataset betreft de seismische velddata in de vorm van tapes (zie paragraaf 3.2). De overige fysieke data betreffen met name de biologische en chemische datatype (paragraaf 3.3). Omdat we weinig zicht hebben op deze, vermoedelijk behoorlijk diverse datasets, laten we ze buiten beschouwing in de kostenberekening (Bijlage D).

In de berekening onderscheiden we vier kostenposten (systeemontwikkeling, inname, beheer en vrijgave), die in wisselende mate van toepassing zijn op de scenario's. De kosten voor systeemontwikkeling en inname zijn eenmalig. De kosten voor beheer en vrijgave keren jaarlijks terug. Voor de jaarlijkse beheerkosten hebben we als vuistregel 10% van de ontwikkelkosten genomen. Verder voorziet geen van de scenario's in verdere opwerking van de data voor specifieke toepassingen, want we gaan ervanuit dat deze verdere opwerking projectmatig zal worden uitgevoerd door de (eind)gebruikers (zie paragraaf 3.2). Dit houdt ook in, dat we in de meeste scenario's geen vrijgavekosten meenemen. Vrijgave zal dan ofwel regulier plaatsvinden via de GDN databank of op projectbasis. Bij elk scenario baseren we de ontwikkelkosten op een (recent) analoog voorbeeld, waarmee de GDN ervaring heeft opgedaan (Bijlage D).

De kosten per scenario (Tabel 3.2) geven een duidelijk beeld op hoofdlijnen van de met borging gemoeide kosten voor zover deze in de scenario's zijn meegenomen. Binnen scenario 1 maken we onderscheid tussen borging inclusief (scenario 1a) en exclusief seismische velddata (scenario 1b) vanwege de hoge kosten, die daarmee verbonden zijn. De kostenschattingen per scenario lopen uiteen van enkele tot tientallen miljoenen voor de doorgerekende periodes. In de eerste fase zijn met name de ontwikkel- en innamekosten hoog. Gaandeweg kunnen de jaarlijkse beheerkosten een behoorlijke kostenpost worden, doordat er steeds meer data is geborgd. De berekende kosten (Tabel 3.2) betreffen een globale inschatting. Voor elk datatype/set moet te zijner tijd bepaald worden welke mate van integratie in de GDN databank en welke vorm van uitgifte gewenst is en welke consequenties dat heeft voor de doorlooptijd en financiering. Ook zal de absolute borgingsinspanning bij datasets met dezelfde borgingswens kunnen verschillen, daar de ene dataset meer complexiteit herbergt dan de ander. Dit niveau van detail valt buiten de reikwijdte van dit onderzoek (zie paragraaf 1.4).

Tabel 3.2: Totale kosten voor ontwikkeling, inname en beheer voor de vier verschillende scenario's. Bij scenario 1, waarbij de datasets al onderdeel zijn van de GDN databank, gaat het niet op datatypes maar voor specifieke datasets. Daarom staan daar alleen de totale kosten vermeld. Scenario 1a: Inname, opslag en vrijgave van data pre 2003 en seismische veld-data. Scenario 1b: Idem zonder seismische veld-data. Scenario 2, 3 en 4: Inname, opslag en vrijgave datatype respectievelijk zonder, met beperkte of met volledige integratie in de GDN databank. Voor bedragen onder de 1 MEuro vermelden we alleen de hoogste kostenschatting. De onzekerheid bedraagt naar schatting 40% op grond van vooral de onzekerheid in het aantal datasets per datatype (Bijlage D).

Scenario	Eenmalige kosten			Jaarlijkse kosten	
	Ontwikkeling (per datatype en/of exploitant)	Inname (per datatype en exploitant)	Totaal (alle datatypes en exploitant)	Beheer (per datatype en exploitant)	Totaal (alle datatypes en exploitanten)
1a			15 MEuro		1,3 MEuro/jaar
1b			0,4 MEuro		0,7 MEuro/jaar
2	0,03 MEuro	0,02 MEuro	2 – 4 MEuro	0,002 MEuro	0,2 – 0,4 MEuro/jaar
3	0,2 MEuro		20 – 40 MEuro	0,02 MEuro	2 – 4 MEuro/jaar
4	0,25 MEuro	0,2 MEuro	36 – 56 MEuro	0,02 MEuro	2 – 4 MEuro/jaar

Inzichten

De combinatie van scenario 1a met scenario 2 is het meest kostenefficiënt als het gaat om het minimaal borgen van de (in de kostenberekening meegenomen) data. Deze minimale vorm van borging is vooral interessant voor de gebruikersgroep, die vertrouwd is met AI/ML (zie paragraaf 3.3). Tegelijk laat het de mogelijkheid open om op een later tijdstip alsnog integratie in de GDN databank te bewerkstelligen om tegemoet te komen aan de behoeften van andere gebruikersgroepen.

Naarmate de data dieper in de GDN databank geïntegreerd wordt (scenario 3 en 4), lopen de kosten verder op (ook al vervallen de kosten voor de al eerder gedane stappen). Opvallend is, dat de kosten van scenario 3 en 4 redelijk overeenkomen, waardoor de netto maatschappelijke baten van scenario 4 (bij uitvoering van een volledige MKBA, zie paragraaf 1.5) waarschijnlijk gunstiger zullen zijn.

3.5 Handelingsperspectief

Aangezien elke exploitant zijn eigen investeringsbeleid hanteert (hoofdstuk 1) en een andere hoeveelheid data beheert (Tabel 0.3), is de tijdsplanning een belangrijk aspect om rekening mee te houden bij het borgen van de data. Als besloten wordt om de data gefaseerd per exploitant te borgen, dan nemen de kosten eveneens gefaseerd toe en wel op basis van de hoeveelheid data bij een specifieke exploitant. Zodoende zal de databorging niet in één keer plaatsvinden, waardoor keuzes over het tijdspad (prioritering) mede bepalend zullen zijn voor de jaarlijkse kosten. Om de prioritering te kunnen ondersteunen vertalen we de scenario's naar zes perspectieven (Tabel 3.3). Hierin nemen we de (verwachte) volgorde van vertrek van exploitanten uit Nederland mee (perspectief 3), omdat een abrupt vertrek van meerdere exploitanten tegelijk als reactie op de gasprijsontwikkeling een borgingsrisico vormt (paragraaf 1.2). Verder schatten we in, dat we met 20% van alle data zeker 80% van alle voor hergebruik belangrijke data kunnen borgen. Hiervoor baseren we ons op het hergebruikspotentieel per datatype (# must haves) per maatschappelijke subthema (Figuur 2.7), waarbij we de top 15 gebruiken. De overige perspectieven zijn gebaseerd op de organisatorische aspecten, die we eerder in dit hoofdstuk hebben onderkend. Het eerste

perspectief biedt de keuze om de seismische velddata wel (perspectief 1a) of niet (perspectief 1b) te betrekken. En het derde perspectief kent vier getrapte mogelijkheden:

- Perspectief 3a: Minimaal de pre-2003 datasets (conform scenario 1b);
- Perspectief 3b: Minimale borging van de overige datasets (conform scenario 2);
- Perspectief 3c: Combinatie van 3a en 3b (alle data van deze bedrijven minimaal geborgd, uitgezonderd de seismische velddata);
- Perspectief 3d: Alle data van deze bedrijven geborgd in de GDN databank.

De overige perspectieven zijn niet verder onderverdeeld.

Tabel 3.3: Overzicht van de zes perspectieven.

Perspectief	Omschrijving	Periode
1	completeren van het bestaande datamodel	aankomende 5 jaar
2	op minimale wijze borgen van de digitale data van alle huidige exploitanten	aankomende 5 jaar
3	borgen van alle data van de in deze periode vertrekkende exploitanten	aankomende 5 jaar
4	borgen van de top 15 in de GDN databank	aankomende 15 jaar
5	op minimale wijze borgen van de digitale data van alle huidige exploitanten, plus de top 15 geborgd in de GDN databank	aankomende 15 jaar
6	borgen van de data van alle huidige exploitanten in de GDN databank	aankomende 15 jaar

Om de kosten van de verschillende perspectieven direct met elkaar te kunnen vergelijken voeren we twee aanvullende bewerkingen uit (Bijlage E):

- De kosten voor de handelingsperspectieven met een tijdhorizon van 5 jaar zijn doorgerekend naar 15 jaar. Dit betreft dan voor de laatste 10 jaar alleen de beheerkosten;
- Er zijn gemiddelde, jaarlijkse kosten berekend over elke 5 jaar en in totaal. De kosten zijn een indicatie (en lopen op als functie van de tijd).

Voor een algehele afweging zijn naast de kosten ook de realisatie van de baten en het behoud van datasets bij de verschillende perspectieven van belang (Tabel 3.4). Gelet op de kosten vallen de perspectieven in een aantal clusters uiteen. Het eerste cluster bestaat uit de vijf perspectieven onderaan de lijst. Tegenover deze een goede indruk van wat er in de komende 5 jaar met een budget tot 10 MEuro bereikt kan worden, afhankelijk van de prioritering. Als de inspanningen zich beperken tot één of een combinatie van deze perspectieven, zullen daarbij gegevens voorgoed verloren gaan. Dit geldt in het bijzonder voor de seismische velddata op tapes.

Het tweede cluster omvat de vier perspectieven, waarvan de gemiddelde kostenschattingen schommelen nabij de 20 en 30 MEuro. Daarbij valt op, dat de kosten van perspectief 4 en 5, die allebei inzetten op de top 15 (zie paragraaf 2.4), redelijk overeenkomen. Perspectief 5 is al aantrekkelijk ten opzichte van perspectief 4 bij relatief lage baten door hergebruik van de digitale data, die buiten de top 15 vallen (vanaf 6 MEuro). Verder omvat dit cluster twee perspectieven (1a en 1c), die zijn gericht op het completeren van de datasets, die passen binnen de huidige datastromen, en tezamen dus slechts een klein deel van de totale data omvatten (zie Bijlage D).

Tabel 3.4: De 6 perspectieven gerangschikt naar afnemende totale kosten over een periode van 15 jaar met een inschatting van de realisatiekans van de baten (expertinschatting van de auteurs). Voor borging van alle data, die momenteel in beheer zijn bij de exploitanten, komen daar nog de kosten bij voor het resterende deel van de fysieke data (kernhuis, inclusief biologische en chemische data, zie paragraaf 3.4). De onzekerheid in de absolute bedragen is aanzienlijk (tenminste 40%, zie Bijlage E). Doordat de onderliggende aannamen bij de perspectieven veelal dezelfde zijn, is de kostenverdeling over de perspectieven indicatief voor de onderlinge kostenverhoudingen.

