



Nederlandse AI Coalitie

Naar First-time-Engineering en Operationalisatie

Inhoudsopgave

1	Inleiding	3
2	Stakeholders en gezamenlijke visie	4
3	Processtappen: Van PoC naar implementatie	5
3.1	Het Strategic Business Options model	5
3.2	AI applicaties en PoC's ontwikkeling	7
3.3	Het vormgeven van een experimenteel datadeel infrastructuur	9
3.4	Operationeel datadelen voor experimentele AI applicaties	11
3.5	Roadmap naar een sterk AI landschap in Nederland	14
4	Randvoorwaarden	15
5	Roadmap	16
6	Referenties	17

1 Inleiding

Dit document beschrijft een stappenplan hoe leden van de NL AIC het delen van data via experiment naar de dagelijkse praktijk kunnen brengen. Het beschrijft het trajectvoorstel van first-time-engineering naar implementatie en representeert deliverable 3 en 4 in het kader van de opdracht 'Verantwoord datadelen ten behoeve van NL AIC, Voorbereidende Fase'.

Op basis van representatieve use cases voor 'Verantwoord datadelen ten behoeve van AI-toepassingen', worden de essentiële componenten voor verantwoord datadelen gedemonstreerd door first-time-engineering in de vorm van Proofs-of-Concept (PoC's). Het doel is daarbij zichtbaar te maken hoe een datadeel implementatie voor AI er uit zou kunnen zien. Met vertegenwoordigers van de toepassingsgebieden zal worden afgestemd welke toepassingen hiervoor worden uitgewerkt.

Daarna volgt een gecontroleerde overgang naar een productie-omgeving waar zal worden ingegaan op de voor productie relevante onderwerpen die niet eerder aan bod zijn gekomen.

In deze deliverable zullen we uitwerken hoe zo'n traject eruit kan zien en wat deelnemers kunnen verwachten.

Aan de hand van dit document kunnen de partners die willen samenwerken zich oriënteren op de onderwerpen die aan bod moeten komen richting implementatie en uiteindelijk operationalisatie van de datadeel oplossing. Daarmee biedt dit document een houvast voor betrokken partijen en is dus feitelijk de voorbereiding voor het daadwerkelijk aan de slag gaan met datadelen in NL AIC.

Er is geen gedetailleerde blauwdruk voor deze route. Het is een gezamenlijk ontwerpproces dat afhangt van de genomen uitgangspunten voor de PoC's en de ervaringen en kennis die gaandeweg opgedaan wordt. Het geeft partners een methodiek om tot een exploiteerbare datadeel infrastructuur te komen die gebaseerd is op een model van TNO. Deze ontwerpaanpak zal nader toegelicht worden in paragraaf 3.1.

In het volgende hoofdstuk 2 gaan we in op de toekomstvisie voor datadelen en stakeholders in een samenwerking. Na een introductie van het Strategic Business Options Model, beschrijven we vervolgens in hoofdstuk 3 het proces van PoC naar operationalisatie in verschillende stappen. Bij iedere stap worden mogelijk relevante strategische open keuzes aangereikt ter inspiratie. In hoofdstuk 4 worden kort de randvoorwaarden voor dit proces besproken, zodat betrokken organisaties weten waar zij aan toe zijn. Tot slot geeft hoofdstuk 5 een eerste high-level versie ontwikkel roadmap voor het traject via first-time-engineering tot aan operationalisatie van datadelen ten behoeve van AI.

2 Stakeholders en gezamenlijke visie

Datadelen is een belangrijke voorwaarde voor het verbeteren van de positie van Nederland op het gebied van AI. Door het cross-organisatorisch delen van data zullen nieuwe inzichten en ongekende AI toepassingen mogelijk worden, waarmee potentiële grote impact te maken is. Echter zijn de risico's ook significant, aangezien de markt nog niet uitontwikkeld is en innovatiekosten hoog zijn.

Technologieën voor (gecontroleerd en veilig) datadelen en AI toepassingen bevinden zich momenteel in de innovatiefase en maken een transitie naar de markt door. De technische ontwikkeling is met name gericht op het verbeteren van de functionaliteit en performance van de datadeel producten en services. Vervolgens worden deze innovaties in de markt geïntroduceerd en is het van belang om snelle, iteratieve leercycli te maken, zodat de producten en services optimaal aangepast kunnen worden aan de opkomende markt en gebruikersbehoeften. De herbruikbaarheid van de PoC's draagt bij aan deze snelle vooruitgang.

Voor bedrijven is het essentieel om in deze fase een hoge mate van flexibiliteit te behouden in hun strategie. Dat betekent enerzijds snelheid maken met het experimenteren en toepassen van nieuwe technologieën, anderzijds niet te vroeg op één innovatie inzetten met als gevaar dat men verkeerd "gokt". Door kennisdeling binnen een actieve community op het gebied van datadelen en AI toepassingen wordt deze innovatiesnelheid bevordert. Een bijkomend voordeel is dat innovatiekosten en risico's onderling gedeeld worden.

Datadelen is van nature een multi-stakeholder aangelegenheid en heeft daardoor een passende governance en business model nodig. Een dergelijke datadeel infrastructuur vertoont karakteristieken van een zogenaamd "*multi-sided platform business model*", doordat het de uitwisseling van data tussen twee of meer onderling afhankelijke groepen mogelijk maakt (bijv. data aanbieders en data gebruikers). Dit beschrijft de logica achter het creëren van waarde voor de gebruikers van de data en AI applicaties die daarop draaien. Als toekomstvisie voor Nederland gaan we daarom in dit document uit van een generieke datadeel infrastructuur (een gemedieerd netwerk) met generieke bouwblokken, die in principe voor alle organisaties en sectoren beschikbaar zijn.

De weg daarheen is een gezamenlijk proces wat niet volledig van te voren te definiëren is. Wel is het mogelijk om bepaalde ontwerpprincipes aan te houden die de kans op succes vergroten. Zo moet men samen bewust zijn van de strategische keuzes die te maken zijn en welke gevolgen deze potentieel kunnen hebben. Stakeholders moeten samen dit pad bewandelen en onderweg niet alleen technische, maar ook organisatorische en strategische keuzes maken, zoals hoe men toegang krijgt tot de data en wie bijdraagt aan het opzetten en in stand houden van de infrastructuur.

