

Architectuurnotitie

WPI model onderzoekswaardigheid

uitkeringsaanvragen

Inhoud

1 Inleiding	1
1.1 Relevante documenten	2
1.2 Raakvlakken	2
1.3 Gerelateerde projecten	2
1.4 Belanghebbenden	3
2 Situatieschets	4
2.1 Context	4
2.2 Doelen	4
2.3 Betrokken systemen	4
2.4 Knelpunten	4
2.5 Toekomstige situatie	5
2.6 Informatiebehoefte	5
3 Kaders	6
3.1 Beleid Gemeente Amsterdam	6
3.2 Uitgangspunten	6
3.3 Architectuurprincipes	7
3.4 Risico's	8
4 Eindsituatie	9
4.1 Doelarchitectuur	9
4.1.1 Toelichting doelarchitectuur	10
4.1.2 Scheiding van rollen	11
4.1.3 Architectuurprincipes	12
4.2 Knelpunten	12
4.3 Tijdelijke situatie	13
4.4 Bijlagen	17

1 Inleiding

In het kader van Informatiegestuurd Handhaven bij WPI wordt gewerkt aan een algoritme waarmee data uit verschillende systemen binnen WPI wordt gecombineerd om te helpen beslissen welke aanvragen Levensonderhoud (Participatiewet) door een Inkomensconsulent worden opgepakt voor de uitvoering van processtap bepalen recht en welke door Handhaving worden opgepakt voor de uitvoering van processtap controle op bijstandsgerechtigheid van de aanvrager.

Het model wordt ontwikkeld door:

- Team Advanced Analytics van WPI

In samenwerking met:

- De werkgroep Informatiegestuurd Handhaving van WPI

Dit document beschrijft het ontwerp van en de afspraken over de informatiearchitectuur.

1.1 Relevante documenten

Titel	Documentnaam, versie en status	Bron
Participatiewet	https://wetten.overheid.nl/BWBR0015703/2022-01-01	Rijksoverheid

1.2 Raakvlakken

De belangrijkste projecten die dit project beïnvloeden, of andersom, zijn:

1. Veranderingen in het systeemlandschap van WPI: Er speelt veel rondom de systemen die WPI gebruikt qua uitfasering/vervanging/herinrichting. Dit geldt zowel voor de systemen die de inputdata voor het model bevatten als het applicatielandschap in zijn geheel.
2. Reorganisatie van WPI en het I-domein.
3. De migratie van gemeentelijke applicaties naar Azure: Het model is een van de eerste data-applicaties die op Azure zal gaan draaien en is zodoende ook een test case. Aangezien nog niet alle Azure-componenten volledig af zijn, beschrijven we op sommige plekken twee architectuursituaties: de beoogde lange-termijnsituatie en de voorlopige tussenoplossing.
4. Dataverkenner voor WPI: De beoogde applicatie om uitkomsten van het model in het proces te laten landen is de dataverkenner voor WPI, die in ontwikkeling is.
5. Publicatie van het algoritme in het publieke gemeentelijke algoritme register:
<https://algoritmeregister.amsterdam.nl/>

1.3 Gerelateerde projecten

1. Project Dataverkenner
2. Project SVWI
3. Datateam Sociaal (DASO)
4. Analyse Services

1.4 Belanghebbenden

Naam	Rol	Opmerking
Wethouder	Politiek verantwoordelijk	
Directeur Inkomen	Gemandateerd directeur	
Epic Owner	Business verantwoordelijke	
Business Product Owner	Producteigenaar	
Product Owner	Producteigenaar	
Team Advanced Analytics WPI	Ontwikkelaar en (intern) leverancier model	
Cockpit I lead	I lead opgave Amsterdammers Financieel Gezond	
Cockpit Architect	Architect opgave Amsterdammers Financieel Gezond	
Architect	Architect opgave Amsterdammers Financieel Gezond	
Architect CCC	Architect CCC	
Privacy officer WPI	Privacy officer	
Security officer WPI	Security officer	
Security officer CCC	Security officer	
Team Dataverkenner	Implementatie modeluitkomsten	
Datateam Sociaal	Dataverancier	
Team Analyse Services	Leverancier ontwikkelomgeving	
Basisinformatie	Adviseur gegevensmanagement	

2 Situatieschets

2.1 Context

In opdracht van de gemeenteraad draagt het college zorg voor een rechtmatige uitvoering van de Participatiewet door onder meer de inzet van preventie en controle. Directie Inkomen is verantwoordelijk voor de beoordeling de rechtmatigheid van aanvragen voor levensonderhoud (bijstandsuitkering). Ontvangen aanvragen worden beoordeeld en afhankelijk van de ontvangen aanvraag kan er aanleiding zijn om nader onderzoek te doen, extra informatie op te vragen, en uiteindelijk te verlenen of te weigeren.