Perspectief	Omschrijving	Gemiddelde schatting van de totale kosten (MEuro)	Realisatiekans baten van hergebruik	Behoud van alle in de scenario's meegenomen datasets?
6	Digitale datasets in 15 jaar geborgd in GDN databank	95	Optimaal voor de betreffende datasets	Ja, behalve pre-2003 data en seismische velddata
3d	Borging van de datasets in GDN databank van vertrekkende exploitanten, in de komende 5 jaar	62	Optimaal voor de betreffende datasets	Ja, maar niet de data van de overige exploitanten
1a	Borgen van pre-2003 en seismische velddata in 5 jaar	31	Optimaal voor de betreffende datasets	Nee, alle digitale data ontbreken
5	In 15 jaar top 15 borgen in GDN databank, overige digitale datasets minimaal geborgd	25	Uitstekend	Ja, behalve pre-2003 data en seismische velddata
1c	Borgen van seismische velddata in 5 jaar	22	Optimaal voor deze ene dataset	Nee, er wordt slechts 1 dataset geborgd
4	In 15 jaar top 15 borgen in GDN databank	19	Uitstekend	Nee, naast pre-2003 data en seismische velddata ontbreekt ook rest digitale data
1b	Borgen pre-2003 data in 5 jaar	9	Beperkt tot redelijk voor deze ene dataset	Nee, er wordt slechts 1 dataset geborgd
2	Minimaal borgen alle digitale datasets in 5 jaar	7	Optimaal voor deze ene dataset	Ja, behalve pre-2003 data en seismische velddata
3c	Minimale borging datasets vertrekkende exploitanten in komende 5 jaar	7	Beperkt	Nee, seismische velddata noch data van overige exploitanten
3a	Minimaal pre-2003 datasets borgen vertrekkende exploitanten in komende 5 jaar	3	Optimaal voor deze ene dataset	Nee, er wordt slechts 1 dataset en niet voor rest exploitanten
3b	Minimaal borgen datasets vertrekkende exploitanten in komende 5 jaar	3	Beperkt	Nee, naast de pre-2003 data en seismische velddata ontbreekt data van overige exploitanten

In het derde cluster vallen eenvoudigweg de twee perspectieven, waar de hoogste kosten (en baten) aan verbonden zijn (65 respectievelijk 95 MEuro). Als de inspanningen zich beperken tot één van deze twee perspectieven, zullen ook daarbij een aanzienlijk deel van de gegevens voorgoed verloren gaan; ofwel de fysieke data, of de data van de overige operators.

De meest kostenefficiënte combinatie voor minimaal borgen van alle in de kostenberekening meegenomen data (perspectief 1a met perspectief 2) kost naar schatting 38 MEuro over een looptijd van 5 jaar. Deze kosten kunnen desgewenst (deels) worden uitgesmeerd over een langere periode (15 jaar). Ter vergelijking, de kosten voor volledige integratie in de GDN databank van dezelfde data (perspectief 1a met perspectief 6) nemen 15 jaar in beslag en lopen op naar $(95+31=)$ 126 MEuro. Als het uitgangspunt is om alle data te borgen, dan komen de kosten voor het resterende deel van de fysieke data (kernhuis, inclusief biologische en chemische data) er nog bij (zie paragraaf 3.4). Dit verschil onderstreept het belang van het meenemen van de maatschappelijke baten in de afweging.

Vertaald naar de jaarlijkse kosten ziet het borgen van alle in de kostenberekening meegenomen data er budgettair ongeveer als volgt uit:

- alle digitale data in de GDN databank kost gemiddeld ruim 6 MEuro/jaar over de aankomende 15 jaar;
- de top 15 in de GDN databank kost gemiddeld krap 2 MEuro/jaar over de aankomende 15 jaar;
- alle digitale data op minimale wijze innemen (en ze alleen op projectmatige basis vrijgeven) kost circa 0,9 MEuro/jaar in de aankomende 5 jaar en daarna ongeveer 0,3 MEuro/jaar;
- het innemen van de pre-2003 data kost krap 2 MEuro/jaar in de aankomende 5 jaar;
- het innemen van zowel de pre-2003 data als de seismische velddata kost ruim 6 MEuro/jaar in de aankomende 5 jaar.

Gedacht vanuit een budget met een jaarlijks plafond van ongeveer 500 k/jaar voor de aankomende 5 jaar is één van de volgende opties haalbaar:

- inname van de pre-2003 data;
- alle digitale data op minimale wijze innemen kan alleen wanneer de kosten aan de lage kant zitten van de huidige inschatting;
- de digitale data van alleen de eerst vertrekkende exploitanten op minimale wijze innemen kan als de kosten aan de gemiddelde/hoge kant zitten van de inschatting.

Gedeeltelijke combinaties van bovenstaande opties kunnen, zolang de kosten aan de lage kant zitten van de huidige inschatting. Bij al deze opties is het essentieel om aanvullend budget te regelen voor na de eerste 5 jaar, bijvoorbeeld om ook de data van de andere exploitanten te kunnen borgen. En daar, waar vrijgave alleen op projectbasis plaats zal vinden, is de realisatiekansen van de baten door hergebruik beperkt (Tabel 3.4).

Inzichten

Voor de prioritering kan worden gekozen tussen zes uitgewerkte perspectieven of combinaties daarvan. Deze berusten op drie overwegingen:

- organisatorische aspecten van borging;
- inspelen op het vertrek van specifieke exploitanten in de aankomende vijf jaar;
- het borgen van de top 15 in de GDN databank conform paragraaf 2.4.

Wij hebben de top 15 afgeleid uit de enquêteresultaten. Desgewenst kan de Staat een eigen top 15 samenstellen. Dit laatste kan bijvoorbeeld door bepaalde maatschappelijke subthema's (Bijlage A) zwaarder te wegen, bepaalde gebruikerswensen te honoreren of bepaalde GDN informatieproducten voorrang te geven (omdat deze producten qua hergebruik breed inzetbaar zijn en amnesie tegengaan).

De perspectieven lopen uiteen voor wat betreft de kosten (3 tot 95 MEuro per perspectief) en de hoeveelheid bestaande gegevens en interpretaties, die verloren zal gaan. Het op minimale wijze borgen van zoveel mogelijk gegevens en interpretaties vereist een combinatie met minimaal twee perspectieven. Dit kost naar schatting krap 40 MEuro, waarvan circa 22 MEuro voor de seismische velddata, en daar komen de kosten voor uitbreiding van de opslagcapaciteit (kernhuis, inclusief de fysieke biologische en chemische data) nog bij. Ter vergelijking, de kosten voor volledige integratie in de GDN databank van dezelfde data - wat de realisatiekansen van hergebruik optimaliseert - lopen op naar 126 MEuro. De onzekerheid in de genoemde bedragen bedraagt naar schatting 40% op grond van vooral de onzekerheid in het aantal datasets per datatype.

Bovenstaande getallen zijn onzeker en berusten op meerdere aannames; door te starten met de inname van data kunnen gaandeweg de belangrijkste aannames aangaande de kosten worden getoetst en daarmee ook de bijbehorende onzekerheden worden verminderd. De seismische velddata vormen vanwege het arbeidsintensieve innameproces een relatief grote kostenpost. Daarom loont het de moeite om te investeren in het optimaliseren van dit proces; zelfs kleine verbeteringen in efficiëntie zullen resulteren in grote kostenbesparingen.

4 Analoge werkwijze in vergelijkbare landen

Om het handelingsperspectief (hoofdstuk 3) te kunnen plaatsen in de internationale context en mogelijk uit te breiden hebben wij de situatie qua mijnbouwdata in een aantal landen op hoofdlijnen geanalyseerd. Hiervoor hebben we vier landen onder de loep genomen: het Verenigd Koninkrijk, Noorwegen, Duitsland en Australië. Alle landen hebben een min of meer actieve mijnbouwsector in koolwaterstoffen, al zijn er grote verschillen in de ontginningsfase. Ook hebben zij allen een helder en uitgewerkt raamwerk om relevante data te borgen en vrij te geven.

Per land hebben we een situatieschets opgesteld op basis van analyse van gepubliceerde documenten juridische richtlijnen en werkdocumenten en interviews met de instanties, die verantwoordelijk zijn voor de borging van mijnbouwdata (Bijlage F). In de analyse hebben we ons gericht op:

- alternatieve filosofieën voor databorging;
- hergebruikspotentieel;
- datatypen en vertrouwelijkheid;
- andere inzichten aangaande regelgeving en richtlijnen.

De gespreksresultaten vatten we eerst samen (paragraaf 4.1) om ze daarna te duiden in het kader van onze vraagstelling (paragraaf 4.2).

4.1 Gespreksresultaten

De resultaten vatten we hieronder samen.

Databeleid

Niet alle gesprekspartners waren in detail op de hoogte van de redenen om een bepaalde databorgingmethodiek op te zetten aangezien de partijen vaak bij uitvoerende instanties werken. In veel gevallen wordt het beleid bij een andere instanties opgesteld en uitgewerkt. De landen, waar een heldere visie is omschreven, gaven aan, dat een goede borging en vrijgave van mijnbouwdata/interpretaties van grote waarde is om het potentieel van de ondergrond optimaal te ontsluiten.

Veel gesprekspartners gaven aan een clause te hebben om te zorgen, dat er zo min mogelijk gegevens verloren te gaan. Bijvoorbeeld door als overheid een optie in te regelen, waarbij de (daartoe aangewezen) databorgende partij het recht krijgt om alle relevante gegevens op te vragen als eis bij teruggaaf van de vergunning (in de vorm van een formeel overdrachtsrapport of middels een formeel overdrachtmoment). In Duitsland is elke vergunninghouder verplicht om bij de databorgende partij toestemming te vragen indien zij data willen vernietigen. Het Verenigd Koninkrijk hanteert een verplichting om een contactpersoon in te stellen voor datazaken bij vergunninghouder/uitvoerder.

Hergebruikspotentieel en vervangingswaarde

Alle gesprekspartners zien de potentie voor hergebruik van de bestaande gegevens en zien dit ook in de praktijk gebeuren, het meest concreet voor nieuwe koolwaterstofprojecten, geothermie en CO₂-opslag. De hergebruikswaarde is volgens de gesprekspartners lastig

concreet te maken. Daarbij gaven ze aan, dat er geen twijfel bestaat dat de vervangingswaarde van de gegevens zeer groot is. Enkele landen hebben recent formele consultatierondes en workshops gehouden om belanghebbenden de kans te geven hun mening te geven over hergebruik van gegevens voor nieuwe toepassingen van de ondergrond.

Een punt van zorg bij verschillende partijen is de toekomstige leesbaarheid van specifieke digitale formats. Wijzelf schatten dit specifieke risico in als beperkt (met enige tijdsinvestering blijven conversies mogelijk, ook als de software niet meer ondersteund wordt). Tegelijk delen we de zorg voor amnesie (zie paragraaf 2.4). Ook zien wij landen met een brede databorging beleid, waar dus veel gegevens worden ingenomen, worstelen met de hoeveelheid aan types gegevens. De vindbaarheid van de gegevens is daar vaak de beperkende factor.

Datatypes en confidentialiteitstermijnen

De focus ligt zonder uitzondering op seismische data en boorgegevens. Enkele landen hebben wetgeving en regelingen ingevoerd, die gelden voor geïnterpreteerde gegevens en informatie (zoals seismische interpretaties, reservoirmodellen etc.), omdat deze informatie een grote hoeveelheid voor hergebruik en nazorg relevante kennis bevat. Dit laatste is een groot verschil met de Nederlandse situatie (in vergelijking met andere landen richt Nederland zich sterk op het innemen van meetgegevens) en kan essentieel zijn voor hergebruik.

Een overzicht van de datatypes en confidentialiteitstermijnen per land (Tabel 4.1) laat zien, dat verschillende landen andere termijnen hanteren; deze zijn zowel langer als korter dan de termijnen in Nederland. De confidentialiteitsduur is in meerdere landen recent aangepast.

Tabel 4.1: Termijnen waar binnen data als vertrouwelijk behandeld moeten worden in vijf verschillende landen.