Belangrijk is daarom van te voren bewust te zijn welke partijen deze keuzes beïnvloeden en wie er bij dit keuzeprocess een plek aan tafel mogen (of zelfs moeten) hebben. Zo is het verstandig om na te denken over toekomstige deelnemers (i.e. tijdens de operationalisatie) en deze zo vroeg mogelijk in het gesprek betrekken. Niet alleen de data aanbieders en data gebruikers, maar juist ook de (potentiële) ondersteunende service providers, zullen vanaf het begin moeten worden betrokken om tot een goede toekomstige rolverdeling te komen.

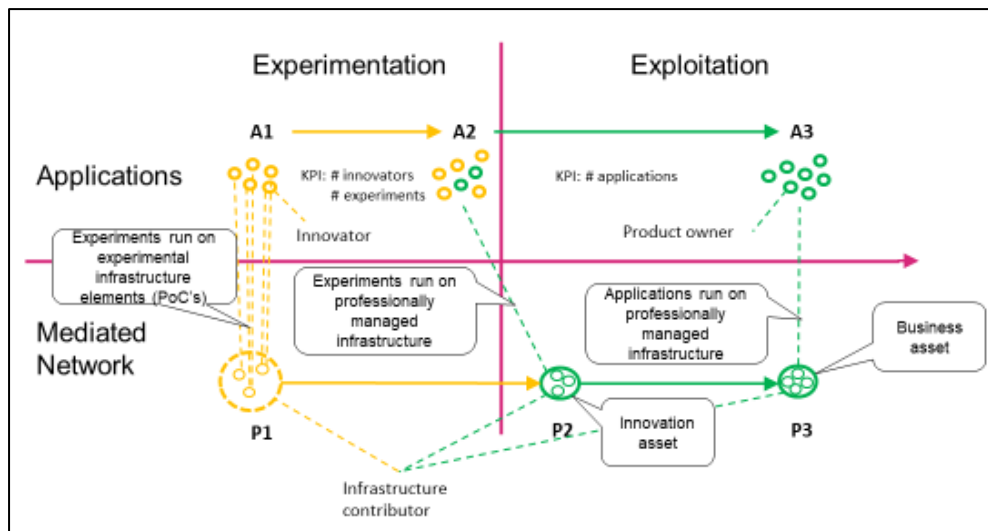
Een belangrijk aspect is hierbij ook om afstemming te houden met (inter-)nationale ontwikkelingen en initiatieven rondom datadelen, zoals benoemd in [1, sectie 1.3]. AI en datadelen houden niet op bij grenzen van sectoren of landen. Aansluiting bij (inter-)nationale ontwikkelingen rondom initiatieven zoals de Data Sharing Coalition iSHARE, IDS en GAIA-X dienen daarom beschouwd en geborgd te worden.

3 Processtappen: Van PoC naar implementatie

3.1 Het Strategic Business Options model

In deze paragraaf gaan we dieper in op het ontwerpprincipe dat beschreven wordt in figuur 1, het "Strategic Business Options Model", dat gebaseerd is op ervaringen die TNO heeft opgedaan [2]. Hieronder leggen we de onderliggende principes uit [3].

Het model beschrijft de ontwikkeling voor digitale, datagedreven diensten van de experimentele (links aangegeven in de figuur) naar de operationele fase ("exploitation", rechts). Het model stelt leden van NL AIC in staat om in een gemeenschappelijke taal over de verschillende fases van experiment naar operationalisatie te delen.



Figuur 1: Strategic Options Model. De mediated network laag stelt de infrastructuur voor die benodigd is om data te delen voor AI applicaties.

In het model zijn twee lagen te onderscheiden:

- **"Mediated Network" (onder):** Hieronder verstaan we de generieke (dadadeel en AI-faciliterende) infrastructuur met herbruikbare bouwblokken die benodigd is voor het veilig en verantwoord delen van data met verschillende partijen. Deze infrastructuur kan zowel fysieke (bijv. hardware) als ook abstracte elementen, zoals een afsprakenstelsel, standaarden, interfaces, architectuur, etc., bevatten.
- **"Applications"(boven):** AI toepassingen die gebruik maken van de data die door deze infrastructuur ter beschikking komt.

Zowel dit gemedieerd netwerk als ook de AI-toepassingen lopen verschillende fases door in hun ontwikkeling:

- **P1:** Een experimentele infrastructuur voor het delen van data, gebaseerd op generieke bouwblokken van de PoC's, om te leren hoe men een dergelijke infrastructuur opzet en onderhoudt. Hieronder valt ook het deployment en management van AI-applicaties op basis van deze data. Ontwikkelaars van AI applicaties moeten er rekening mee houden dat deze infrastructuur nog niet aan al hun eisen, zoals beschikbaarheid, stabiliteit, veiligheid, zal voldoen.
- **P2:** Een datadeel infrastructuur die voldoende professionaliteit heeft zodat deze als dienst aan gebruikers kan worden aangeboden die willen experimenteren met het

ontwikkelen en draaien van AI-applicaties. Deze AI applicaties stellen nieuwe eisen aan het datadelen zoals toegang, controle, veiligheid en feedback. Het is een *innovatie asset* die innovatieve AI-applicaties op basis van gedeelde data mogelijk maakt.

- **P3:** Dit is een operationele, generieke, infrastructuur ("Mediated Network") voor het delen van data ten behoeve van exploitierbare AI-toepassingen. Het is een *business asset*, die aan andere eisen moet voldoen dan P2, omdat er exploitierbare applicaties van afhankelijk zijn. Ook zullen andere bedrijfsrisico's en overeenkomsten van toepassing zijn dan bij P2.

In lijn met de verschillende fases in de ontwikkeling van de generieke infrastructuur doorlopen ook de AI-toepassingen een groeitraject:

- **A1:** Experimentele AI-toepassingen, voornamelijk bedoeld om vaardigheden op te ontwikkelen met het delen van data.
- **A2:** Experimentele AI-toepassingen, voornamelijk bedoeld om ervaring op te doen met het ontwikkelen van AI-applicaties ter voorbereiding op productie.
- **A3:** Exploiteerbare AI-toepassingen die voldoende betrouwbaar en stabiel zijn om aan eindgebruikers beschikbaar te stellen in ruil voor iets van monetaire waarde.

Om impact te genereren is het noodzakelijk dat AI-applicaties de transformatie maken van de A2- naar de A3-fase. Echter het kan niet uitgesloten worden dat applicaties ontwikkeld op P1 of P2 vervolgens gebruik zullen maken van een andere (publiek toegankelijke) datadeel infrastructuur dan P3 (mits deze bestaat). Ook kan het voorkomen dat verschillende instanties van P1 en P2 naast elkaar zullen bestaan, juist omdat deze ontwikkeling een continu proces is en de afzonderlijke infrastructuren ook steeds verbeterd worden met bijvoorbeeld nieuwe features. De overgang tussen de verschillende fases is daarom niet als harde scheiding te zien.