Bij de beoordeling van de bijstandsuitkering wordt expertise van de afdeling Toezicht en Handhaving ingezet. Dit betekent dat de afdeling Handhaving Werk en Inkomen (HWI) een aanvraag ook bekijkt als daar een aanleiding/ een signaal voor is. Voor bijstandsaanvragen geldt dat een dergelijk signaal wordt gegeven door de verantwoordelijke inkomens consulent (IC) die de aanvraag in behandeling heeft. Dit is momenteel grotendeels een handmatig proces. De handhavingsmedewerker en de IC bespreken de aanvraag middels een consultgesprek, waarna de IC vervolgens verder gaat met de beoordelen van de aanvraag. De handhavingsmedewerker geeft enkel een advies met betrekking tot de beoordeling vanuit diens expertise bij de aanvraag. Het uitvoeren van een nader onderzoek draagt bij aan het voorkomen van bijstandsschulden als gevolg van een ten onrechte verstrekte uitkering ontstaan en dat maatschappelijk geld wordt verstrekt aan mensen die geen recht hebben op een uitkering.

2.2 Doelen

Het doel van dit product is het minimaliseren van onrechtmatige verstrekking van bijstand, waardoor we kunnen voorkomen dat er ten onrechte uitkering worden verstrekt en er bijstandsschulden ontstaan bij terugvordering van een onterecht verstrekte uitkering. Daarnaast blijven we ervoor zorgen dat publiek geld wordt ingezet waar nodig en het sociale stelsel betaalbaar blijft.

2.3 Betrokken systemen

Voor het voeden van het model worden gegevens uit de systemen Socrates, Sherlock, Raak en GALO gebruikt. Hierin zit ook een deel doorgeleverde data uit de BRP. Het data warehouse van WPI is betrokken als leverancier van de gegevens uit deze bronsystemen.

De resultaten van het model zullen worden gebruikt in de Dataverkenner applicatie voor WPI.

2.4 Knelpunten

We willen onze wettelijke taak meer uitvoeren in lijn met de visie van WPI: efficiënter, effectiever, proportioneeler, transparanter, preventief. Daarnaast zijn er nog 3 knelpunten gesignaleerd bij het analyseren van de doelen en het huidige proces.

Complex en tijdrovend: Het analyseren en beoordelen van de rechtmatigheid van de aanvraag is een complexe en tijdrovende taak doordat veel informatie uit verschillende bronnen bekeken en vergeleken moet worden. Daarnaast moet de werknemer ondertussen ook alert zijn op mogelijke risico's of (bewuste) onjuistheden of onvolledigheden.

Doorzetten handhaving verminderd: We zien dat aanvragen steeds minder worden doorgezet door IC'ers naar handhaving. Doorzetten is geen doel op zich maar het is wel een signaal dat we mogelijk meer onrechtmatige uitkeringen verstekken. Wij zien hierbij verschillende oorzaken:

- De werkwijze in het primaire proces is veranderd: Er is focus op doelmatigheid en om dit te bereiken wordt er geïnvesteerd in het klantcontact. Hierdoor is er minder focus op rechtmatigheid, zodat het klantcontact niet verstoord wordt met een handhavingsonderzoek en vragen over rechtmatigheid.
- Hierdoor wordt niet iedere klant op dezelfde manier behandeld. Dit is sterk afhankelijk van wie de klant in zijn of haar caseload heeft. Daarnaast is dit ook mogelijk omdat er geen eenduidige werkinstructies zijn.
- De inzet van Handhaving is afgelopen jaren proportioneler geworden. Dit heeft het imago van Handhaving aangetast, de samenwerking met het primaire proces is hierdoor verminderd. Sommige IC'ers denken dat de handhaving gaat niet meer op huisbezoek gaat, dus doorzetten geen zin heeft.

Inefficiënt en niet optimaal effectief werkproces: Er is onduidelijkheid over de rollen van de inkomensconsulent en handhaving en de scheiding daartussen. Hierdoor ontstaat het uitvoeren van dubbel werk. Zowel het proces als de systemen ondersteunen hier niet in. Dit leidt tot een hogere doorlooptijd.

Er is geen eenduidige en standaard werkwijze waarop een aanvraag wordt gescreend of overgedragen wordt. Dit kan leiden tot een niet gelijke aanpak van elke aanvraag en daarmee een lagere effectiviteit.

Het reguliere werkproces ondersteunt niet bij het doel van Handhaving om onrechtmatige instroom preventief te voorkomen.

2.5 Toekomstige situatie

Het afgelopen jaar is gewerkt aan een voorspellend model, een Explainable Boosting Machine algoritme, wat het mogelijk maakt om geautomatiseerd de taak om te bepalen of een aanvraag al dan niet onderzoekwaardig is kan ondersteunen. Dit model wordt toegevoegd aan het werkproces en dient als extra filter voor onderzoekwaardige aanvragen. Een medewerker kan vervolgens mede op basis van de voorspelling van het model bepalen welke vragen al dan niet nader worden onderzocht.