Type	Vertrouwelijkheidstermijn				
	NL	DE	UK	NO (recent verkort)	AUS
Metadata	Meteen	3 maanden	Meteen	?	?
Productie data	3 maanden	Niet van toepassing	2 maanden	?	Niet van toepassing
Putgegevens (logs, kernen etc.)	5 jaar	10 jaar	2 jaar	2 jaar	2 jaar
Reguliere seismiek	5 jaar	10 jaar	5/10/15 jaar (verschillende categorieën)	10 jaar	3 jaar
Multiclient seismiek	10 jaar	10 jaar	Ongespecificeerd	10 jaar	15 jaar
Geïnterpreteerde data en informatie	Niet van toepassing	Oneindig	Zelden gebruikt, niet geformaliseerd	5 jaar	4 jaar putgegevens, 5 jaar seismiek

4.2 Inzichten

De gespreksresultaten bevestigen (hoofdstuk 2), dat:

- alle datatypen moeilijk reproduceerbaar zijn en dat als opnieuw verzamelen wel kan, dan tegen hoge kosten;
- de zorg over het verlies van de benodigde kennis en vaardigheden voor hergebruik in de periode tot 2050 door andere landen gedeeld wordt, zij het soms op andere punten;
- het hergebruikspotentieel aanzienlijk is, ook relatief ten opzichte van de verwachte kosten.

Net als andere landen is het in Nederland gehanteerde uitgangspunt op het gebied van mijnbouwgegevens en interpretaties (dat de exploitanten deze gegevens gedurende de operationele vergunningsfase zelf beheren en ze op aanvraag overdragen aan de Staat, zie paragraaf 1.1) in principe erop gericht om te zorgen, dat er zo min mogelijk gegevens verloren te gaan. Daarbij valt op, dat Nederland hiervoor geen harde clause heeft ingebouwd, waar dat elders gangbaar is.

Een veelgebruikte clause geeft een daartoe aangewezen databorgende partij het recht om alle relevante gegevens op te vragen als eis bij teruggaaf van de vergunning (in de vorm van een formeel overdrachtsrapport of middels een formeel overdrachtsmoment).

5 Bevindingen en aanbevelingen

Het vraagstuk van een sector brede overdracht van gegevens vormt om meerdere redenen een uitdaging (paragraaf 1.3). Daarom hebben we in dit onderzoek gekeken naar de baten, kosten en organisatorische aspecten (paragraaf 2.4), hebben we het handelingsperspectief in kaart gebracht (paragraaf 3.5), waarbij we voor zes perspectieven de kosten hebben uitgewerkt (Tabel 3.4) en zijn nagegaan hoe dit vraagstuk in een aantal andere landen wordt opgepakt (hoofdstuk 4). Op grond van de opgedane inzichten (samengevat in paragraaf 5.1) komen we tot twee hoofdaanbevelingen (paragraaf 5.2 en 5.3) en definiëren we een aantal pilots (paragraaf 5.4).

5.1 Inzichten

Voor wat betreft de baten, kosten en organisatorische aspecten zijn we tot de volgende bevindingen gekomen (hoofdstuk 2 en hoofdstuk 4):

- Het hergebruikspotentieel van alle datatypen is aanzienlijk (ook relatief ten opzichte van de verwachte kosten) en verschilt per maatschappelijk subthema;
- Voor alle datatypen geldt, dat ze moeilijk reproduceerbaar zijn. En als opnieuw verzamelen wel kan, dan tegen zeer hoge kosten. Tegelijkertijd is verlies van de benodigde kennis en vaardigheden voor hergebruik in de periode tot 2050 reëel aangezien de opmars naar dit verlies reeds is geobserveerd in de laatste jaren;
- Naar schatting beheren de in Nederland werkzame olie- en gasexploitanten momenteel tussen de 30 tot 60% van de totale hoeveelheid unieke historische en actuele mijnbouwgegevens en de daarvan afgeleide interpretaties. Het borgen en afdoende ontsluiten voor hergebruik daarvan vraagt een (aanzienlijk) grotere inspanning dan wat nu nodig is voor de huidige data, die GDN beheert.

Daarom is het essentieel om zoveel mogelijk gegevens en interpretaties te borgen, stapsgewijs en planmatig.

Voor een stapsgewijze en planmatige borging zijn een aantal organisatorische aspecten essentieel (hoofdstuk 3):

- De wijze van inname, beheer en vrijgave van de fysieke data hangt af van door KGG te maken keuzes aangaande één of meerdere opslaglocaties, wat te doen met de niet-unieke, minder kwalitatieve data en de rol van digitalisatie bij vrijgave. Deze keuzes zijn relevant voor het tijdig kunnen organiseren van de benodigde opslagcapaciteit;
- Het innemen, controleren en opslaan van de seismische velddata tapes is een arbeidsintensief proces, dat niet gemakkelijk is om op te schalen (zeker niet onder tijdsdruk). Dit vormt een risico bij abrupt vertrek van (meerdere) exploitanten;
- Data, die worden ingenomen via een online innameportaal, zijn beter gestructureerd, tijdiger en eenvoudiger te controleren op kwaliteit. Daarom heeft deze wijze van inname indien mogelijk onze voorkeur;
- Duidelijkheid vooraf over de betrouwbaarheid van nieuwe datatypen en de termijnen daarvoor is een vereiste voor uitgifte in het publieke domein;

- Naast het opschalen van de huidige werkprocessen is het van belang om innovatieve werkprocessen en andere modi van uitgifte te onderzoeken, zodat we kunnen aansluiten bij de opkomende gebruikersgroepen. Daarbij denken we bijvoorbeeld aan het rechtstreeks aanbieden van bestaande, ongestructureerde data aan ervaren AI/ML gebruikers.

Voor de overdracht van zowel de fysieke als de digitale datasets is medewerking van de exploitant van belang, vooral voor de datatypes, waar de GDN tot op heden minder ervaring mee heeft opgebouwd (bepaalde datasets binnen de operationele en chemische datatypes). Dit geldt in het bijzonder voor het veilig stellen van de metagegevens (waaronder de herkomst vanwege de vele datatransfers gerelateerd aan fusies en overnames van exploitanten in het verleden) en de kwaliteitscontrole (vaststellen kwaliteitslabel voor hergebruik).

Het handelingsperspectief bestaat uit zes uitgewerkte perspectieven of combinaties daarvan, die berusten op drie overwegingen:

- organisatorische aspecten van borging;
- inspelen op het vertrek van specifieke exploitanten in de aankomende vijf jaar;
- het borgen van de top 15 in de GDN databank conform paragraaf 2.4.

Wij hebben een top 15 afgeleid uit de enquêteresultaten. Desgewenst kan de Staat een eigen top 15 samenstellen. Dit laatste kan bijvoorbeeld door bepaalde maatschappelijke subthema's (Bijlage A) zwaarder te wegen, bepaalde gebruikerswensen te honoreren of bepaalde GDN informatieproducten voorrang te geven (omdat deze producten qua hergebruik breed inzetbaar zijn en amnesie tegengaan).

De perspectieven lopen uiteen voor wat betreft de kosten (3 tot 95 MEuro per perspectief) en de hoeveelheid bestaande gegevens en interpretaties, die verloren zal gaan. Het op minimale wijze borgen van zoveel mogelijk gegevens en interpretaties vereist een combinatie met minimaal twee perspectieven. In de meest kostenefficiënte combinatie kost dit naar schatting krap 40 MEuro, waarvan circa 22 MEuro voor de seismische velddata, en daar komen de kosten voor uitbreiding van de opslagcapaciteit (kernhuis, inclusief de fysieke biologische en chemische data) nog bij. Ter vergelijking, de kosten voor volledige integratie in de GDN databank van dezelfde data - wat de realisatiekansen van hergebruik optimaliseert - lopen op naar 126 MEuro. De onzekerheid in de genoemde bedragen bedraagt naar schatting 40% op grond van vooral de onzekerheid in het aantal datasets per datatype.

Bovenstaande getallen zijn onzeker en berusten op meerdere aannames; door te starten met de inname van data kunnen gaandeweg de belangrijkste aannames aangaande de kosten worden getoetst en daarmee ook de bijbehorende onzekerheden worden verminderd. De seismische velddata vormen vanwege het arbeidsintensieve innameproces een relatief grote kostenpost. Daarom loont het de moeite om te investeren in het optimaliseren van dit proces; zelfs kleine verbeteringen in efficiëntie zullen resulteren in grote kostenbesparingen.

Op deze kostenschattingen zijn ook andere nuancerings van toepassing (hoofdstuk 3). Zo zal de databorging niet in één keer plaatsvinden, waardoor keuzes over het tijdspad (prioritering) mede bepalend zullen zijn voor de jaarlijkse kosten. En voor elk datatype/set moet te zijner tijd bepaald worden welke mate van integratie in de GDN databank en welke vorm van uitgifte gewenst is en welke consequenties dat heeft voor de doorlooptijd en financiering. Ook zal de absolute borgingsinspanning bij datasets met dezelfde borgingswensen kunnen verschillen, daar de ene dataset meer complexiteit herbergt dan de ander. Dit niveau van detail valt buiten de reikwijdte van dit onderzoek. Ook voorzien de kostenschattingen niet in verdere opwerking van de data naar aangepaste of nieuwe informatieproducten voor specifieke toepassingen, want we gaan ervanuit dat dit projectmatig zal worden uitgevoerd door de (eind)gebruikers.

Vrijgave voor hergebruik vindt dus ofwel regulier plaats via de GDN databank of op projectbasis. We merken daarbij op, dat opwerking het verloren gaan van kennis tegen gaat en het beoogde hergebruik stimuleert.

De GDN is aangesloten bij diverse standaarden voor data en dataverwerking, die onder meer van ons vragen dat we streven naar behoud van originele data (hoofdstuk 3). De Staat heeft momenteel geen harde clause ingebouwd in de wet om te zorgen, dat er zo min mogelijk gegevens verloren gaan, waar dit elders gangbaar is (hoofdstuk 4). Een door andere landen veelgebruikte clause geeft een daartoe aangewezen databorgende partij het recht om alle relevante gegevens op te vragen als eis bij teruggaaf van de vergunning (in de vorm van een formeel overdrachtsrapport/overdrachtsmoment). Dat zo'n clause niet overdreven is, blijkt uit het feit dat op dit moment naar schatting 8 tot 15% van de unieke gegevens (afhankelijk van het datatype) in de categorie missende of ontbrekende data valt (Figuur 2.5).

5.2 Voorkom dat mijnbouwgegevens verloren gaan

Op grond van de opgedane inzichten (paragraaf 5.1) is de eerste hoofdaanbeveling te voorkomen dat mijnbouwgegevens verloren gaan. Hiervoor bevelen we aan om:

- een daartoe aangewezen databorgende partij het recht te geven om alle relevante gegevens op te vragen als eis bij teruggaaf van de vergunning en te voorzien in een proportionele sanctie als de exploitant hieraan niet voldoet. Zo kan formeel worden vastgesteld, dat de exploitant aan al zijn verplichtingen op het vlak van dataoverdracht heeft voldaan;
- exploitanten te verplichten om inventarislijst te maken van de digitale en fysieke gegevens en bijbehorende interpretaties, die zij momenteel beheren, inclusief missende en ontbrekende gegevens (bijvoorbeeld als gevolg van overnames in het verleden) en deze lijst periodiek te actualiseren;
- voldoende middelen vrij te maken, zodat deze gegevens en interpretaties daadwerkelijk op minimale wijze kunnen worden geborgd (25-60 MEuro in de komende 5 tot 15 jaar), en eventuele opties voor co-sponsoring door potentiële gebruikers te benutten;
- binnen afzienbare termijn keuzes te maken over uitbreiding van de opslagcapaciteit voor fysieke data (aangaande één of meerdere opslaglocaties, wat te doen met de niet-unieke, minder kwalitatieve data en de rol van digitalisatie bij vrijgave), zodat de benodigde opslagcapaciteit tijdig kan worden georganiseerd;
- afspraken te maken met de sector over getemporeerde overdracht van de (fysieke) gegevens en interpretaties;
- maatschappelijke bewaarcriteria op stellen. Tegenover de verplichting tot het bewaren van originele data staan andere overwegingen. Gaan we deze data voor eeuwig bewaren (met oneindig oplopende beheerkosten plus de gerelateerde klimaat- en milieubelasting), tot ze verkeren in een vooraf bepaalde staat van degradatie of onbruik, of is er sprake van natuurlijke bewaartermijnen (1, 2 of 3 generaties)?