Verder kunnen er in iedere fase verschillende partijen betrokken zijn, omdat deze verschillende functies en dus waardeproposities hebben. We onderscheiden de volgende business rollen:

- **Infrastructure Contributor:** De rol van "infrastructuur contributor" is voor partijen weggelegd die technische componenten beschikbaar stellen die samen het gemedieerde netwerk om data te delen tot een functioneel geheel maken, denk aan authenticatie, opslag, algoritmes, ID management, service brokering, etc. Deze rol zal zeer waarschijnlijk door verschillende (commerciële) partijen naast elkaar ingenomen worden.
- **Application Innovator:** Dit is een partij die één of meerdere AI-applicaties (A2) wil ontwikkelen om daarmee te experimenteren.
- **Product Owner¹:** Deze partij wil een AI-application exploiteren (A3) op basis van een betrouwbare datadeel infrastructuur.

Elk van de rollen kan door meerdere partijen / organisaties worden ingevuld. Bovendien moet er een organisatie zijn die het voortouw neemt in het vormgeven en uitrollen van het gemedieerde datadeel netwerk. Deze is verantwoordelijk voor enerzijds de ontwikkeling (en standaardisatie) van de architectuur en anderzijds voor het aanjagen van de operationalisatie van generieke bouwblokken door de 'Infrastructure Contributors' en adoptie ervan door de markt van data aanbieders en gebruikers. Deze rol is vergelijkbaar met de rol van de IDS Association (voor IDS ontwikkeling en adoptie) en iShareworks (voor iSHARE ontwikkeling en adoptie). Deze organisatie kan besluiten (een deel van) deze taken uit te besteden aan een derde partij, denk aan bijv. de ontwikkeling en beheer van de architectuurprincipes en

¹ Hoewel de term Product Owner ook in agile ontwikkeling gebruikt wordt, is deze vorm van ontwikkeling geen vereiste. Deze rol kan ook vertaald worden als product uitbater.

standaarden. Deze coördinerende rol kan o.a. door een groep partners samen ingevuld worden bijv. in een onafhankelijke stichting. Dit samenwerkingsverband zal het aanspreekpunt zijn voor de hierboven genoemde rollen om te zorgen dat het gehele netwerk van datadeel bouwblokken optimaal functioneert.

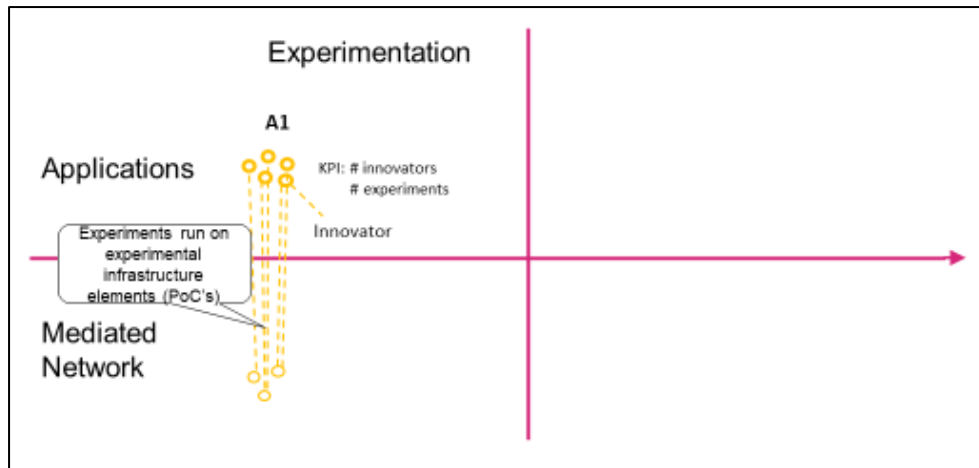
In de volgende paragrafen geven we op hoofdlijnen een procesbeschrijving waarbij we bovenstaande fases zullen doorlopen. Zoals aangegeven in hoofdstuk 2 zal aan dit proces gezamenlijk door het consortium (in een brede definitie als de groep die het systeem realiseert) verder invulling gegeven moeten worden. In iedere paragraaf wordt een aanzet gegeven van strategische keuzes die op dat moment relevant kunnen zijn. Het is aan het consortium om te besluiten welke keuzes gemaakt moeten worden. Deze strategische keuzes zijn opgedeeld in drie categorieën: technisch, organisatorisch en ecosysteem:

- *Technologie*: Het vormgeven van de architectuur benodigd voor het delen van data, het definiëren van interfaces en informatiemodellen en het demonstreren en realiseren van
- *Organisatie (governance)*: Deze is gericht op de (door)ontwikkeling van de voorgestelde aanpak, architecturen, interfaces en standaarden. Dit bevat zowel de procesinrichting hiervoor door middel van een aansturende organisatie en de inrichting van een change (advisory) board als de technische roadmap.
- *Ecosysteem*: Deze is gericht op de adoptie door organisaties en marktpartijen. Aangezien we hier te maken hebben met een implementatie door meerdere partijen, spreken we over een ecosysteem. Het is van belang dit ecosysteem zo goed mogelijk te faciliteren, ter voorbereiding en realisatie van groei. Door de opschaling van zowel partijen die verantwoord data met elkaar delen als van AI-toepassingen zelf wordt de positieve impact voor de eindgebruikers in steeds grotere mate gerealiseerd.

De keuze voor deze categorieën vloeien voort uit de aanname dat de beoogde eindsituatie een open, gefedereerde, infrastructuur is die datadelen ten behoeve van AI mogelijk maakt, met verschillende rollen die elk een generieke bouwblok als dienst operationeel in de markt aanbiedt.

3.2 AI applicaties en PoC's ontwikkeling

De ideeën, architecturen en concepten rondom datadelen voor AI zullen door middel van proof-of-concepts (PoC's) worden vormgegeven. De PoC's vertegenwoordigen elementen voor datadelen die generiek toepasbaar zijn. Samen met de sectoren worden één of meerdere toepassing scenario's gekozen om datadelen onder AI te realiseren (zie ook Figuur 2). Men moet zich er van bewust zijn dat de AI applicaties in deze fase vooral bedoeld zijn om de PoC's tot stand te laten komen.