Het huidige processen en de toevoeging van het model (rood) aan het proces staan hier versimpeld weergegeven. Het model wordt als extra filter ingezet op basis waarvan aanvragen direct door de medewerker vooronderzoek (MV) worden opgepakt ipv eerst door de IC. De MV beslist dan of de aanvraag onderzoekwaardig is of niet. Zo ja, dan wordt de aanvraag doorgezet naar de HH-Specialist. Zo niet, dan wordt het doorgezet naar de IC.



Voor de technische toekomstige situatie, zie de solution outline van de doelarchitectuur. Het nieuwe proces is weergegeven in bijlage 1.

2.6 Informatiebehoefte

In onderstaande afbeelding is de informatiebehoefte van de eindgebruikers te zien. Het is tevens ook het voorbeeld van het gewenste eindproduct voor WPI. In dit overzicht is te zien dat het gaat om gegevens van de aanvragers en de score op onderzoekwaardigheid van de aanvraag. Vervolgens worden de

belangrijkste 'features' ofwel variabele getoond bij de risico indicatie met een toelichting wat dit betekent.

LO aanvraag

Administratienummer: 00000000

Hieronder zie je de drie belangrijkste datapunten die hebben geleid tot de uitkomst van het algoritme. Deze staan op volgorde van relevantie. Een + betekent dat het datapunt de aanvraag meer onderzoekswaardig maakt; een - betekent dat het datapunt de aanvraag minder onderzoekswaardig maakt.

Let op: Deze uitkomst geeft je richting voor je onderzoek, maar is geen absolute waarheid: gebruik altijd je eigen expertise om een aanvraag te behandelen.

Belangrijkste datapunten

5 Dagen sinds verhuizing +	Aantal dagen sinds de aanvrager voor het laatst verhuisd is. Als er geen adres bekend is of alleen een postadres, dan wordt deze feature zo gevuld alsof de aanvrager heel lang geleden voor het laatste verhuisd is.
60% Percentage maatregel +	Een feit kan leiden tot een maatregel: een korting op de uitkering. Deze feature geeft het gemiddelde kortingspercentage aan van alle feiten in het jaar voorafgaand aan onderzoek. Dit geeft een indicatie van hoe ernstig de begane feiten waren.
Ja Eerder levensonderhoud	Is nul wanneer er geen maatregelen waren in het jaar voor de aanvraag of wanneer de aanvrager in het jaar voor de aanvraag niet in onze systemen voorkomt.
Ja Eerder levensonderhoud	Heeft klant in het jaar voorafgaand aan de aanvraag hetzelfde product aangevraagd dat hij/zij nu aanvraagt?

3 Kaders

3.1 Beleid Gemeente Amsterdam

In deze paragraaf staan architectuurprincipes van de Gemeente Amsterdam die van toepassing zijn op het onderwerp.

Het algoritme wordt gepubliceerd op het algoritmeregister. Algoritmes mogen niet discrimineren. Het lerend vermogen van het algoritme wordt altijd door een mens beoordeeld, om te voorkomen dat het algoritme discriminerende factoren benadrukt (zie Belastingdienst met etniciteit en voorkomen).

Streef er naar de TADA waarden te hanteren: de principes van het "TADA manifest" zijn te vinden op de website van TADA: <https://tada.city/over-ons/>
Deze waarden gaan vooral over transparantie van gegevens die een gemeente gebruikt.

De taken en verantwoordelijkheden bij de ontwikkeling van algoritmes zijn beschreven door de CIO-office in het document 'taken en verantwoordelijkheden bij het gebruik van algoritmen door de gemeente Amsterdam'. Dit document is op 20 januari 2022 vastgesteld door het GMT.

3.2 Uitgangspunten

Architectuur: alle WPI modellen worden op dezelfde manier vormgegeven, zodat de ontwikkelingen gelijksoortig zijn en de resultaten gelijksoortig zijn en dus gelijksoortig ingezet kunnen worden in de toolset van WPI. Dit zal voor alle betrokken partijen maximale efficiency opleveren.

Gebruikmaken van gestandaardiseerd aanvraagmodel:

- GBI standaard (BPMN) voor aanvraag beslisboom en gegevenselementen waar check op plaatst kan vinden.
- WPI gegevensmodel wat ontwikkeld wordt in team 8 en 10 uit de opgavetrein AFG, datateam sociaal.
- Bij het gebruik van API's moet er voldaan worden aan API de standaarden:
<https://www.gemmaonline.nl/index.php/API-standaarden>

3.3 Architectuurprincipes

Vanuit de centrale architectuurprincipes van de gemeente Amsterdam (versie 1.0, 24 sept 2021) zijn de volgende principes van toepassing, met daarbij in 1 zin waarom het principe voor het WPI model van belang is. Details zijn per principe terug te vinden in het document Architectuurprincipes v1.0.pdf:

1. Principe 1: Stel de BOB en medewerker centraal
De rechten van de bewoner staat centraal bij het beoordelen van de aanvraag. De IC medewerker hoeft minder controles uit te voeren en kan vertrouwen op het juist uitvoeren van het algoritme.
2. Principe 2: Borg de kwaliteit van (digitale) producten en diensten
De kwaliteit van het aanvraagproces en recht bepalen wordt door het algoritme product uniform en transparant gewaarborgd.
3. Principe 3: Standaardiseer waar het kan
Wat we doen voor het ene model doen we ook voor een eventueel volgend model. Dat is geheel aansluitend bij de in de vorige paragraaf genoemde doelstelling.
4. Principe 4: Stel digitale oplossingen samen op basis van componenten
Maak alle onderdelen die nodig zijn, modulair, dus zorg dat ze los van elkaar kunnen bestaan.
5. Principe 5: Digitaliseer producten en diensten en de voortbrenging daarvan
Zorg ervoor dat alle tot nu toe handmatige onderzoeken, controles en het uitprinten en op straat meenemen van een veelheid aan documenten, niet meer nodig is.
6. Principe 6: Beschouw data als waardevol
Neem afdoende maatregelen om de gebruikte data veilig in te zetten en gebruiken.
7. Principe 7: Wees open
Algoritmes worden gepubliceerd op Amsterdam.nl.

Vanuit de ontwikkelingen binnen WPI zijn de volgende principes van belang:

0. **Service georiënteerd:** opzet van de oplossing zoveel als mogelijk in de vorm van (micro)services die ieder voor zich de potentie hebben om hergebruikt te kunnen worden.
1. **Digitaal werken** waar mogelijk. Dit komt voort uit onder meer de principes rondom de "Moderne Overheid – digitale transformatie in Amsterdam".
2. **Data is altijd ontkoppeld van applicaties.** Data wordt altijd bij de bron opgehaald, de applicatie maakt dit mogelijk doordat data via API's ontsloten kan worden.
3. **Data eenmalig opslaan**
Ieder gegeven wordt eenmalig opgeslagen en van daaruit geraadpleegd of onderhouden. Dit betekent dat gegevens worden gebruikt, vanuit de ene bron waarin ze bestaan. Als het wettelijk noodzakelijk is gegevens toch in een eigen dossier te bewaren, dan dient geborgd te worden dat die gegevens niet aanpasbaar zijn, zodat de integriteit van de gegevens bewaard wordt.

4. **Privacy** richtlijnen volgen en dus data alleen gebruiken vanuit de juiste doelbinding. De doelbinding van een medewerker binnen een bedrijfsproces, bepaalt of data benaderd mag worden en hoe deze gebruikt mag worden.
5. **Systemen hebben discrete en afgebakende functies. Net zoals bij mensen wordt er functiescheiding (vier ogen principe) toegepast waar dit nodig is.**
Er is altijd menselijke interactie op het model mogelijk: op alle momenten, moet het mogelijk zijn via een toegankelijke user interface en begrijpelijke rapportage om het algoritme te monitoren zodat het voor een mens te controleren is. Hoe we hier precies invulling aan gaan geven en wat er mogelijk is met Azure gaan we volgende PI verder uitzoeken.
6. **Informatiemodellen en gegevensuitwisseling zijn gestandaardiseerd.**
Werk met standaarden die gangbaar zijn in de gemeente en die voldoen aan de uitgangspunten en principes. Wat gedaan wordt als inrichting voor het WPI-model zal een voorbeeld zijn voor eventuele volgende modellen.

3.4 Risico's

Architectuur:

1. Doelarchitectuur is nog niet definitief.

Privacy:

Voor een overzicht van de privacy risico's en de getroffen maatregelen, zie de DPIA. Daarnaast is er nog een specifiek infrastructuur privacy risico dat wij hieronder beschrijven.

1. Er moet nog een risicoanalyse worden uitgevoerd en de verwerkersovereenkomst met Azure moet worden gefinaliseerd voor het Azure Platform¹⁰. Deze worden in de loop van Q3 afgerond. Idealiter waren deze documenten al opgesteld. Uit de DTIA kunnen namelijk nog onbekende risico's worden geïdentificeerd. Desondanks is door de eindverantwoordelijke van het Azure Platform besloten dat het platform kan worden gebruikt door gemeente, omdat de gemeente wel in controle is op dit onderwerp. Dat heeft er mee te maken dat de mogelijk nieuw geïdentificeerde risico's grotendeels kunnen worden gemitigeerd. Als echter toch wordt geconstateerd dat veilige gegevensverwerking niet kan worden gegarandeerd doordat geen mitigerende maatregelen kunnen worden getroffen, dan zal dit direct worden gecommuniceerd naar de betrokken directeuren. De kans hierop wordt als klein beschouwd.

