

Bijlage 7a - Proof of Concept Dataplatform

Inleiding

Onderdeel van deze aanbesteding is een verificatie van de winnaar middels een Proof of Concept. Na de voorlopige gunning moet de winnende leverancier in de PoC aantonen dat de software daadwerkelijk kan wat in de aanbesteding beweerd wordt. Dit duurt ongeveer 4 weken en na de verificatie wordt de PoC-omgeving verwijderd. Bij een positief resultaat volgt een definitieve gunning en zal direct gestart worden met de implementatie.

Voor de PoC zijn specifieke cases uitgekozen, gericht op de huidige (data)problemen binnen de organisatie. Na overleg is er bewust gekozen voor het gebruik van echte cases en echte data om de volgende redenen:

- De focus in de PoC moet liggen op de werking van de software, door gebruik te maken van bestaande cases waar we veel kennis van hebben, verzanden we niet in de uitdagingen van de cases zelf.
- Het ontwikkelen van fake data voor deze cases zou veel tijd en moeite kosten, met daarbij de vraag of de data dan ook nog realistisch en kwalitatief in orde is. Dit kan de werking van de PoC verstoren, daarom wordt er gekozen voor echte data.

Doel van de PoC

Verifiëren van een deel van de eisen bij de voorlopige winnaar. Positieve verificatie is een voorwaarde voor de definitieve gunning.

Looptijd

De PoC loopt van 15 maart t/m 14 april 2023. Op woensdag 15 maart zal de Kick Off plaatsvinden met alle deelnemers aan de PoC op het stadskantoor te 's-Hertogenbosch.

Casussen

Voor de PoC hebben we gekozen voor data van schulddienstverlening (Allegro/CiVision), telefoniecentrale (OSCC) en Meldingen Openbare Ruimte (Signalen).

Voor schulddienstverlening maakt gemeente 's-Hertogenbosch gebruik van Allegro. In dit systeem wordt data vastgelegd over deze processen. Verzoek is om een casus de implementeren waarbij gegevens van budgetbeheer ontsloten worden. Deze gegevens dienen verrijkt te worden met een referentietabel, waarbij de historische teamindeling inzichtelijk wordt gemaakt.

De software van onze digitale telefoniecentrale (OSCC) levert metadata op in de vorm van CSV-bestanden waarmee telefoongesprekken van bijvoorbeeld het call-center worden geanalyseerd. Deze dataset vraagt veel (handmatige) nabewerking voor ze gebruikt kan worden. In de case willen we testen in hoeverre dit geautomatiseerd kan worden in het platform.

Voor meldingen in de openbare ruimte door inwoners gebruiken we een specifieke gemeentelijke toepassing (Signalen) die met behulp van algoritmes de meldingen deels automatisch classificeert en toewijst voor behandeling. Dit systeem kan zichzelf verbeteren via machine learning. Dit vereist nu wel veel handmatige handelingen, waarbij meerdere losse export-bestanden moeten worden samengevoegd, ontdebeld, gefilterd en weer ge-upload. In de case willen we beproeven of het platform deze handelingen kan verminderen en/of automatiseren.

In de onderstaande tabel is uitgewerkt welke cases we willen uitvoeren en welke eisen daarmee beproefd worden. Cases 1 t/m 7 (en onderliggende subnummers) worden uitgevoerd met de data van schulddienstverlening. Cases 8 en 9 met data van de telefoniecentrale en case 10 met data van meldingen openbare ruimte.

#	Beschrijving	eis #
1.0	Als data engineer via een <u>visueel gemodelleerde</u> dataflow <u>de interne bron</u> voor Schulddienstverlening op ingestelde tijdstippen verversen en verwerken in een dataset voor verdere analyse voor schuldontwikkeling.	12 14
1.1	Als data engineer de dataset verrijken met contextuele informatie. Deze data wordt ingelezen uit een CSV-file die eerst geprepareerd en gecontroleerd wordt op kwaliteit (referentietabel).	11 15
1.2	Als data engineer de data lineage inzichtelijk maken.	18
1.3	Als data engineer een tweede versie van een flow maken en deze deployen.	13
1.4	Als data engineer monitoren van de flow en de opbouw van de dataset.	26
1.5	Als data engineer historische dataset opbouwen van schuldontwikkeling bij burgers (interne bron houdt nl. alleen status quo bij). Er kan gerapporteerd worden op individueel case-niveau en op geaggregeerd niveau, afhankelijk van de rol van de raadpleger.	17 20 40
1.6	Als data engineer gevoelige persoonsgegevens kunnen maskeren	35
1.7	Als data engineer geautomatiseerde DQ-validaties laten uitvoeren in de flow op de brondata en de data owner hierover informeren	30 32
1.8	Als data steward inzicht hebben in de ontwikkeling van data kwaliteit van de brondata.	33

1.9	Als data engineer het dataproduct aanbieden als dataproduct aan Cognos en PowerBI via API. Het dataproduct is beschikbaar in de vorm van stermodel en als geo-product (plotting schuldlasten per wijk bijv.)	22 23 27
1.10	Als data steward het dataproduct administreren in de data catalogus inclusief metadata (classificatie, privacy, betekenis, DQ bronsystemen, e.d.)	50
1.11	Als data consument het dataproduct kunnen raadplegen in de data catalogus inclusief metadata (classificatie, privacy, betekenis, DQ bronsystemen, e.d.)	10
1.12	Als data consument alleen de data uit dataset getoond krijgen waar ik op basis van mijn rol recht op heb.	5
2.0	Als data engineer wil ik data uit een bron verversen en verwerken met behulp van een API uit de bron Civision.	19
3.0	Als <i>niet-data engineer</i> wil ik onbekende data kunnen preppen & kunnen interpreteren: in dit geval uit het telefoniebestand OSCC	1
3.1	Als data scientist wil ik onbekende data kunnen preppen & kunnen interpreteren: in dit geval uit het telefoniebestand OSCC	42
4.0	Als data scientist wil ik het AI-model voor de herleiding in de signalen-database regelmatig her-trainen, waarbij ik de datakwaliteit van de toetsdata/uitkomsten meer geautomatiseerd kan controleren/verbeteren	50