Door te starten met de inname van data kunnen gaandeweg de belangrijkste aannames aangaande de kosten worden getoetst en daarmee ook de bijbehorende onzekerheden worden verminderd (zie paragraaf 5.4).

5.3 Stimuleer hergebruik

Op grond van de opgedane inzichten (paragraaf 5.1) luidt de tweede hoofdaanbeveling hergebruik van de data te stimuleren. Hiervoor bevelen we aan om:

- hergebruik, opwerking en uitgifte in het publieke domein van de op minimale wijze geborgde data op projectbasis organisatorisch en financieel aantrekkelijk te maken (bijvoorbeeld via de Topsectoren innovatiehelix⁸);
- rekening te houden met de zorgen van exploitanten over eventuele juridische aansprakelijkheid bij hergebruik van hun interpretaties door derden. De NAM (persoonlijke communicatie⁹) stelt in dit kader ook voor om de communicatie over door de Staat vrijgegeven interpretaties van exploitanten te laten verlopen via een daartoe aangewezen partij (niet de exploitanten zelf, branchevereniging ElementNL zou wel kunnen);
- regelmatig in gesprek te gaan met specifieke (potentiële) gebruikersgroepen, zoals uitvoeringsinstanties, toezichthouders, de decentrale overheden en het MKB, over hun gebruikerswensen en welke drempels zij in de praktijk ervaren;
- om innovatieve modi van uitgifte te onderzoeken, die (beter) aansluiten bij opkomende gebruikersgroepen, zoals de AI/ML experts (zie paragraaf 5.4);
- bij de overdracht waar mogelijk voorrang te geven aan de top 15. Desgewenst kan de Staat een eigen top 15 samenstellen. Dit laatste kan bijvoorbeeld door bepaalde maatschappelijke subthema's (Bijlage A) zwaarder te wegen, bepaalde gebruikerswensen te honoreren of bepaalde GDN informatieproducten voorrang te geven (omdat deze producten qua hergebruik breed inzetbaar zijn en amnesie tegengaan).

Mocht het hergebruik op projectbasis of via innovatieve uitgiftemodi stagneren, dan is het mogelijk om alsnog over te schakelen naar het stimuleren van hergebruik door gedeeltelijke of volledige integratie in de GDN databank (althans voor zover de data minimaal geborgd is, zie paragraaf 5.2).

5.4 Pilots

We stellen voor om een aantal pilots op te starten¹⁰. Daarmee kunnen we naast het maken van betere inschattingen van de kosten (paragraaf 5.2) gericht ervaring opdoen met alle aspecten van inname, beheer en vrijgave (paragraaf 5.3) in afstemming met de exploitanten en de betrokken overheden en uitvoeringsinstanties.

Concreet hebben de voorgestelde pilots tot doel om met beperkte middelen en in korte tijd:

- te toetsen in hoeverre de huidige raming van de kosten, de hoeveelheid gegevens en informatie, de borgingsinspanning en andere inhoudelijke bevindingen (hoofdstukken 2 en 3) aansluiten bij de realiteit;
- te komen tot een kosteneffectieve werkwijze, die geschikt is voor opschaling naar andere datatypes of exploitanten;
- voor één of meer maatschappelijke subthema's te toetsen welk hergebruikspotentieel (Figuur 2.7) de buitenwereld de verschillende datatypes toedicht;
- criteria op te stellen om het gewenste minimale opwerkingsniveau per datatype te bepalen rekening houdend met de gebruikerswensen (voor zover nu bekend);

⁸ www.rvo.nl

⁹ De NAM ziet bij hergebruik ook een meer actieve rol weggelegd voor EBN of TNO en denkt daarbij concreet aan het inzien van de vrijgegeven rapporten, het verstrekken of verder verwerken van de interpretaties zonder juridische consequenties en het verzorgen van de communicatie hierover.

¹⁰ Inmiddels heeft KGG voor de aankomende 5 jaar ongeveer 2,5 MEuro beschikbaar gesteld (0,5 MEuro per jaar binnen het zgn. PIO).

- in beeld te brengen waar eventuele synergie kan worden behaald tussen de verschillende datatypes, of binnen de GDN databank;
- te inventariseren welke opties er zijn voor co-sponsoring.

Daarbij willen we minimaal een verbeterd inzicht krijgen in de werkelijke kosten van:

1. perspectief 1: het innemen van de pre-2003 data (met/zonder de seismische velddata);
2. perspectief 2: alleen inname en minimale borging;
3. perspectief 4: borging van alleen de top 15 data in de GDN databank;

en in de mogelijkheden om afspraken te maken tussen de GDN en het ministerie, evenals tussen de GDN en de exploitanten, aangaande het op gecontroleerde wijze aanleveren van data¹¹ (bij vertrek).

In afwachting van de door KGG te maken keuzes over uitbreiding van de opslagcapaciteit voor fysieke data (paragraaf 5.2) hebben we geen pilots opgenomen over de fysieke data (anders dan pilot D over de seismische datatypes).

Pilot A De digitale data van de NAM

Dit betreft de versnelde overdracht van meerdere (niet alle), overwegend digitale datasets. In deze pilot willen we expliciet kijken naar:

- Kosten volgens perspectief 1;
- Kosten volgens perspectief 2;
- Toetsing op inschatting van hoeveelheid data t.o.v. het gemiddelde;
- Gesprek over afspraken, die gemaakt kunnen worden over overdracht van data en in welke mate deze breder te gebruiken zijn.

Pilot B Een exploitant met relatief weinig digitale data

Bij pilot A gaat het om een grote hoeveelheid data, waarbij inname en beheer meteen een grote kostenpost zal zijn. Daarom stellen we voor om de andere aannames verder te toetsen bij een exploitant met een relatief kleine hoeveelheid data (Spirit, Jetex, Dana, Kistos, Petrogas of TAQA). Daarbij lijkt het ons raadzaam om te starten met een exploitant met een hoge vertrekans op relatief korte termijn. Bij deze pilot stellen we voor om de volgende aspecten te toetsen:

- Kosten volgens perspectief 4;
- Toetsing op inschatting van hoeveelheid data t.o.v. het gemiddelde;
- Gesprek over afspraken, die gemaakt kunnen worden over overdracht van data en in welke mate deze breder te gebruiken zijn.

Pilot C Verkenning van een voor GDN relatief onbekende dataset

We stellen voor te beginnen met de dataset locatie-onderzoeksgegevens, die worden ingewonnen voorafgaand aan de installatie van platforms en andere faciliteiten. Voor deze dataset willen we de hoeveelheid en complexiteit van de data vaststellen en een innameproces opzetten. In een later stadium kan een vergelijkbare aanpak worden gevolgd voor enkele specifieke datasets van het chemische of geïnterpreteerde datatype. Hiermee kan op beperkte schaal:

1. worden getoetst welk hergebruikspotentieel de buitenwereld deze dataset toedicht (concreet: in relatie tot de aanleg van windparken op zee, zie Bijlage C);
2. criteria worden opgesteld om het gewenste minimale opwerkingsniveau te bepalen rekening houdend met de gebruikerswensen (voor zover nu bekend);

¹¹ Met name de inname van grote datasets, zoals seismische herbewerkingen, is gebaat bij een gestructureerd protocol. Dit wordt ook onderkend door de gassector, zie de OSDU standaard voor data uitwisseling.

Pilot D Slimme oplossingen voor de seismische datatapes

De crux van deze pilot is het terugdringen van het handmatige karakter van het innameproces. We stellen voor om de creativiteit van experts te prikkelen door deze uitdaging te behandelen als een wedstrijd, of door het met de juiste innovatieprikkels in de markt te zetten (waarbij het bedrijf, dat binnen de gestelde kwaliteitseisen de goedkoopste oplossing ontwikkelt, wordt beloond).

Pilot E Informatieproducten: stel de referentiemonsters veilig

De fysieke referentiemonsters zijn essentieel om het stratigrafische raamwerk, dat de basis vormt van de geologie in Nederland, wetenschappelijk te borgen (respondentenvoorstel, zie Bijlage C). Daarvoor is noodzakelijk, dat we in beeld hebben welke monsters het betreft en bij wie ze momenteel in beheer zijn. Deze pilot betreft dus een combinatie van een inventarisatie en het daadwerkelijk inwinnen van (een beperkte hoeveelheid) data bij meerdere exploitanten.

Pilot F Stel de gegevens over de initiële toestand van de ondergrond veilig

Deze pilot betreft het verzamelen en veiligstellen van de gegevens over de initiële toestand van de ondergrond (direct na aanboren), omdat dit de randvoorwaarden in numerieke en analoge modelstudies bepaalt (respondentenvoorstel, zie Bijlage C). Ook deze pilot betreft een combinatie van een inventarisatie en het daadwerkelijk inwinnen van (een beperkte hoeveelheid) data bij meerdere exploitanten.

Pilot G Alternatieve modi van uitgifte

Bij deze pilot denken we aan het ontwikkelen van een prototype om de meerwaarde en haalbaarheid te onderzoeken van getrapte integratieniveaus in de GDN-databank als tussenvorm voor enerzijds de reguliere integratie in de GDN-databank en anderzijds de optie ‘beschikbaar op aanvraag’. (Deze laatste optie werkt voor potentiële gebruikers vaak drempelverhogend, wat de baten van hergebruik in negatieve zin kan beïnvloeden. Daarom willen we deze optie liever niet gebruiken.) Eén zo’n traptree betreft het uitgeven van bestaande, ongestructureerde data door middel van (semi-)automatische categorisering. Een diepgaander onderzoek naar de gebruikerservaring kan een logisch vervolg zijn.

Pilot H Inventarisatie van relevante inzichten uit lopende projecten

Hergebruik is niet nieuw en daarom kunnen we leren uit lopende projecten bij GDN en elders. Een goed voorbeeld is de omgang met de privacywetgeving bij hergebruik van schadegegevens, dat aan de orde is bij het lopende project Nazorg steenkolenwinning Limburg (zie paragraaf 3.2). Andere voorbeelden zijn:

1. het ontwikkelen van een kwaliteitslabel bij geothermie om vervuiling door verouderde interpretaties tegen te gaan;
2. het digitaliseren van microscooppreparaten bij de NPD in Noorwegen;
3. de vloeistofmonsters, die zijn ingewonnen in het SCAN project.

In deze pilot wordt beoordeeld in hoeverre dergelijke ervaringen (voor zover bij ons bekend) van meerwaarde kunnen zijn voor het voorliggende vraagstuk.