Figuur 2: PoC's worden ontwikkeld als generieke bouwblokken voor het datadelen voor verschillende AI toepassingen (A1).

De exacte invulling van de benodigde activiteiten voor de deelnemende partijen in deze zogenoemde A1-fase zal in grote mate afhankelijk zijn van de gekozen AI toepassingen en de te ontwikkelen PoC's. Hieronder wordt besproken welke strategische keuzes in deze fase relevant kunnen zijn. Zie ook Tabel 1 voor een niet-uitputtende opsomming hiervan.

De eerste relevante keuze is welke AI applicaties gerealiseerd zullen worden. Hierbij moet men zich vooraf bedenken waarom deze toepassingen in de toekomst zinvol lijken. Als voorbeeld: het automatisch managen van gebouw-utiliteiten voor o.a. de regeling van klimaat kan potentieel grote energiebesparingen opleveren wat zowel positief bijdraagt aan de klimaatdoelstellingen als ook kostenbesparingen met zich meebrengt. Nadenken over de toekomstige waarde van een dergelijke AI applicatie zorgt er voor dat het datadelen als middel ingezet wordt om maatschappelijke impact te bereiken. Het is van belang om het doel van deze toepassingen vooraf zo helder mogelijk te beschrijven.

Verder is de ontwikkeling in grote mate afhankelijk van de partijen die gezamenlijk deze experimenten willen uitvoeren. Voor het gemak noemen we dit verder het kernteam. Dit is het moment om vast te stellen wat de individuele doelstellingen en belangen van deelnemers in het kernteam zijn. Voor sommige partijen kan het extra waardevol zijn om ervaring op te doen met het ontwikkelen van AI applicaties ("Innovator"), terwijl anderen juist op zoek zijn naar ervaring met het runnen en faciliteren van een (onderdeel van een) datadeel infrastructuur ("Infrastructuur Contributor").

Het is nuttig voor leden van het kernteam om na te denken welke rol (zie 3.1) zij in de toekomst willen innemen. Tevens moet het kernteam nagaan of en zo ja, welke elementen / partijen / rollen nog missen. Zo kan het vroegtijdig betrekken van de eindgebruikers een goed beeld opleveren van de benodigde functionaliteiten van de AI toepassing. Bovendien moet het kernteam bedenken hoe en waar vandaan zij voldoende data krijgen om in deze experimentele fase met elkaar te delen. Dit kan al een eerste inzicht geven in hoe datadeel concepten de bereidheid om (gevoelige) data te delen kan vergroten.

Het kernteam moet tot een samenwerkingsvorm komen die vooral snel en effectief is. Men moet zich bewust zijn van elkaars commitment en de tijd en resources die iedere partij bereid is in deze experimenten te steken. Ten slotte hangen er nog geen bedrijfsprocessen direct van de uitkomsten af, waardoor de meeste bijdragen op vrijwillige basis zullen zijn. Belangrijk in deze fase is het gezamenlijk leren en het uitwisselen van inzichten en ervaringen. Het kernteam moet daarvoor een manier vinden om dit borgen.

Gezien de beperkte resources is het belangrijk om focus te houden welke minimale functionaliteiten voor het delen van data in deze fase gedemonstreerd zullen worden.

Tevens kan men vooraf bedenken waar eventuele uitdagingen te verwachten zijn en of men het belangrijk vindt om deze nu aan te pakken of tot een later stadium te bewaren. Ook moeten ontwikkelaars nu al strategische keuzes maken wat betreft technologie en architectuur die de herbruikbaarheid en flexibiliteit in de toekomst zullen beïnvloeden. Mogelijk worden er verschillende instanties van PoC's naast elkaar ontwikkeld om juist de verschillen tussen technische scenario's helder te krijgen.

Tabel 1 Niet uitputtende lijst van strategische keuzes voor fase A1.

Technisch	Organisatorisch	Ecosysteem
Welke technische eigenschappen op het gebied van datadelen moeten ontwikkeld worden om de AI toepassingen te verwerklijken?	Hoe gaat het consortium praktisch samenwerken? (communicatie, meetings, documentatie, etc.)	Voor welke AI toepassingen worden PoC's ontwikkeld? Waarom is de verwachte impact van deze toepassingen het grootst?
Waar worden technische uitdagingen voorzien? Hoe gaat men hier mee om (oplossen of workaround)?	Wie zijn onderdeel van de experimenten (kernteam)? Wie worden nauw betrokken, wie op een afstand?	Wie zijn de eindgebruikers van de beoogde AI toepassingen (A3)? Hoe worden zij betrokken?
Welke technische keuzes beïnvloeden de herbruikbaarheid van de experimenten? Hoe behoud men flexibiliteit? (bijv. keuze voor een programmeertaal)	Wie levert welke resources (hardware, software, locaties, ontwikkelaars, etc.)?	Welke waarde wordt voor de eindgebruikers gecreëerd? Wat is het toekomstige business model van deze AI toepassingen?
...	Hoe wordt (experimentele) snelheid gecreëerd?	Wie zijn de 'innovators', de partijen die experimentele ervaring met deze AI toepassingen op willen doen?
	Hoe worden learnings met elkaar en andere leden van NL AIC gedeeld?	...

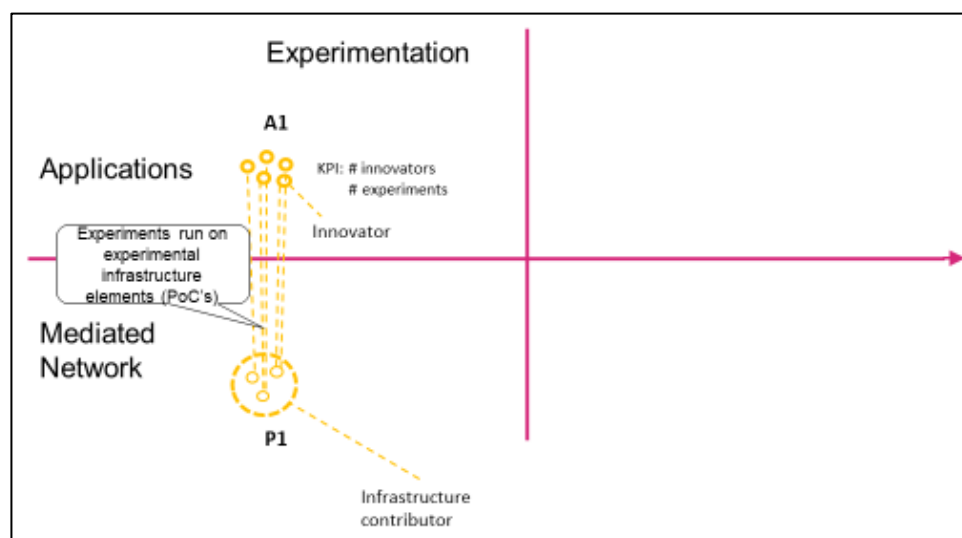
3.3 Het vormgeven van een experimenteel datadeel infrastructuur