Security:

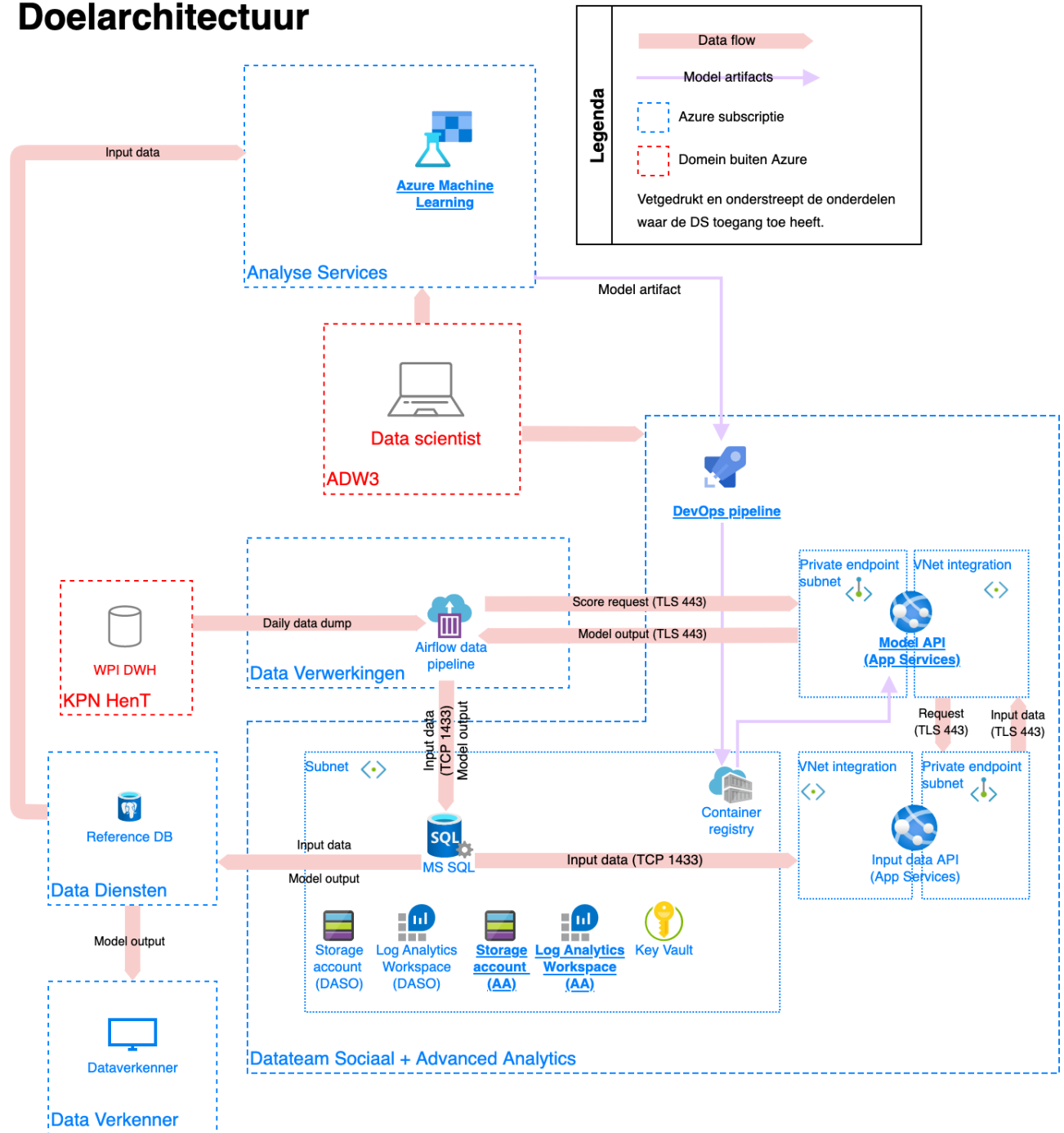
1. Eerder werd voor dit product de VAO (Veilige Analyse Omgeving) gebruikt. Nu gebruiken we, zoals beschreven staat in de doelarchitectuur de Analyse Service op Azure. Het handmatige proces om data van de VAO naar de Azure ML omgeving te verplaatsen is kwetsbaar en had mogelijk kunnen leiden tot gegevensverlies en daarmee imagoschade voor de gemeente Amsterdam. Er zijn een aantal maatregelen genomen (werken in ADW-omgeving persoonlijk account, aggregatie van data, hashing, verwijdering data en beperkte toegang) om dit proces zo veilig mogelijk te laten verlopen. De kans op gegevensverlies is klein. Door de getroffen maatregelen is de impact ook minimaal, omdat de data onherkenbaar is gemaakt en praktisch niet te 'hacken'¹¹. De migratie heeft inmiddels succesvol plaatsgevonden en de VAO wordt niet meer gebruikt. Meer informatie hiervoor is te vinden in het document 'Dataketen migreren data naar Azure'.
2. Data bevindt zich tijdelijk in zowel de VAO als in de Azure omgeving. De data voor de ontwikkeling van het model zal tijdelijk zowel in de VAO als in de Azure omgeving staan. De

Azure omgeving is nieuw voor de developers. Ze zullen het al gemaakte model werkend moeten krijgen op deze nieuwe omgeving en het verder gaan doorontwikkelen. Hierbij kunnen onbedoeld fouten worden gemaakt. Door de gehashte data beschikbaar te hebben in de VAO, is er een vangnet om eventuele fouten te herstellen. De VAO die gebruikt werd staat nu op de planning om ontmanteld te worden in Q3 2022. Dan wordt de hele omgeving en alle data verwijderd.