Pilot I Evaluatie van opties om de lange termijn overdracht van data te realiseren

In deze pilot willen we een evaluatie uitvoeren naar de opties om gecontroleerde wijze van aanleveren van data te realiseren. Dit kan bijvoorbeeld door een op een afspraken met de mijnbouwmaatschappijen, een convenant of wettelijke regelgeving.

Referenties

Adrichem, W. van, Baat, P. de, Duijn, R. van, Elfrinkhof, S. van, Maas, M., Scheenaart, N., Struijk, B., Thoen, H., Wiezer, H., Programma van eisen herontwikkeling kernhuis Zeist, TNO rapport (intern document), 21 maart 2023, projectnummer RHDHV: BI2854, uitvoering: Royal HaskoningDHV

A.G. Akinyemi, M. Sun, A.J.G. Gray, Data integration for offshore decommissioning waste management, Automation in Construction, Volume 109, 2020, 103010, ISSN 0926-5805, <https://doi.org/10.1016/j.autcon.2019.103010>.
(<https://www.sciencedirect.com/science/article/pii/S0926580518304059>)

Beardsmore, G., Ali, A., Brommer, M., Chen, H., Conti, P., Falcone, G., Fridriksson, T., Hogarth, R., Mijnlief, H., Rumberg, G., Usshe, G., UNFC Resource Classification Applied to Real World Geothermal Projects: Outcomes from an Education and Implementation Pilot Program, Proceedings World Geothermal Congress 2020+1 Reykjavik, Iceland, April - October 2021, 9 p.
<https://www.geothermal-energy.org/pdf/IGAstandard/WGC/2020/03035.pdf>

Chapela, J., Data agility: Innovative approaches to subsurface data management, First Break, volume 41, issue 12, 2023, p. 69-72

DeMarco, K., Streamlining energy and production data management from field to processing, First Break, volume 41, issue 12, 2023, p. 45-55

Eszter Békési, Maartje Struijk, Damien Bonté, Hans Veldkamp, Jon Limberger, Peter A. Fokker, Mark Vrijlandt, Jan-Diederik van Wees, An updated geothermal model of the Dutch subsurface based on inversion of temperature data, Geothermics, Volume 88, 2020, 101880, ISSN 0375-6505,
<https://doi.org/10.1016/j.geothermics.2020.101880>.
<https://www.sciencedirect.com/science/article/pii/S0375650519304006>

Esteves Martins, J., Van Linden, E., De Bruin, G., 2021, Effects of former coal mining: upward drillings and sinkhole formation. TNO 2021 R12776, p. 24

EZK, 2022, Delfstoffen en aardwarmte in Nederland. Een overzicht van opsporings- en winningsactiviteiten en ondergrondse opslag. Jaarverslag.
<https://www.nlog.nl/sites/default/files/2024-02/Jaarverslag%202022%20-%20Delfstoffen%20en%20aardwarmte%20in%20Nederland.pdf>

J.R. Etherington, J.E. Ritter, The 2007 SPE/WPC/AAPG/SPEE Petroleum Resources Management System (PRMS), J Can Pet Technol 47 (08). Paper Number: PETSOC-08-08-15-GE.
<https://doi.org/10.2118/08-08-15-GE>

D. C., Farrow, L. C. Brooks, S. Hyun, R.mJ. Tibshirani, D. S. Burke, R. Rosenfeld, A human judgment approach to epidemiological forecasting, March 10, 2017
<https://doi.org/10.1371/journal.pcbi.1005248>

N. Galan, Knowledge loss induced by organizational member turnover: a review of empirical literature, synthesis and future research directions (Part I), The Learning Organization, Volume 30, Issue 2, 2023, Pages 117-136, ISSN 0969-6474,
<https://doi.org/10.1108/TLO-09-2022-0107>.

(<https://www.sciencedirect.com/science/article/pii/S0969647423000587>)

Hanton, C., Anantharamu, V., Automating the data flow – how AI and ML can reimagine subsurface data management, First Break, volume 41, issue 12, 2023, p. 47-50

Mark Helm, 2 - The usage of memory in current systems, Editor(s): Andrea Redaelli, Fabio Pellizzer, In Woodhead Publishing Series in Electronic and Optical Materials, Semiconductor Memories and Systems, Woodhead Publishing, 2022, Pages 27-54, ISBN 9780128207581,
<https://doi.org/10.1016/B978-0-12-820758-1.00008-X>.

(<https://www.sciencedirect.com/science/article/pii/B978012820758100008X>)

Hulsker, W., Bos, D., Schütte, H. Totaalrapport kosten en baten Basisregistratie Ondergrond (BRO). Eindrapport, Ecorys, 2021, p. 37

IenW en EZK, 2018, Structuurvisie Ondergrond,

<https://www.rijksoverheid.nl/documenten/rapporten/2018/06/11/structuurvisie-ondergrond>

IenW, LNV en EZK, 2020, Het akkoord voor de Noordzee – Afspraken tussen Rijk en stakeholders tot 2030 met een doorkijk naar de ontwikkeling van windenergie op de lange termijn,

<https://open.overheid.nl/documenten/ronl-99d46f4b-1d45-49cd-a979-2ce8bf737e22/pdf>

F. Jakobsen, P. Britze, H. Thöle, F. Jähne-Klingberg, H. Doornenbal, G.-J. Vis, Harmonized stratigraphic chart for the North Sea area NL-DE-DK, 3D Geomodeling for Europe Project number: GeoE.171.005, https://geoera.eu/wp-content/uploads/2021/11/3DGEO-EU_D3.3_Harmonized-stratigraphic-chart-for-the-North-Sea-area-NL-DE-DK.pdf

E. Karger, J. Teperowski Monrad, B. Mellers, Ph. JE. Tetlock, Reciprocal Scoring: A Method for Forecasting Unanswerable Questions, 2021, available at SSRN:
<https://ssrn.com/abstract=3954498> or <http://dx.doi.org/10.2139/ssrn.3954498>

Klammer, A. and Gueldenberg, S. (2019), "Unlearning and forgetting in organizations: a systematic review of literature", Journal of Knowledge Management, Vol. 23 No. 5, pp. 860-888. <https://doi.org/10.1108/JKM-05-2018-0277>

KNAW; NFO; NWO; TO2-federatie; Vereniging Hogescholen; VSNU (2018): Nederlandse gedragscode wetenschappelijke integriteit. DANS. <https://doi.org/10.17026/dans-2cj-nvwu>

K.T. Lance, Y.P. Georgiadou, A.K. Bregt, Evaluation of the Dutch subsurface geoportal: What lies beneath?, Computers, Environment and Urban Systems 35 (2011) p. 150-158

A. L. Leal-Rodríguez, S. Eldridge, J. L. Roldán, A. Genaro Leal-Millán, J. Ortega-Gutiérrez, Organizational unlearning, innovation outcomes, and performance: The moderating effect of firm size, Journal of Business Research, Volume 68, Issue 4, 2015, Pages 803-809, ISSN 0148-2963, <https://doi.org/10.1016/j.jbusres.2014.11.032>.

(<https://www.sciencedirect.com/science/article/pii/S0148296314003865>)

Lewis, J., New, S., Allard, J., Ancira, V., How the latest SEG-Y revision will improve data management, First Break, volume 41, issue 12, 2023, p. 65-68

Liakt, M., Hassan, M., A systematic review of literature on organizational unlearning and forgetting, J. Contemporary Issues in Business and Government, Vol. 27, no. 2. <https://cibg.org.au/>. Doi: 10.47750/cibg.2021.27.02.133

Maljers D., Stafleu J., van der Meulen M.J., Dambrink RM. Advances in constructing regional geological voxel models, illustrated by their application in aggregate resource assessments. Netherlands Journal of Geosciences - Geologie en Mijnbouw. 2015;94(3):257-270. doi:10.1017/njg.2014.46

Louise Manning, Wyn Morris, Ian Birchmore, Organisational forgetting: The food safety risk associated with unintentional knowledge loss, Trends in Food Science & Technology, Volume 118, Part A, 2021, Pages 242-251, ISSN 0924-2244, <https://doi.org/10.1016/j.tifs.2021.08.028>.
(<https://www.sciencedirect.com/science/article/pii/S0924224421005148>)

McAndrew, Th., Reich, N. G., An expert judgement model to predict early stages of the COVID-19 pandemic in the United States. PLOS computational biology. September 23, 2022
<https://doi.org/10.1371/journal.pcbi.1010485>

Meekes, J.A.C., Seismische data – DINO Nieuwe stijl, TNO 2015 R11830, 25 p.

Metaculus. (2020, June 2). A preliminary look at metaculus and expert forecasts.
<https://www.metaculus.com/news/2020/06/02/LRT/>

van der Meulen M.J., Doornenbal J.C., Gunnink J.L., et al. 3D geology in a 2D country: perspectives for geological surveying in the Netherlands. Netherlands Journal of Geosciences - Geologie en Mijnbouw. 2013;92(4):217-241. doi:10.1017/S0016774600000184

Parker, D., Lawrie, M., Hudson, P., A framework for understanding the development of organisational safety culture, Safety Science, Volume 44, Issue 6, 2006, Pages 551-562, ISSN 0925-7535, <https://doi.org/10.1016/j.ssci.2005.10.004>.
(<https://www.sciencedirect.com/science/article/pii/S0925753505001219>)

Mehvish Rashid, Paul M. Clarke, Rory V. O'Connor, A systematic examination of knowledge loss in open source software projects, International Journal of Information Management, Volume 46, 2019, Pages 104-123, ISSN 0268-4012, <https://doi.org/10.1016/j.ijinfomgt.2018.11.015>.
(<https://www.sciencedirect.com/science/article/pii/S0268401217310095>)

S. Santhakumar, C. Heuberger-Austin, H. Meerman, A. Faaij, Technological learning potential of offshore wind technology and underlying cost drivers, Sustainable Energy Technologies and Assessments, Volume 60, 2023, 103545, ISSN 2213-1388, <https://doi.org/10.1016/j.seta.2023.103545>.
(<https://www.sciencedirect.com/science/article/pii/S2213138823005386>)

Tetlock, P. (2005). Expert political judgment: How good is it? How can we know? Princeton University Press.

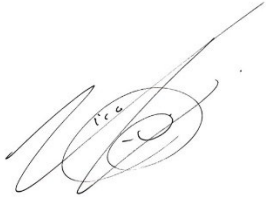
TNO (2021). GDN Knowledge Management Strategy 2022-2025. Final concept. E. De Boever, J. Griffioen, I. Kroon en J. D. van Wees met ondersteuning door W. Visser. Intern document.

TNO (2023). Menukaart 2.0.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016).
<https://doi.org/10.1038/sdata.2016.18>

Ondertekening

TNO › Energy & Materials Transition › Utrecht 22 oktober 2024

A handwritten signature in black ink, appearing to be 'Nico Zornig', is centered within a light gray rectangular box. The signature is fluid and cursive.

Nico Zornig
Research Manager

Ingrid Kroon
Auteur

Bijlage A

Maatschappelijke thema's

Deze thema's zijn opgesteld in het kader van het GDN Kennisplan (TNO, 2021).

Tabel 0.1: Maatschappelijke thema's.