Nadat eerste ervaringen met de ontwikkeling van verschillende PoC's opgedaan zijn kunnen partijen starten met het creëren van "P1", één experimenteel onderling gedeelde infrastructuur voor het runnen van meerdere experimentele applicaties. Waar de vorige PoC's alleen bedoeld waren om aan te tonen dat het concept van datadelen in één specifieke AI-toepassing mogelijk is, kijken we hier naar de ontwikkeling van een prototype voor het delen van data. Om te bepalen hoe deze 'minimal viable infrastructure' er technisch uitziet zal men ook hier een vorm moeten vinden om inzichten uit de PoC's te delen.

Het is niet noodzakelijk te wachten met het finaliseren van fase A1 alvorens te beginnen met het definiëren en uitvoeren van fase P1. Zodra eerste ervaringen tussen PoC's onderling

uitgewisseld kunnen worden, kan men beginnen met identificeren hoe deze generieke componenten elkaar kunnen aanvullen en wat er nog mist om deze in een experimenteel datadeel infrastructuur P1 te bundelen (zie

Figuur 3). Ook moet men zich afvragen of alle benodigde componenten voor een dergelijk prototype aanwezig zijn c.q. door een partij omgezet kunnen worden. Zo niet, dan moet deze partij gevonden en betrokken worden.



Figuur 3: In deze fase word gekeken of de verschillende PoC's kunnen worden gecombineerd tot één experimentele datadeel infrastructuur (P1). Het is zinvol de toekomstige 'infrastructuur contributor(s)' hierbij te betrekken.

Er vanuit gaande dat iedere PoC een eigen kernteam heeft bestaande uit verschillende partijen, is dit het moment waarop het consortium van partijen uit alle kernteams voor het eerst bij elkaar komt. Partijen krijgen bijv. door de infrastructuur P1 ineens toegang tot nieuwe datasets of delen deze juist ineens met onvoorzien partijen. Randvoorwaarden voor datadelen die voor één PoC acceptabel waren, zijn dat niet noodzakelijk voor de partijen die betrokken zijn bij de infrastructuur P1.

Daarom moet er nu nagedacht worden over acceptabele voorwaarden voor het gebruik van de gezamenlijke datadeel infrastructuur en de beschikbare data zelf. Om de snelheid te bewaren is het nuttig om naar bestaande afsprakenstelsels te kijken zoals IDS en iShare om het wiel niet op nieuw te moeten uitvinden. Bovenal moet er een samenwerkingsvorm gevonden worden waarin snel beslissingen genomen kunnen worden over deze en andere strategische kwesties.

Dit is een voorproefje van het complexe organisatorische vraagstuk wat op de deelnemers af komt bij operationalisatie van een dergelijke infrastructuur. Derhalve moet men ook alvast vastleggen hoe men omgaat met het eventuele vertrekken van een partij uit het consortium. Wanneer partijen niet bereid zijn hun individuele belangen opzij te zetten voor het algemene belang van het datadeel consortium, is dit een situatie die zeer realistisch is.

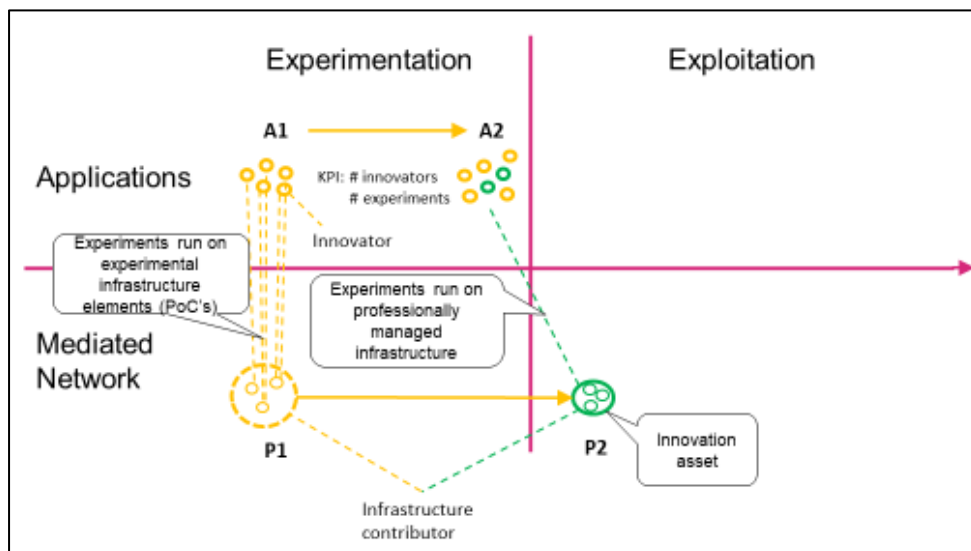
Tot slot is het belangrijk hoe de perceptie van het experimentele infrastructuur is bij de gebruikers. Hun feedback geeft belangrijke aanwijzingen voor de verdere ontwikkeling. Wat werkt goed en wat moet verbeterd worden? Met name hoe de bereidheid van het delen van data verder vergroot kan worden is een belangrijk aandachtspunt voor het latere succes van een dergelijke infrastructuur. Tevens is het nuttig om helderheid te krijgen in de ervaring van 'Infrastructuur Contributors' en hoe zij hun rol in een toekomstig gemedieerd datadeel netwerk zien. Dit is al een eerste aanzet voor de volgende fase. Tabel 2 geeft een overzicht van strategische keuzes zoals hierboven besproken, ter inspiratie voor het toekomstige consortium.

Tabel 2 Niet-uitputtend overzicht van strategische vragen die het consortium zich zou moeten stellen bij het opzetten van een experimenteel datadeel infrastructuur.