4 Eindsituatie

4.1 Doelarchitectuur

Doelarchitectuur



4.1.1 Toelichting doelarchitectuur

De doelarchitectuur bestaat uit twee delen: development (waar het model ontwikkeld wordt) en deployment (waar het model gedraaid wordt). Alle onderdelen draaien op Azure.

Het development-deel wordt verzorgd door Analyse Services. Het deployment-deel wordt verzorgd door team AA in samenwerking met Datateam Sociaal.

Dit document en de bijgevoegde diagrammen beschrijven de functionele samenhang tussen alle componenten die nodig zijn om het model te ontwikkelen en in productie te draaien. Het beschrijft niet de

technische (infrastructurele) details en beveiligingsdetails, omdat veel van deze componenten worden beheerd door andere teams en los zijn gecertificeerd. Voor deze details verwijzen we daarom naar:

- De [wiki](#) van Analyse Services voor het development-deel, of neem contact op met de Product Owner.
- De [wiki](#) van Datateam Sociaal voor het deployment-deel, of neem contact op met de Product Owner.

Development en deployment zijn met elkaar verbonden doordat in Analyse Services een getraind model opgeslagen wordt. Dit bestand wordt gedownload naar Azure DevOps binnen de subscriptie van team AA en Datateam Sociaal (in het diagram: "model artifact"). Dit gebeurt middels een service connection van DASO. De service principal van deze service connection is door AS toegevoegd aan een Entitlement Management groep met de rollen die nodig zijn om de model export uit te voeren.

Nadat het bestand is gedownload, wordt er een Docker image omheen gebouwd met een API waarmee het model kan worden aangeroepen. Deze image wordt opgeslagen in de container registry en gedeployed naar Azure App Services. Hiervoor is team AA verantwoordelijk.

De benodigde data wordt aangevoerd door een data pijplijn, beheerd door Datateam Sociaal. De pijplijn draait op de beheerde Airflow instance van Data Verwerkingen. Deze bestaat uit de volgende stappen, met daarbij aangegeven bij welk team de verantwoordelijkheid ligt:

1. Een dagelijkse, incrementele data dump wordt aangeleverd uit het data warehouse van WPI¹ met alle nieuwe aanvragen van de dag ervoor en de benodigde gegevens om deze aanvragen te beoordelen (in het diagram: "daily data dump"). **[DASO]**
2. De verwerkte data dump wordt door Datateam Sociaal gehasht waar nodig, opgeslagen in een SQL-database en beschikbaar gemaakt via een API (in het diagram: "input data API"). **[DASO]**
3. Zodra dit gebeurd is, wordt in de pijplijn de API van het model aangeroepen om voorspellingen te maken voor de nieuwe aanvragen (in het diagram: "score request"). **[DASO]**
4. Het model haalt de data op uit de genoemde database (in het diagram: uitwisseling tussen "model API" en "input data API"). **[team AA]**
5. De API van het model antwoordt met de voorspellingen, de belangrijkste factoren die eraan hebben bijgedragen en metadata over de voorspelling, zoals wanneer deze gemaakt is en welke versie van het model gebruikt is (in het diagram: "model output"). **[team AA]**
6. De pijplijn schrijft deze informatie terug naar de centrale datahub bij Data Diensten. **[DASO]**

Hiervandaan kan de Dataverkenner de informatie ophalen en tonen aan de medewerkers. Ook kan er in het geval van een vraag of bezwaar door een burger worden terug gekeken hoe een voorspelling tot stand is gekomen.

Alle componenten in de subscriptie van DASO en AA zijn beschikbaar in OTAP-omgevingen.

4.1.2 Scheiding van rollen

Binnen de subscriptie van DASO en team AA is een scheiding van rollen. De data scientist heeft toegang tot de volgende componenten (vetgedrukt en onderstreept weergegeven in het diagram):

- DevOps pipelines: alleen rechten om de deployment pipeline van het model af te trappen.
- Container registry: alleen indirect via de pipeline om de images van de model API te beheren.
- Model API op App Services.

¹ Data warehouse WPI is de afgesproken data bron, totdat in de gewenste situatie de data via API's op de bronapplicatie Socrates kan worden ontsloten conform het principe dat data bij de bron wordt ontsloten.

- Log Analytics Workspace en Storage Account: aparte workspace en achterliggende storage account voor alléén de logs die worden gegenereerd door het model.
- Git repositories die relevant zijn voor het model (niet ingetekend op diagram).

Zie voor meer details en de rechten van DASO de autorisatiematrix.