Hoofdthema	Subthema
Digitale ondergrond	Digitale transformatie in gegevensbeheer
Energie Transitie	Energietransitie versnellen
Energie Transitie	Energieopslag
Energie Transitie	Geothermische energie
Energie Transitie	Opties voor hergebruik op zee
Energie Transitie	Uitfaseren aardgas
Energie Transitie	Schaarse grondstoffen
Energie Transitie	Ruimtelijke ordening
Informatieproducten	Informatieproducten
Veilige en leefbare delta	Klimaat effecten
Veilige en leefbare delta	Grondwatervoorraden
Veilig en leefbare delta	Infrastructuur
Veilig en leefbare delta	Veerkrachtige bodem
Veilige en leefbare delta	Veiligheid
Ondergrondse risico's	Lekkage en integriteit van voorzieningen
Ondergrondse risico's	Acceptatie door het publiek
Ondergrondse risico's	Seismiciteit
Ondergrondse risico's	Bodemdaling

Bijlage B

De enquêtevragen

Zie bestand BijlageB.pdf

Bijlage C

De enquête-antwoorden

Zie bestand BijlageC.pdf.

Bij open vraag 54 zijn de respondenten in de gelegenheid gesteld om per datatype aan te geven, waar wat hen betreft de grootste uitdaging op het gebied van databorging ligt. Hun antwoorden hebben we hieronder per datatype samengevat.

Bij alle datatypen wordt de inspanning, die nodig is om de gegevens te harmoniseren (inclusief een vorm van kwaliteitscontrole) als grootste uitdaging benoemd. De respondenten verwachten, dat deze inspanning het grootst zal zijn bij de biologische data (vanwege de uitzonderlijke diversiteit aan subdatasets en opslagvormen) en minder omvangrijk bij de operationele en mijnbouweffectengegevens.

Biologische data

De zorgen betreffen het overdrachtsproces, het beschikbare aantal experts om de bijbehorende werkzaamheden uit te voeren, de kans dat waardevolle kennis en vaardigheden verloren gaan en de kosten voor hernieuwde inwinning. De respondenten geven het projectteam de volgende adviezen mee: stel minimaal de fysieke referentiemonsters van de belangrijkste stratigrafische periodes veilig, digitaliseer de microscoopplaten (net als de NPD momenteel doet in Noorwegen) en maak bij inname systematisch onderscheid tussen ruwe, opgewerkte en geïnterpreteerde gegevens. De respondenten wijzen op de mogelijkheid om met deze gegevens het bestaande stratigrafische raamwerk en de daarvan afgeleide informatieproducten te verrijken.

Chemische data

De zorgen betreffen de kosten voor hernieuwde inwinning, de beperkte reproductiemogelijkheden, de opslagveiligheid (vooral van de vloeistofmonsters) en verhoogde boorrisico's bij hernieuwde inwinning - zeker in de afwezigheid van deze data. De respondenten geven het projectteam de volgende adviezen mee: achterhaal welke laboratoria betrokken waren en welke methoden zijn gebruikt, zodat de data goed kan worden geduid, maak gebruik van de ervaring bij milieuadviesbureaus en houdt rekening met het toenemende gebruik van andere vloeistoffen in de ondergrond (H₂ en CO₂). De respondenten wijzen op de mogelijkheid om met deze gegevens een centrale databank te ontwikkelen.

Schadedata

De zorgen betreffen de juridische kant (met name de privacywetgeving), de maatschappelijke gevoeligheid van schadegegevens en daarvan afgeleide interpretaties en de beperkte reproductiemogelijkheden. Wat betreft de teloorgang van kennis merkt een respondent op, dat historische schadegegevens van specifieke gebouwen hun waarde veelal verliezen op moment dat het gebouw gesloopt of vervangen wordt - en in 2050 zal dat voor veel van de huidige gebouwen het geval zijn. Tot die tijd zijn dergelijke gegevens juist waardevol (vooral voor gebouwen uit de jaren 60-80).

De respondenten geven het projectteam de volgende adviezen mee: ga eerst na hoeveel schadegegevens daadwerkelijk in het bezit zijn van de exploitanten, betrek de uitvoeringsorganisaties (met een formele rol in schadeafhandeling), sla om hergebruik eenvoudiger te maken de gegevens bij voorkeur op in een digitale kaartendatabank, bewaar (of stimuleer het behoud van) de originele gegevens (foto's etc.) met het oog op toekomstige technologische innovaties (AI) en neem omgevingsinformatie (inclusief de ondergrond) waar mogelijk mee, want het is technisch lastig om aan te tonen of schade wel/niet/deels samenhangt met ondergrondse activiteiten en processen. De respondenten wijzen op de mogelijkheid om de gebouwde omgeving robuuster te maken met behulp van deze gegevens, dit geldt met name voor nieuwbouw.

Geïnterpreteerde data

De zorgen betreffen de beperkte reproduceerbaarheid van de interpretaties (ook in relatie tot verlies van kennis en ervaring), de kosten voor databeheer en de herleidbaarheid van informatie (brongegevens, versiebeheer). De respondenten geven het projectteam de volgende adviezen mee: stel -bij voorkeur in samenspraak met de oorspronkelijke eigenaar- een kwaliteitslabel op om vervuiling door verouderde interpretaties tegen te gaan (hier wordt mee gepionierd bij geothermie), AI-reconstructies kunnen nuttig blijken maar vereisen 'oude' voorbeelden, die daarvoor beschikbaar moeten blijven, breukinformatie vereist extra aandacht (omdat de breukenstratigrafie een nieuwe ontwikkeling is), net als het borgen van de numerieke modellen van exploitanten en overweeg om ook dit type informatie bij derden (adviesbureaus) mee te nemen. De respondenten wijzen op de mogelijkheid om de interpretaties deels te integreren in de GDN informatieproducten en ze daarmee breed toegankelijk te maken. Eén respondent verwoordt het als volgt: "Integrating research insights/data from previously performed studies in existing databases and models [...] can in my opinion be classified as relatively low effort for high reward."

Mijnbouweffectendata

De zorgen betreffen de beperkte reproduceerbaarheid, de benodigde expertcapaciteit en de kosten voor databeheer. Dit betreft veelal tijdseriegegevens, waarbij de kans dat een aanzienlijk deel van de gegevens überhaupt niet wordt opgeslagen of direct na eerste analyse wordt weggegooid aanzienlijk is. De respondenten geven het projectteam de volgende adviezen mee: maak onderscheid tussen incident-gedreven en meer continue meetreeksen, behoud de metadata (in het bijzonder de laboratoriummetingen en de gemaakte keuzes bij de verschillende bewerkingsstappen), bewaar de originele gegevens zorgvuldig met het oog op technologische innovaties (AI) en stem af met de andere organisaties, die dit datatype beheren (in het bijzonder RWS en KNMI). Tenslotte wijzen de respondenten op de omvangrijke hoeveelheid meetgegevens (InSAR, LiDAR, GPS etc.), die momenteel voor monitoringsdoeleinden wordt geproduceerd (waarbij deze nieuwe datasets overigens niet noodzakelijkerwijs eenvoudiger zijn om te beheren dan de historische gegevens).

Operationele data

Ook hier maken meerdere respondenten (14) zich zorgen over de inspanning, die nodig is om de geïnterpreteerde gegevens te harmoniseren. Waar de ene respondent aangeeft: "... this type of data is very important for operators. Most data are therefore in more or less the same format", zegt de andere: "The original operator might have details on e.g. the old exploration wells and workovers but if taken over, I doubt how this type of data is transferred to successive owners. Particularly relevant for the pre-digital age where everything was on paper...at best these are now scanned pdf's".

Andere zorgen betreffen de vrijwel afwezige reproduceerbaarheid en de kans dat kennis en vaardigheden voorgoed verloren gaan na beëindiging van de gaswinning. Een aantal

respondenten geven het projectteam een advies met de volgende strekking mee: geef om de kosten voor borging en hergebruik te drukken expliciet aandacht aan innovaties op het gebied van dataverwerking (AI) en stem dit af met EBN en SodM, stel eerst de locatie-onderzoeksgegevens, die zijn ingewonnen voorafgaand aan de installatie van platforms en andere faciliteiten (en momenteel van belang bij de aanleg van windparken op zee) veilig, geef voldoende aandacht aan het bovengrondse deel van de operationele datasets en aan veranderingen in de putstatus/afwerking.

Fysische data

De zorgen betreffen de vrijwel afwezige reproduceerbaarheid, de kosten voor databeheer, de overdracht, het behoud van metadata en de doorgaande veroudering van de originele dataopslagmedia. Het algemene beeld wordt door de volgende respondentenuitspraak goed samengevat: "Data are unique, highly valuable and costs are extremely high to replace them. Archiving current dataset is much cheaper." De respondenten geven het projectteam de volgende adviezen mee: stel minimaal alle gegevens over de initiële toestand van de ondergrond (direct na aanboren) veilig (want dit bepaalt de randvoorwaarden in modelstudies), anticipeer op de volgende generatie technologie door de originele gegevens veilig te stellen (prestacks, angle stacks, logs, site/HD surveys etc.), leg metadata (locatieprecisie!) en tussenstappen systematisch vast en maak daarbij verstandige keuzes om de omvang van de databestanden werkbaar te houden (hiermee vergeleken is de omvang van de BRO beperkt!); houdt vanwege die omvang rekening met de noodzaak om nieuwe datastandaarden (API's) te ontwikkelen en leg de fysische parameters bij voorkeur vast in een formele databank om de informatie direct door te kunnen linken naar de GDN informatieproducten, zoals ThermoGIS.

Bijlage D

Kostenberekening

Scenario 1 Inname, opslag en vrijgave data binnen het bestaande datamodel

In scenario 1 gaat het om datasets (geen datatypen), die passen binnen de huidige datastromen. Daardoor zijn er geen extra kosten nodig voor systeemontwikkeling en kan er gebruik worden gemaakt van de reguliere vrijgaveprocessen. We beperken de kostenberekening hier tot de inname- en beheerskosten van twee datasets (de pre-2003 data en de seismische veld-data op tapes). We schatten in dat dit maximaal 10% van de totale data omvat (en hebben dit niet verdisconteerd in de rest van de berekeningen).

Pre-2003 datasets

Uitgangspunt voor onze kostenschatting van de pre-2003 dataset zijn de huidige kosten bij de GDN voor het beheer en vrijgave van de post-2003 data, die is opgebouwd over een periode van twintig jaar (circa 750 kEuro/jaar). Daarbij nemen we aan dat de pre-2003 dataset qua omvang hiermee vergelijkbaar is; de totale periode is langer, maar in de beginjaren is er minder data ingewonnen dan later. Ook nemen we aan, dat ongeveer de helft van de data digitaal aanwezig is en de andere (oudere) helft alleen in papieren vorm; waardoor de kosten voor de extra digitalisatieslag grofweg twee keer zo hoog zullen uitvallen. Verder passen we de vuistregel toe, dat de gewone innamekosten ongeveer 10% van de jaarlijkse beheerskosten bedragen. Dat leidt tot de volgende kostenschatting voor de pre-2003 data:

- Extra kosten beheer en vrijgave (jaarlijkse kosten):
 - Huidige inspanning beheer en vrijgave van huidige data x 0,9 (minus deel inname).
 - Deze kosten zijn ingeschat op: 680 kEuro/jaar.
- Extra kosten inname (eenmalige kosten):
 - Huidige inspanning beheer en vrijgave van huidige data x 0,1. Deze kosten zijn ingeschat op: 75 kEuro.
 - Plus extra inspanning digitaliseren van nog niet digitale data. Deze kosten zijn ingeschat op (0,5 x de totale inspanningskosten): 38 kEuro.
 - Plus extra inspanning om specifieke data van exploitanten te verkrijgen (inclusief een eenvoudig vorm van kwaliteitscontrole). Deze kosten zijn gemiddeld ingeschat op 25 kEuro per exploitant (waarbij de bedrijven met veel data natuurlijk meer inspanning kosten dan die met minder data).
 - Dat is een totaal van ongeveer 370 kEuro eenmalige kosten.