Technisch	Organisatorisch	Ecosysteem
Hoe ziet een 'minimal viable infrastructure' er uit?	Hoe word uitwisseling van ervaringen tussen verschillende PoC's gewaarborgd? (bijv. zelfde ontwikkelaars inzetten, regelmatig overleg, etc.)	Wie zijn de gebruikers en wat is hun perceptie van het datadeel infrastructuur P1?
Welke technische componenten uit de PoC's kunnen gecombineerd worden?	Onder welke voorwaarden mag er gebruik gemaakt worden van P1 en de data?	Welke partijen zijn nemen de rol van 'infrastructure contributor' in? Hoe ervaren zij dit? Hoe zien zij dit in de toekomst?
Welke componenten missen nog?	Hoe gaat men om met het uitstappen van een partij uit het consortium?	Hoe kan de bereidheid om data te delen over deze infrastructuur vergroot worden?

3.4 Operationeel datadelen voor experimentele AI applicaties

Wanneer de voorgaande experimentele fase ruim over de helft is, moeten partijen na gaan denken over het vervolg, de eerste stap richting operationalisatie. Hiervoor nemen deelnemers actief deel aan gezamenlijke sessies om een roadmap op te stellen voor een exploitbaar datadeel infrastructuur "P2" (zie Figuur 4). Strategische onderwerpen en activiteiten die aan bod moeten komen bij de uitvoering worden in deze roadmap verder uitgewerkt. In deze paragraaf geven we voorbeelden van dergelijke strategische vragen (zie ook Tabel 3), maar de exacte invulling hangt af van de ervaringen en keuzes die gedaan zijn in de voorgaande fases.



Figuur 4: Op basis van de ervaringen opgedaan met A1 en P1 wordt een roadmap opgesteld voor het exploitbaar datadeel infrastructuur P2. Ook nieuwe functionaliteiten benodigd voor het klaarstomen van experimentele AI toepassingen (A2, geel) voor operationeel gebruik (A2, groen) worden meegenomen.

Het doel van deze operationele infrastructuur is om een datadeel infrastructuur ter beschikking te stellen voor derde partijen om te experimenteren met AI applicaties. Hierbij gaan we ervan uit dat dit een noodzakelijke tussenstap is om tot een gemedieerd datadeel netwerk te komen voor AI applicaties in productie, omdat hierbij nog veel strengere eisen komen kijken. Het is voor de hand liggend om in deze fase gebruik te maken van standaarden en best practices voor de opschaling naar P3.

Uiteraard wordt P2 gebaseerd op ervaringen die opgedaan zijn met P1. Technisch gezien komen er bij een operationele infrastructuur meer eisen kijken omdat het betrouwbaar moet zijn voor haar eindgebruikers. Niet alleen moet het helder zijn hoe de architectuur opgezet en de infrastructuur onderhouden wordt, maar bijv. ook hoe de operationele ontwikkel en dienstprocessen voor release en deployment geregeld worden. Gebruikers zullen verwachtingen hebben aan de stabiliteit en performance van de services (vastgelegd in SLA's), en aanvullende services zoals een helpdesk.

Een belangrijke keuze in dit stadium is die hoe het gemedieerde datadeel netwerk in de toekomst zal worden gecoördineerd. Deze kan bijv. door een aparte stichting worden overgenomen, met een operationeel bestuur en een adviesraad. Belangrijk is vast te leggen hoe deze opgezet wordt en hoe de governance eruit ziet. De verantwoordelijkheden, rechten en plichten van de deelnemende partijen hieromtrent moeten goed vastgelegd worden, waarvoor zowel intern als mogelijk ook extern juridisch expertise ingeschakeld zal moeten worden. De stichting kan besluiten operationele taken zoals die van administrator en technische coördinatie uit te besteden aan derde partijen. Hierbij kan geleerd worden van de governance structuur zoals die inmiddels is vormgegeven in vergelijkbare datadeel initiatieven, zoals door de IDS Association (voor IDS ontwikkeling en adoptie) en iShareworks (voor iSHARE ontwikkeling en adoptie).

Bij al deze punten moeten individuele betrokken partijen afwegen welke rol zij hierin willen spelen, wat hun doelen zijn en welke investeringen zij bereid zijn te doen. De afgevaardigde die (tot nu toe) betrokken is geweest bij het datadelen initiatief zal wellicht hiervoor intern bij zijn eigen organisatie commitment moeten halen van het hoger management. De impact van betrokkenheid bij deze fase zal ingrijpender zijn dan die in de voorgaande fase, waar experimenten nog los van operationele bedrijfsprocessen stonden.

Partijen moeten zich afvragen welke rol zij voor zichzelf zien. Willen zij hoofdzakelijk experimenteren met het ontwikkelen van AI toepassingen (Innovator voor A2 applicaties), willen zij inhoudelijke elementen beschikbaar stellen aan het datadeel netwerk (Infrastructure Contributor) of willen zij (ook) zeggenschap hebben over de coördinatie daarvan? (zie Figuur 4) Het kan verhelderend zijn om een aantal toekomstscenario's uit te werken waarin partijen verschillende rollen innemen om te zien hoe de machtsverhoudingen zullen liggen. Uiteindelijk wil men een scenario dat voor alle partijen acceptabel is, om bijv. een vendor lock-in te voorkomen.

Ook moet de doelgroep voor het nieuwe exploitierbare data deel netwerk bepaald worden. Uiteindelijk moet opschaling van gebruikers er voor zorgen dat het datadelen en het experimenteren daarmee t.b.v. AI toepassingen een succes wordt. Hiervoor moet men in een vroeg stadium al nadenken over de waardepropositie en het business model van dit gemedieerd datadeel netwerk en hoe potentiële gebruikers te bereiken zijn om adoptie van het netwerk te vergroten.

Dit is ook het moment om vast te leggen op welke manier inkomsten met de infrastructuur gegenereerd worden en hoe kosten en baten onderling in het ecosysteem verdeeld worden. Dit is een complex multi-stakeholder vraagstuk waarbij expertise benodigd zal zijn om tot een passende oplossing te komen.

Anderzijds is het belangrijk om, in het kader van veilig en betrouwbaar datadelen, goed vast te leggen aan welke voorwaarden partijen moeten voldoen om toegang te verkrijgen tot het

datadeel netwerk. Hoe omgegaan wordt met het uitstappen is in dat opzicht een net zo belangrijk vraagstuk met strategische consequenties voor de infrastructuur.

In Tabel 3 is een (niet volledige) opsomming van mogelijke strategische vraagstukken die aan de orde zullen komen. Welke keuzes precies relevant zijn voor deze fase moet samen in het consortium besloten worden.

Tabel 3 Niet-uitputtende lijst van strategische vragen die beantwoord moeten worden voor he operationeel datadelen (P2) voor experimentele AI toepassingen (A2).