4.1.3 Architectuurprincipes

Een aantal keuzes draagt bij aan het borgen van de stedelijke architectuurprincipes:

- Door het gebruik van bouwblokken en centrale diensten als Analyse Services, Data Diensten en Data Verwerkingen kan deze architectuur grotendeels hergebruikt worden voor eventuele volgende modellen.
- Door het gebruik van APIs voor de datalevering en het model is data ontkoppeld van de applicatie en kunnen aanpassingen aan beide kanten in veel gevallen worden gemaakt zonder wederzijdse afhankelijkheden.
- Het gestandaardiseerde Amsterdam Schema wordt gebruikt voor het verwerken van de dagelijkse datastroom vanuit het data warehouse.
- Data wordt op één centrale plek opgeslagen door de model output via de Data Hub beschikbaar te maken aan de Data Verkenner.

GBI modellen worden gebruikt voor het gegevensmodel van de aanvraag. De GBI modellen betreffen BPMN en DMN modellen. De modellen van GBI zijn nog in ontwikkeling. Het model voor uitkeringsaanvragen is een kleine schakel binnen het bredere BPMN model. Het team en de architecten houden periodiek de vinger aan de pols om op de hoogte te zijn van wederzijdse ontwikkelingen.

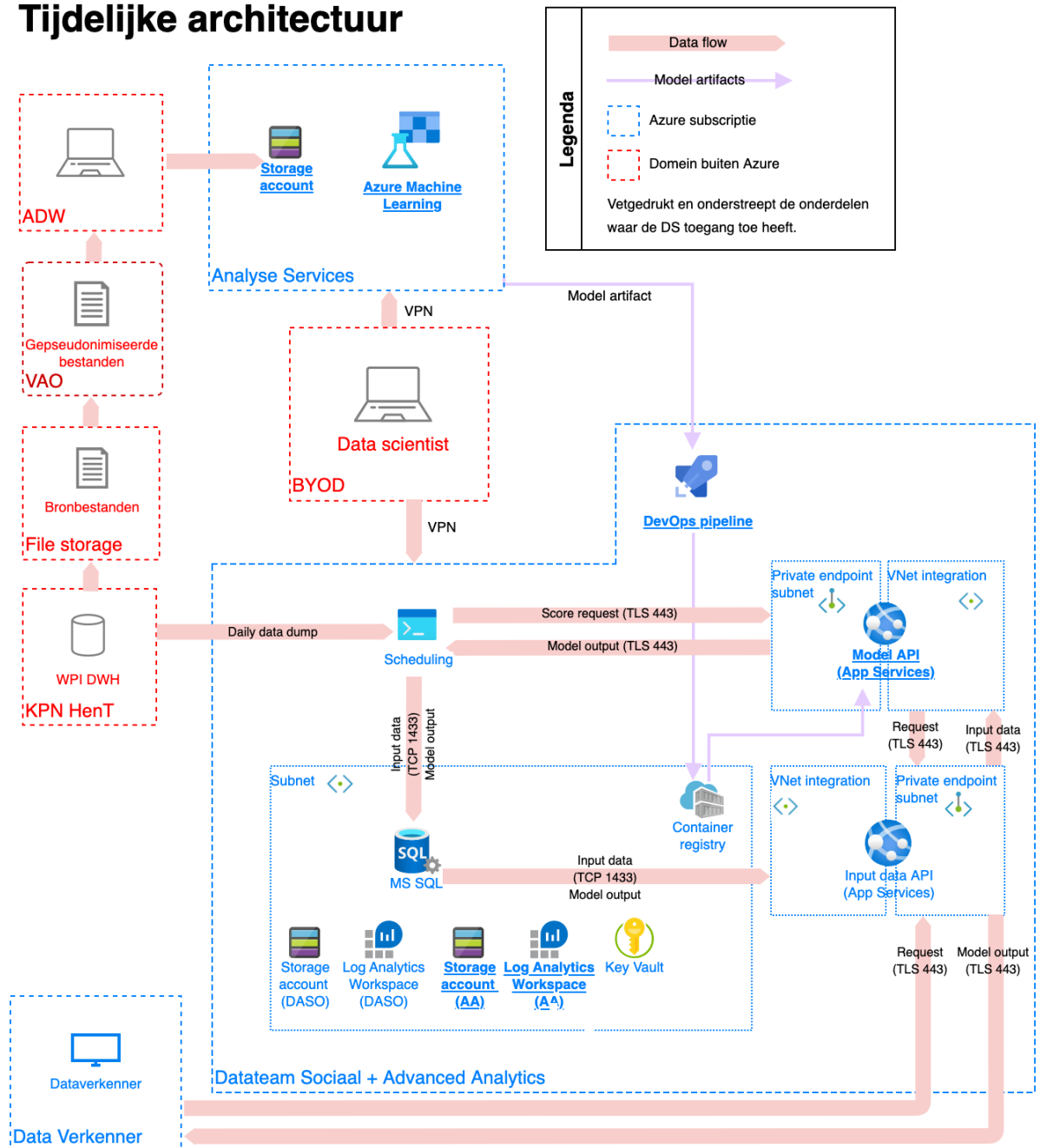
4.2 Knelpunten

Het data platform op Azure is nog niet uitontwikkeld. Het is daarom van belang om goed in de gaten te houden welke functionaliteit wanneer beschikbaar komt en welke functionaliteit we tijdelijk anders moeten oplossen.