Seismische veld-datatapes

Als uitgangspunt gebruiken we de huidige kosten bij de GDN voor het innemen en beheren van deze ruwe seismische velddata. Het innemen van deze dataset is dermate arbeidsintensief, dat dit de geschatte, totale kosten domineert. Voor de beheerskosten kijken we naar de huidige beheerskosten voor de al aanwezige seismiek en dit is doorberekend met de verwachte toename van de hoeveelheid data. Op dit moment beheert de GDN 5% van de seismische velddata; de overige 95% is nog in beheer bij de exploitanten. Dat leidt tot de volgende kostenschatting:

- Extra kosten inname (eenmalig):
- De innamekosten voor ~5% van de tapes, die al zijn overdragen aan GDN, zijn grofweg 800 kEuro. De inname van de overige 95% kost dan ~15 MEuro.
- Extra kosten beheer (jaarlijks):

- De jaarlijkse beheerkosten zijn geschat met behulp van de jaarlijkse beheerkosten, die nu voor seismiek gemaakt worden (dit is grofweg 100 kEuro/jaar) en gecorrigeerd voor het deel dat nu naar schatting wordt besteed aan alleen het beheer van de velddata (ongeveer 30 kEuro/jaar). Daarmee komen we uit op ~600 kEuro.

Scenario 2 Inname, opslag en vrijgave data zonder integratie in de GDN databank

Dit scenario betreft alle digitale datasets van alle zeven datatypen. Naast de inname- en beheerkosten zijn voor dit scenario ook de kosten voor systeemontwikkeling relevant, want hoewel we in dit scenario de data 'as is' innemen met een beperkte vorm van kwaliteitscontrole blijft een lichte vorm van standaardisatie nodig zolang de datasets van elk datatype door verschillende exploitanten apart en in hun eigen structuur aangeleverd worden. Hiermee kunnen we de data op een uniforme manier (lokaal) opslaan en we kunnen metadata extraheren. Daarna kan de data worden vrijgegeven als aparte databank via een uitgifteportaal (bijvoorbeeld nlog.nl). De databank kan afhankelijk van de gebruikerswensen statisch worden aangeboden of periodiek geactualiseerd worden, waarbij de laatste optie arbeidsintensiever en dus duurder is. We gaan hier uit van een statisch aanbod. De vrijgave doen we op projectbasis en daarom nemen we in deze berekening geen vrijgavekosten mee.

Als voorbeeld voor de kosten van een 'gemiddelde digitale dataset' gebruiken we onze ervaringen met een geprepareerde dataset met drukdata, waarbij we de berekening uitsplitsen per kostenpost:

- Eenmalige inname: 10 - 20 kEuro
- Jaarlijkse beheerkosten: 1 - 2 kEuro

Om tot de cumulatieve eenmalige kosten te komen voor alle digitale datasets van alle datatypen vermenigvuldigen we deze eenmalige kosten voor de inname van één gemiddelde dataset eerst met de totale, relatieve inname- en beheerinspanning voor alle datatypes samen (ten opzichte van een gemiddelde dataset voor de digitale data), ofwel een factor 16 (Tabel 0.2).

Deze factor is primair gebaseerd op twee variabelen, te weten de kwalitatieve inschatting van de borgingskosten per datatype tot het jaar 2050 (Figuur 2.6) en de kwalitatief ingeschatte, unieke hoeveelheid data per datatype (Figuur 2.4). De in de enquête gebruikte meetlat voor de borgingskosten neemt de 'huidige (jaarlijkse) gegevensbeheerinspanning van de GDN voor hetzelfde datatype' als referentie. We nemen aan, dat deze referentie gelijk staat aan 30 x de huidige inspanning voor een gemiddelde dataset. De onzekerheid rondom deze aanname is aanzienlijk (15 tot 50 x de huidige inspanning voor een gemiddelde dataset). De kwalitatief ingeschatte, unieke hoeveelheid data per datatype (Figuur 2.4) vertalen we als volgt naar een kwantitatieve schatting:

- Erg groot: 4 x gemiddeld
- Groot: 3 x gemiddeld
- Redelijk: 2x gemiddeld
- Gemiddeld
- Klein: 0,5 x gemiddeld
- Erg klein: 0,25 x gemiddeld

Hieruit volgen de hoeveelheid data en de borgingsinspanning per datatype ten opzichte van een gemiddelde dataset (Tabel 0.2). Als laatste stap maken we onderscheid tussen de fysieke en de digitale data met als aanname (conform paragraaf 3.3), dat voor de biologische en chemische data zowel de hoeveelheid data als de borgingsinspanning voor de digitale data met een factor 3 daalt (Tabel 0.2). De fysieke biologische en chemische data laten we in de kostenberekening (voor nu) buiten beschouwing (zie hoofdstuk 5). Op deze wijze komen we tot een factor 16 voor de digitale data (Tabel 0.2).

Daarnaast omvat dit scenario de unieke hoeveelheid digitale data, die aanwezig is bij alle exploitanten. Daarom vermenigvuldigen we nogmaals met de totale, relatieve borgingsinspanning voor alle exploitanten samen (ten opzichte van een gemiddelde hoeveelheid uniek beschikbare data per exploitant). Deze factor voor de totale inspanning voor alle bedrijven gezamenlijk is (afgerond) 12, gebaseerd op de hoeveelheid data per exploitant en de daarvan afgeleide, relatieve inname- en beheerinspanning ten opzichte van een exploitant met een gemiddelde hoeveelheid unieke data (Tabel 0.2). De hoeveelheid data per exploitant baseren we op onze expertinschatting en (deels vertrouwelijke) informatie in onder meer de winningsplannen. Om deze schatting te kunnen kwantificeren zijn intensieve contacten met de betreffende exploitanten nodig. Dat valt buiten de reikwijdte van dit onderzoek (zie paragraaf 1.4).

Tabel 0.2: Kwantificatie van borgingsinspanning per datatype ten opzichte van een gemiddelde dataset.

Datatype	Hoeveelheid data			Borgingsinspanning		
	Alles	Digitaal	Fysiek	Alles	Digitaal	Fysiek
Biologisch	3	1	2	30	10	20
Chemisch	3	1	2	3	1	2
Schade	1	1	-	3	3	-
Geïnterpreteerd	4	4	-	30	30	-
Mijnbouw	2	2	-	10	10	-
Operationeel	3	3	-	5	5	-
Fysisch	4	4	-	5	5	-
Totaal	20	16	4	89	64	22

Tabel 0.3: Kwalitatieve inschatting van de hoeveelheid uniek beschikbare data per exploitant en de bijbehorende inname- en beheerinspanning ten opzichte van een gemiddelde hoeveelheid unieke data per exploitant op alfabetische volgorde. Merk op dat Spirit en Jetex hierin niet zijn meegenomen; Spirit produceert niet meer en Jetex beschikt alleen over opsporingsvergunningen.

Exploitant	Hoeveelheid data	Inname- en beheerinspanning
DANA	Laag	0,5
Kistos	Zeer laag	0,25
NAM offshore	Zeer hoog	2
NAM onshore	Zeer hoog	2
Neptune	Hoog	1,5
ONE-Dyas	Gemiddeld	1
Petrogas	Laag	0,5
TAQA	Laag	0,5
Total	Hoog	1,5
Vermillion	Gemiddeld	1
Wintershall	Hoog	1,5
Totale inname en beheerinspanning t.o.v. gemiddelde inname		12

Inname

De geschatte kosten voor eenmalige inname per exploitant en per datatype:

- 10 - 20 kEuro per gemiddelde dataset
- Voor alle datatypes: 16 maal de inname- en beheerinspanning ten opzichte van een gemiddelde dataset (Tabel 0.2)
- Voor alle exploitanten: 12 maal de borgingsinspanning per datatype ten opzichte van een gemiddelde dataset (Tabel 0.3)
- *Innamekosten * factor 16 * factor 12 geeft een totaal van 1,96 – 3.92 MEuro*

Op basis van de voor de analoge dataset gemaakte kosten schatten we de eenmalige kosten voor de (lichte) standaardisatie per datatype (we nemen aan, dat er geen extra kosten zijn voor de verschillende exploitanten) op 25 kEuro. De totale kosten voor de zeven datatypes samen bedraagt dan 175 kEuro.

Beheer

Geschatte jaarlijkse beheerkosten per exploitant en datatype:

- 10% van de inname- en ontwikkelkosten:
- *Totaal: 196 – 392 kEuro/jaar*

Samengevat zijn de totale eenmalige ontwikkel- en innamekosten en de jaarlijkse beheerinnamen in scenario 2:

- Ontwikkel en inname: 2 – 4 MEuro
- Beheer: 200 – 400 kEuro/jaar

Scenario 3: Inname, opslag en vrijgave met beperkte integratie in de GDN databank

Bij dit scenario vinden de bewerkingsstappen van de data uit scenario 2 plaats, met daarbij enkele extra stappen, waardoor er meer mogelijkheden zijn voor datavrijgave. Bijvoorbeeld door de dataset op te delen in losse ‘pakketjes’ van data, die op een logische manier bij elkaar horen. Zo kan een dataset met meetgegevens van boorgaten opgedeeld worden in de meetgegevens per boorgat en dit ‘pakketje’ kan via de koppeling aan het boorgatobject als los bestand (excel, pdf) worden opgeslagen in de GDN databank. Deze bestanden kunnen dan via de bestaande uitgifteportalen onder het object boorgaten gevonden en gedownload worden. Deze manier van borgen biedt gebruikers (vooral als ze weinig vertrouwd zijn met AI/ML, zie paragraaf 3.3) meer mogelijkheden in vrijgave, zoals het benaderen van data via een kaartviewer.

Een goed voorbeeld van deze aanpak is de put- en pompproevendataset, welke te benaderen is via broloket.nl. De kosten hiervan waren ongeveer 10x zo hoog als bij het vorige scenario. Dat betekent voor de totale eenmalige ontwikkel- en innamekosten en de jaarlijkse beheerinnamen dan ongeveer:

- Ontwikkel en inname: 20 – 40 MEuro
- Beheer: 2 – 4 MEuro/jaar

Als gekozen wordt om voor een exploitant of een datatype/set de data eerst in te nemen via scenario 1, dan kunnen deze gemaakte kosten bij latere opwerking in mindering kunnen worden gebracht op de kosten van scenario 3.

Scenario 4: Inname, opslag en vrijgave met volledige integratie in de GDN databank

Ook bij dit scenario vinden de bewerkingsstappen van de data uit scenario 2 plaats. Vervolgens wordt het datamodel van de GDN databank aangepast of uitgebreid, zodanig dat de

data volledig geïntegreerd kunnen worden. De data wordt net als in scenario 3 eerst opgedeeld in 'pakketjes', die vervolgens worden gemigreerd naar de GDN databank. Door deze manier van borgen komt de samenhang met de andere datasets binnen de databank tot stand en daardoor kan de data op een groot aantal manieren worden vrijgegeven en gebruikt (als onderdeel van een analyse, in een kaartviewer etc.).