Technisch	Organisatorisch	Ecosysteem
Hoe wordt de datadeel infrastructuur opgezet en in stand gehouden? Welke standaarden worden toegepast?	Operationeel bestuur en administrator: wie nemen deze rollen in? Voor hoe lang? Hoe worden afgevaardigden gekozen?	Voor welke doelgroep is het gemedieerde datadeel netwerk? Over welke kanalen kunnen zij bereikt worden?
Hoe wordt (technisch) toegang verschaft tot de datadeel infrastructuur? Hoe tot de data?	Hoe word het cross-organisatorisch delen van resources onderling wettelijk geregeld?	Hoe wordt het datadeel netwerk en geïntegreerde functionaliteiten onder de aandacht gebracht bij de doelgroep?
Waar wordt toezicht op gehouden op de infrastructuur? Hoe wordt dit toezicht gegarandeerd?	Welk budget en resources zijn benodigd om het gemedieerde datadeel netwerk te coördineren?	Welke partijen mogen gebruik maken van de datadeel infrastructuur? Aan welke voorwaarden of standaarden moeten zij voldoen?
Hoe wordt het release/ deployment management geregeld?	Hoe wordt de infrastructuur gefinancierd? Wie overneemt de budget verantwoordelijkheid?	Hoe kunnen partijen gebruik maken van het datadeel infrastructuur? Wat mogen zij wel en niet met de data doen? Hoe kunnen zij op een betrouwbare manier hun eigen data ter beschikking stellen?
Welke testing tools of andere services worden geïntegreerd met het datadeel infrastructuur en hoe?	Wat moeten partijen intern bij hun eigen organisatie regelen? Welk commitment moeten ze verkrijgen? Wat is de impact op de interne business?	Hoe is het off-boarding van partijen geregeld? Wat gebeurt er met de data en AI applicaties die afhankelijk zijn van een partij wanneer deze geen deel meer wil uitmaken van het gemedieerde datadeel netwerk?
Wordt er een helpdesk aangeboden? Aan welke Service Level	Welke contracten moeten worden opgesteld? Waar wordt	Hoe gaat het verdienmodel van de datadeel infrastructuur eruit zien? Hoe wordt er

Agreements (SLA) moet worden voldaan?	juridisch expertise vandaan gehaald?	voldoende aanbod van data gecreëerd?
Is een manual beschikbaar en zo ja, op welke manier?

3.5 Roadmap naar een sterk AI landschap in Nederland

Uiteindelijk is het doel om een sterk AI-landschap te creëren, waarin organisaties door samenwerking impactvolle AI-toepassingen op de markt brengen. Dit is weergegeven in Figuur 1 met de applicaties A3 (rechtsboven). Hiervoor zal een operationele datadeel infrastructuur nodig zijn (P3, rechtsonder), die betrouwbaar genoeg is om commerciële AI toepassingen hier van afhankelijk te laten zijn. Technisch gezien zullen er hierdoor hogere eisen gesteld worden aan de infrastructuur wat betreft stabiliteit, performance en veiligheid.

O.b.v. ervaringen met de ontwikkeling van experimentele AI-applicaties (A2) op P2 en datadeel behoeftes voor exploitierbare AI applicaties A3 wordt een architectuur voor een datadeel infrastructuur P3 opgesteld. De functionaliteiten die gerealiseerd worden bepalen in grote mate de aantrekkingskracht van het gemedieerde datadeel netwerk op de eindgebruikers. Aan de groei en opschaling van dit netwerk moet voldoende aandacht worden besteed om te waarborgen dat er waarde gecreëerd wordt door het beschikbaar maken en delen van data.

Aangekomen bij deze volledig operationele fase gaan we ervan uit dat men voldoende ervaring opgedaan heeft met het professioneel opzetten en runnen van een datadeel infrastructuur in een multi-stakeholder setting. De gedeelde kennisopbouw en uitwisseling van ervaringen opgebouwd in het voorgaande traject zijn hiervoor onontbeerlijk. Wat betreft governance en organisatie zal men voor de infrastructuur P3 zich in grote mate baseren op wat wel en niet werkte bij P2. De complicatie dat de AI-toepassingen draaiend op de infrastructuur nu in productie worden genomen zal er voor zorgen dat er aanvullende afspraken gemaakt en vastgelegd moeten worden.

Ter inspiratie zijn in

enkele strategische aandachtspunten genoemd, die in deze fase aan bod komen. Ervan uitgaand dat vele punten uit Tabel 3 hier wellicht ook relevant zullen zijn.

Tabel 4 Overzicht van enkele strategische vraagstukken die in de exploitatie-fase (operationele datadeel infrastructuur AI toepassingen in productie) aan bod komen.

Technisch	Organisatorisch	Ecosysteem
Welke extra functionaliteiten zijn benodigd wanneer de data voor AI toepassingen in productie gebruikt worden?	Hoe wordt P3 ingericht en georganiseerd t.o.v. P2?	Op welke behoeftes van A2 Innovators en A3 product exploiters wordt ingespeeld met P3?
Welke technische maatregelen zijn nodig om de stabiliteit en veiligheid van de infrastructuur te garanderen?	Welke onderwerpen moeten extra vastgelegd worden in contracten?	Hoe wordt voor verdere opschaling en groei gezorgd? (data en AI toepassingen?)

4 Randvoorwaarden

Dit document geeft een indruk van wat er op organisaties afkomt qua activiteiten en beslissingen bij een traject via first-time-engineering naar operationalisatie van datadelen ten behoeve van AI. Het is nuttig om vooraf helder te hebben wat van hen verwacht wordt zodat zij zich hierop in kunnen stellen. Deelnemende organisaties zullen bijv. voldoende resources ter beschikking moeten stellen, hetzij in de vorm van ontwikkelaars en managers, hetzij in de vorm van hard- of software. De mate van commitment van de deelnemende bedrijven bepaalt tevens het succes van het traject richting operationalisatie en exploitatie.

Om tot een generiek en gemedieerd datadeel netwerk in Nederland te komen moeten partijen in ongewone combinaties met elkaar samenwerken in een consortium. Dit houdt in dat zij niet alleen een voor hen nieuwe vorm van samenwerking moeten ontdekken, maar tevens gezamenlijk belangrijke beslissingen zullen moeten nemen over de strategie en de voortgang van de datadeel infrastructuur. Leden van de NL AIC moeten zich bewust zijn dat dit een lastig proces is dat wellicht 'buiten hun comfortzone' valt. Soms zullen bijv. individuele belangen van stakeholders ondergeschikt zijn aan het gezamenlijk belang. Hoe zij hier eventueel mee omgaan, kunnen partijen het best van tevoren overwegen.