Daarnaast kunnen databronnen nog niet ontsloten worden bij de bron. Dit betekent dat er voorlopig dagelijkse dumps van databestanden worden gebruikt. In het kader van dataminimalisatie wordt hiervoor een selectie gemaakt van de personen, kolommen en tijdspanne die nodig zijn om de aanvragen te scoren.

4.3 Tijdelijke situatie

Tijdelijke architectuur



Helaas kan nog niet worden voldaan aan de doelsituatie doordat nog niet alle CCC-bouwblokken gereed zijn. Er is door de productmanager en epic owner besloten dat er toch voortgang gemaakt kan worden in een tijdelijke doelsituatie waarbij gekeken wordt naar mogelijk aanvullende maatregelen.

De tijdelijke situatie wordt niet onder architectuur ingericht. Dit is een exceptie op onder architectuur werken. Periodiek wordt gekeken wat de vordering is richting de doelarchitectuur.

De tijdelijke situatie wijkt op de volgende punten af van de doelarchitectuur:

- Het is nog niet mogelijk om de data voor het ontwikkelen van het model op te halen bij Data Diensten. Daarom wordt tijdelijk gewerkt met een dump in de 'data brievenbus' (een Azure storage account) van team AA op Analyse Services. Deze dump is reeds gepseudonimiseerd in de

Veilige Analyse Omgeving (VAO). De brievenbus is vanuit ADW te bereiken. Vanaf dat punt verloopt het ontwikkelen van het model zoals in de doelarchitectuur. Deze dump bevat meer gegevens dan uiteindelijk in het model gebruikt worden, vanwege het experimentele karakter van het ontwikkelen van machine learning modellen. Dit is hoe wij in 2022 de gepseudonimiseerde data op de Analyse Services hebben gekregen.

- Als we in de toekomst het model opnieuw willen trainen, zullen we bovenstaande manier niet gebruiken, omdat de VAO niet meer in gebruik is. Voor de nieuwe dataset zal de data geëxporteerd worden vanuit het DWH en gehasht worden door team DASO. Deze gehashte dump wordt geüpload door DASO in het storage account van team AA in de Analyse Services. Om dit mogelijk te maken moet team Analyse Services de authenticatie en het netwerk zo instellen dat dit mogelijk is voor team DASO. Vanuit daar kan team AA de nieuwe gehashte data gebruiken om een nieuwe versie van het model te trainen.

Mocht deze manier toch niet optimaal zijn, dan zou ook nog de huidige API van DASO gebruikt kunnen worden. DASO moet dan een endpoint creëren waarmee team AA de data in verschillende batches kan ophalen en kan opslaan in het storage account in de Analyse Service subscriptie. Dit moet wel zo ingericht worden dat het netwerk niet overbelast raakt. Om dit mogelijk te maken zou team Analyse Services de ACL moeten updaten en API-bevragingen moeten toestaan vanuit de Analyse Services subscriptie naar de DASO subscriptie.

- Omdat Data Diensten nog niet gereed is, wordt data nog niet opgeslagen in de hub van Data Diensten, maar in een eigen SQL-database van DASO. Om dezelfde reden haalt Dataverkenner de model output op bij de API van DASO.
- Het is nog niet mogelijk om de data pijplijn te draaien op de Airflow van Data Verwerkingen. Datateam Sociaal richt daarom tijdelijk een eigen alternatief in (details nader te bepalen) op zo'n manier dat het zo eenvoudig mogelijk is om in de toekomst te migreren naar Data Verwerkingen. Hetzelfde geldt voor de bouwblokken die nog niet te gebruiken zijn. Meer informatie hierover is te vinden op de [wiki van Datateam Sociaal](#).
- ADW₃ is nog niet volledig uitgerold en het is niet zeker dat het alle vereisten voor het werk van de data scientist vervuld. Totdat deze twee problemen zijn opgelost, maakt de data scientist verbinding met Azure via zijn/haar eigen laptop (BYOD) en de VET VPN.

4.3 Afkortingen en begrippen

ADW	Amsterdamse Digitale Werkplek
Team AA	Team Advanced Analytics (team 9, AFG)
AFG	Amsterdammers Financieel Gezond
AS	Team Analyse Services
DASO	Datateam Sociaal (team 10, AFG)
Model	Het 'gedeployde' model voor uitkeringsaanvragen binnen de eigen container
VAO	Veilige Analyse Omgeving

4.4 Bijlagen

Bijlage 1 - BPMN plaat nieuwe situatie