Momenteel is de GDN databank in ontwikkeling en zijn de eerste dataobjecten vanuit DINO gemigreerd naar de nieuwe omgeving, die meer lijkt op de BRO-omgeving. Het is te vroeg om een goede vergelijking te kunnen maken op basis van de ervaringen tot nu toe, al zal het voor sommige datatypes eenvoudiger zijn dan voor andere. Om tot een eerste inschatting te komen, gebruiken we de ingeschatte borgingsinspanning voor alle datatypes (Tabel 0.2) plus de aanname, dat de GDN databank aanpassen voor een gemiddelde datatype ongeveer 250 kEuro kost. Het totaal voor ontwikkelingskosten komt dan uit op ongeveer 16 MEuro.

De kosten voor inname en beheer zijn verder vergelijkbaar met scenario 2. De totale kosten bedragen:

- Ontwikkelkosten: circa 16 MEuro
- Innamekosten: 20 – 40 MEuro
- Beheerkosten: 2 – 4 MEuro/jaar

Bijlage E

Perspectieven

In aanvulling op de aannames en de onzekerheid in de gedane kostenschattingen (Tabel 3.2) gelden voor de berekeningen de volgende aannames:

- De prioritering gaat ervanuit dat er een jaar de tijd is om bij GDN de benodigde voorbereidingen te kunnen treffen. Dit is vooral van belang bij een tijdhorizon van vijf jaar;
- We nemen aan dat alle exploitanten binnen vijftien jaar vertrekken en we hanteren een geleidelijk profiel van de vertrekkende exploitanten als functie van de tijd¹², resulterend in een inname- en beheerinspanning ten opzichte van het gemiddelde;
- We nemen de gemiddelde cijfers van de in hoofdstuk 3 bepaalde kosten;
- We houden geen rekening met toekomstige inflatie;
- We houden geen rekening met eventuele kostenveranderingen door innovaties (zoals AI, OCR en opslagcapaciteit)
- De beheerkosten worden op jaarbasis gemaakt voor de totale hoeveelheid data, die in een specifiek perspectief geborgd wordt.

Voor sommige perspectieven zijn daarboven extra aannames nodig. Deze worden vermeld in de betreffende paragraaf.

Merk op, dat we de bedragen pas aan het eind afronden.

Perspectief 1: Borgen van pre-2003 en seismische velddata in 5 jaar

Voor dit perspectief (scenario 1) geldt dat de data in een periode van 5 jaar worden geborgd. Als we er vanuit gaan, dat we na een jaar voorbereiding in de volgende 4 jaar de data lineair innemen (en dus gaan beheren), dan zijn de totale kosten over deze periode zoals gegeven in onderstaande tabellen, afhankelijk van de keuze om de seismische velddata wel (scenario 1a) of niet (scenario 1b) te betrekken (Tabel 0.4).

Tabel 0.4: Kosten voor ontwikkeling, inname en beheer voor perspectief 1; scenario 1a.

Jaar	Eenmalige ontwikkelkosten (kEuro)	Inname van alle data (kEuro)	Beheerskosten (kEuro)	Totaal (kEuro)
1	0	0	0	0
2	0	3850	163	4013
3	0	3850	488	4338
4	0	3850	813	4663
5	0	3850	1138	4988
Totaal	0	15400	2602	18002

¹² Eerste 5 jaar: 5 x gemiddelde; 6^e – 10^e jaar: 4 x het gemiddelde; 11^e – 15^e jaar: 3,25 x het gemiddelde

Perspectief 2: Borgen overige data in 5 jaar

Conform scenario 2 nemen we de digitale data 'as is' met beperkte kwaliteitscontrole, we slaan de data lokaal op en we geven de data als databank vrij, zonder link/integratie met andere datasets. Daarbij borgen we de data van alle huidige exploitanten in Nederland in de aankomende 5 jaar met als aanname, dat we in het eerste jaar met name met de voorbereidende werkzaamheden van het systeem bezig zijn (en 0,25 van het geheel innemen) en in de vier jaar erop we steeds 3 x (tot de 12,25 dat bij de exploitanten ligt, zie Tabel 0.3) van de innameactiviteiten uitvoeren. Op basis van de in Tabel 3.2 gegeven bedragen komen we dan tot de in Tabel 0.5 gegeven kosten.

Tabel 0.5: Kosten voor ontwikkeling, inname en beheer voor perspectief 2.

Jaar	Eenmalige ontwikkelkosten (kEuro)	Inname van alle data (kEuro)	Beheerskosten (kEuro)	Totaal (kEuro)
1	175	60	3	238
2	0	735	43	778
3	0	735	116	851
4	0	735	200	935
5	0	735	263	998
Totaal	175	3000	615	3790

Perspectief 3: Alles borgen van die exploitanten, die in de aankomende 5 jaar vertrekken

- Dit perspectief speelt in op de noodzaak om van die bedrijven, die de aankomende 5 jaar vertrekken, zo veel mogelijk data te borgen en met het oog op hergebruik vrij te kunnen geven (wellicht) via de GDN databank. Dat leidt tot 4 getrapte mogelijkheden: Perspectief 3a: Minimaal de pre-2003 datasets (conform scenario 1b)
- Perspectief 3b: Minimale borging van de overige datasets (conform scenario 2)
- Perspectief 3c: Combinatie van 3a en 3b (alle data van deze bedrijven minimaal geborgd, uitgezonderd de seismische velddata)
- Perspectief 3d: Alle data van deze bedrijven geborgd in de GDN databank.

De relatieve hoeveelheid data van de in de aankomende 5 jaar vertrekkende exploitanten schatten we op 41% (5/12,25, zie Tabel 0.3). Met dit gemiddelde en de berekende kosten in paragraaf 4.1 en 4.2 en Tabel 3.2 leidt dat tot de in Tabel 0.6 gegeven jaarlijkse en totale kosten.

Tabel 0.6: Kosten voor ontwikkeling, inname en beheer voor perspectief 3a - 3d.

Jaar	3a (Keuro)	3b (Keuro)	3c (Keuro)	3d(Keuro)
1	0	98	98	16000
2	73	319	392	3331
3	142	349	491	3844
4	212	383	595	4356
5	282	409	691	4869
Totaal	709	1554	2263	32400

Perspectief 4: Top 15 in 15 jaar geborgd in de GDN databank

Dit perspectief berust op de gedachte, dat we met 20% van alle data zeker 80% van alle voor hergebruik belangrijke data borgen. Hiervoor prioriteren we op basis van de relatie tussen het hergebruikspotentieel per datatype (# must haves) per maatschappelijke subthema door de top 15 uit paragraaf 2.4 te gebruiken.

We nemen aan, dat de top 15 een totale borgingsinspanning behoeft van 20% van de totale behoefte¹³. We gebruiken de kostenschattingen voor scenario 4 in Tabel 3.2 en benutten het eerste jaar voor ontwikkeling van het systeem. Verder rekenen we met totalen per 5 jaar (per jaar geeft een schijnnaauwkeurigheid). In Tabel 0.7 geven we het overzicht van de berekende kosten.

Tabel 0.7: Kosten voor ontwikkeling, inname en beheer voor perspectief 4.

Jaar	Eenmalige ontwikkelkosten (kEuro)	Inname van alle data (kEuro)	Beheerskosten (kEuro)	Totaal (kEuro)
1	3200	0	0	3200
5	0	2450	1225	3675
10	0	1960	3430	5390
15	0	1590	5205	6795
Totaal	3200	6000	9860	19060

Perspectief 5: Alles minimaal geborgd in 15 jaar plus de top 15 geborgd in de GDN databank

Dit perspectief combineert de gedachte, dat we met 20% van alle data zeker 80% van alle voor hergebruik belangrijke data borgen in de GDN databank, waarbij de resterende data minimaal wordt geborgd (scenario 2) en trekt hiervoor 15 jaar uit.

Hiervoor gebruiken we de getallen uit paragraaf 4.2 en 4.4, waarbij we verrekenen dat de kosten vanuit paragraaf 4.2 niet over 5 maar 15 jaar wordt verspreid, en dat 20% van de data vanuit scenario 2 al wordt meegerekend met het scenario uit 4.4. Dat leidt tot onderstaande tabel 0.8 met het overzicht van de kosten.

Tabel 0.8: Kosten voor ontwikkeling, inname en beheer voor perspectief 5.

Jaar	Eenmalige ontwikkelkosten (kEuro)	Inname van alle data (kEuro)	Beheerskosten (kEuro)	Totaal (kEuro)
1	4600	0	0	4600
5	0	3430	1470	4900
10	0	2744	4116	6860
15	0	2226	6246	8472
Totaal	4600	8400	11832	24832

¹³ Het is te complex om dit te berekenen (uit Figuur 2.6 in paragraaf 2.4), onder meer omdat er een verband moet worden gezocht tussen de borgingsinspanning per item en het deel dat dat vertegenwoordigd in een datatype.

Perspectief 6: Alles in 15 jaar geborgd in de GDN databank

Dit laatste perspectief geeft een overzicht van de kosten, die gemoeid zijn met het borgen van de digitale data (conform scenario 4) in de GDN databank over een periode van 15 jaar. Hiervoor gebruiken we de getallen uit paragraaf 4.4 en vermenigvuldigen dit met 5. In Tabel 0.9 de resultaten.

Tabel 0.9: Kosten voor ontwikkeling, inname en beheer voor perspectief 4.

Jaar	Eenmalige ontwikkelkosten (kEuro)	Inname van alle data (kEuro)	Beheerskosten (kEuro)	Totaal (kEuro)
1	16000	0	0	16000
5	0	12225	6125	18350
10	0	9800	17150	26950
15	0	7950	26025	33975
Totaal	16000	30000	49300	95300

Overzicht van de kosten per perspectief

In Tabel 0.10 vergelijken we de kosten van de verschillende perspectieven.

Tabel 0.10: Vergelijking van de kosten van de verschillende perspectieven op basis van de berekening in Bijlage E. De onzekerheid in de absolute bedragen is aanzienlijk (ruim meer dan 40%, zie ook Tabel 3.2). Doordat de onderliggende aannamen bij de perspectieven veelal dezelfde zijn, is de relatieve verdeling van de kosten over de perspectieven indicatief voor de onderlinge kostenverhoudingen.

Handelings- perspectief	Kosten (kEuro per tijdperiode in jaren vanaf start)								
	1	2-5		6-10		11-15			
		Totaal	Jaarlijks	Totaal	Jaarlijks	Totaal	Jaarlijks	Totaal	Jaarlijks
1a	0	18002	4500	6500	1300	6500	1300	31002	2067
1b	0	1730	433	3400	680	3400	680	8530	569
2	238	3615	904	1500	300	1500	300	6853	457
3a	0	709	177	1394	279	1394	279	3497	233
3b	98	1456	364	777	155	777	155	3108	207
3c	98	2165	541	2171	434	2171	434	6605	440
3d	16000	16400	4100	15000	3000	15000	3000	62400	4160
4	3200	3675	735	5390	1078	6795	1359	19060	1271
5	4600	4900	980	6860	1372	8472	1694	24832	1655
6	16000	18350	3670	26950	5390	33975	6795	95300	6353

Bijlage F

Interviews

Zie bestand BijlageF.pdf.

Energy & Materials Transition

Princetonlaan 6
3584 CB Utrecht
www.tno.nl