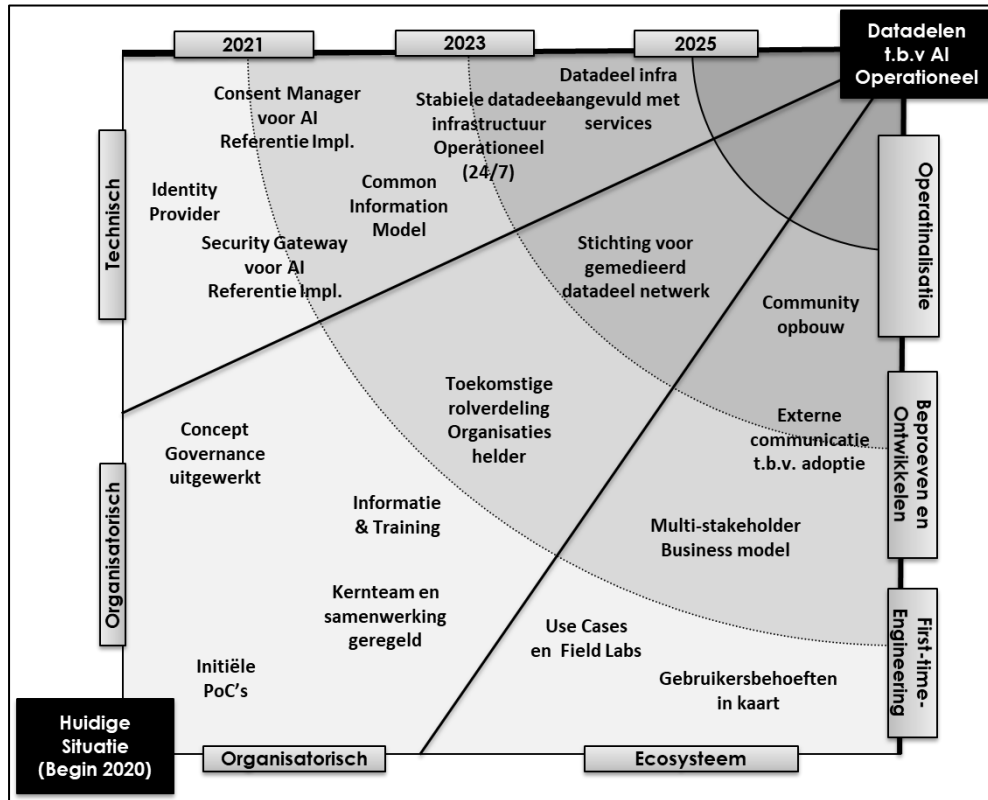
Men moet bewust zijn van het feit dat, zolang de activiteiten nog op de experimentele fase van toepassing zijn (PoC's en experimentele AI toepassingen) de voortgang gebaseerd is op vrijwillige bijdrages van communityleden. Dit houdt in dat men niet alleen geregeld samen moet komen in het consortium voor besprekingen en uitvoerend werk, maar tevens in de eigen interne organisatie werk zal moeten verrichten. Bijv. het regelen van commitment van het hoger management in de eigen organisatie voor de inzet van resources, het opvragen van informatie bij technische en legal afdeling, etc.. Zodra de overgang gemaakt wordt naar operationalisatie en exploitatie, moeten bestaande bedrijfsprocessen gefuseerd worden met de nieuwe, wat van grote impact zal zijn op de individuele bedrijfsstrategie. Het is van belang om dit bewustzijn en de betrokkenheid van interne stakeholders zo vroeg mogelijk in gang te zetten.

En zoals aangegeven in hoofdstuk 2, is een belangrijk aspect hierbij ook om afstemming te houden met (inter-)nationale ontwikkelingen en initiatieven rondom datadelen, zoals de Data Sharing Coalition, iSHARE, IDS en GAIA-X.

Het pad naar een sterk AI-landschap en een deugdelijke datadeel infrastructuur zal dynamisch zijn en vraagt inzet en betrokkenheid van alle deelnemers. Men moet zich bewust zijn dat innovatie een proces is dat met vallen en opstaan gepaard gaat. Als organisatie maakt men deel uit van een groter geheel. Bedrijven die dit avontuur aan durven te gaan staan echter ook veel nieuwe kansen te wachten.

5 Roadmap

Op basis van het traject van first-time-engineering naar operationalisatie zoals geschetst in deze rapportage is een eerste versie van de ontwikkel roadmap voor datadelen ten behoeve van AI grafisch weergegeven in Figuur 5. In lijn met de aanpak en tabellen in hoofdstuk 3 categoriseert het de ontwikkelingsactiviteiten in 'technisch', 'organisatorisch' en 'ecosysteem'.



Figuur 5: Ontwikkelroadmap datadelen voor AI.

Zoals de figuur laat zien wordt voor het doorlopen van het ontwikkelproces van first-time-engineering naar operationalisatie (het traject van P1 naar P3) een periode van 4 tot 5 jaar voorzien. Afhankelijk van investeringen, scope keuzes en complexiteit van de implementaties zal deze periode variëren.

Als eerste vervolgstap voor 2020 zullen de initiële PoC's worden opgezet en uitgevoerd als eerste stap van een experimentele infrastructuur. De opzet en aanpak hiervoor is beschreven in [1].

6 Referenties

- [1] Nederlandse AI Coalitie (NL AIC) Werkgroep Datadelen. 'Datadelen voor AI - Initiële verkenning door de werkgroep 'datadelen' ten behoeve van de Nederlandse AI'
- [2] L. van Piggelen, A. Zandee, T. Vreuls, Douwe van de Ruit (KPN), Mischa Vos (Rabobank), Frank Berkers (TNO), Oskar van Deventer (TNO, editor) "Techruption Consortium Blockchain – blueprint for a Dutch consortium blockchain" Techruption (2018).
- [3] Frank Berkers, Laura van den Aarssen, M. Oskar van Deventer, "A strategic business options model for consortium blockchain
-

Auteurs

L.G. (Laura) van den Aarssen, PHD
F.T.H.M. (Frank) Berkers, MSc
H.J.M. (Harrie) Bastiaansen, PHD
F.T.A. (Frans) van Ette, MSc

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook, zonder voorafgaande toestemming van de NL AI Coalitie.