

## Opdrachtformulering WOB verzoek “Shell Papers”

### Aanleiding:

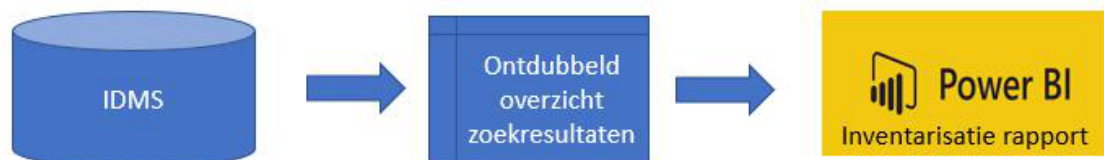
De provincie Zuid-Holland heeft een WOB-verzoek ontvangen van Platform Authentieke Journalistiek (PAJ) en “Follow the Money” waarin gevraagd wordt om inzicht te geven in alle documenten rondom Shell, de dochterondernemingen van Shell en aan Shell gelieerde lobbyisten. Deze vraag legt een zware belasting op het WOB-team en alle betrokken afdelingen om binnen de tijdslimieten van de WOB-procedure te leveren aan de indiener. Het verzoek omvat alle documenten die zijn ontvangen of opgemaakt voor de dagtekening van het WOB-verzoek.

### Voorstel:

De afdeling I&A denkt het WOB-team en de vakafdelingen te kunnen ondersteunen door een slimme inzet van beschikbare IT-middelen en van specialistische kennis op het gebied van analyse van ongestructureerde data. Het scrumteam Vernieuwing Data Warehouse (vDWH) is de trekker van deze oplossing.

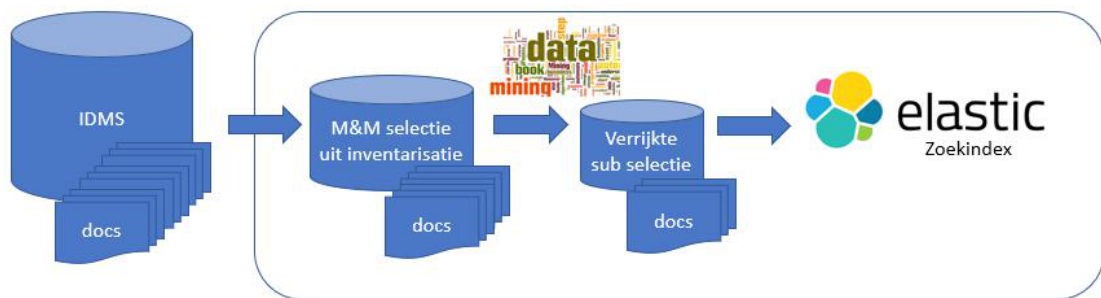
Het voorstel bestaat uit de volgende stappen:

1. Bevragen van IDMS met relevante zoektermen om te komen tot een eerste **inventarisatie** van de totale omvang van alle potentieel relevante documenten, het bijbehorend aantal pagina's en verdeling naar type documenten. Dit overzicht presentabel maken middels een Power BI rapportage.



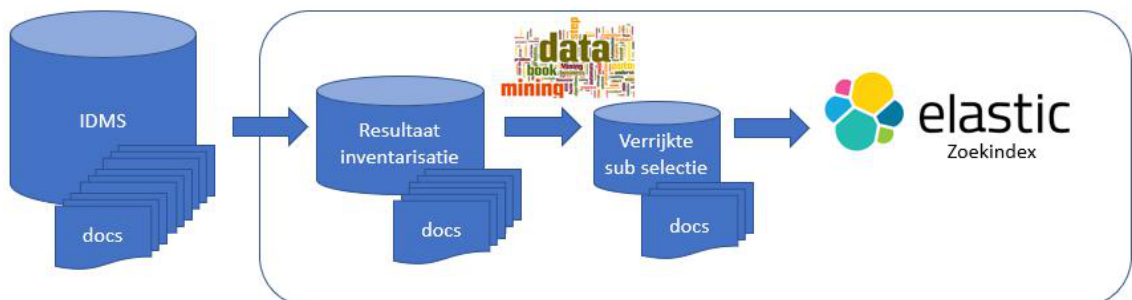
Consequentie: deze stap kan worden gerealiseerd door middel van reeds begrote inzet van het vDWH en IDMS team en vraagt maximaal 5 werkdagen aan inzet van professionals uit beide teams.

2. Uitvoeren van een **Proof of Concept** voor de afdeling Milieu en Mobiliteit (DBI) om potentieel relevante documenten uit het IDMS te laden naar een analyse omgeving in de Microsoft Azure (Cloud) omgeving. Vanuit deze analyse omgeving zal het vDWH-team de IDMS documenten verrijken door middel van textmining analyses. Met deze tekstmining analyses kan de inhoud van het document worden doorzocht op het voorkomen van Shell, de dochterondernemingen van Shell en aan Shell gelieerde lobbyisten (inclusief variaties op de schrijfwijzen van deze entiteiten). Bij het document wordt vervolgens als aanvullende metadata “tags” van de documenten bijgehouden welke van de relevante trefwoorden gevonden zijn. Vervolgens worden de met tags verrijkte documenten ingeladen in een zoekindex (Elasticsearch) waar ze kunnen worden gerangschikt op relevantie en kunnen worden geanalyseerd om de meest belangrijke documenten te vinden.

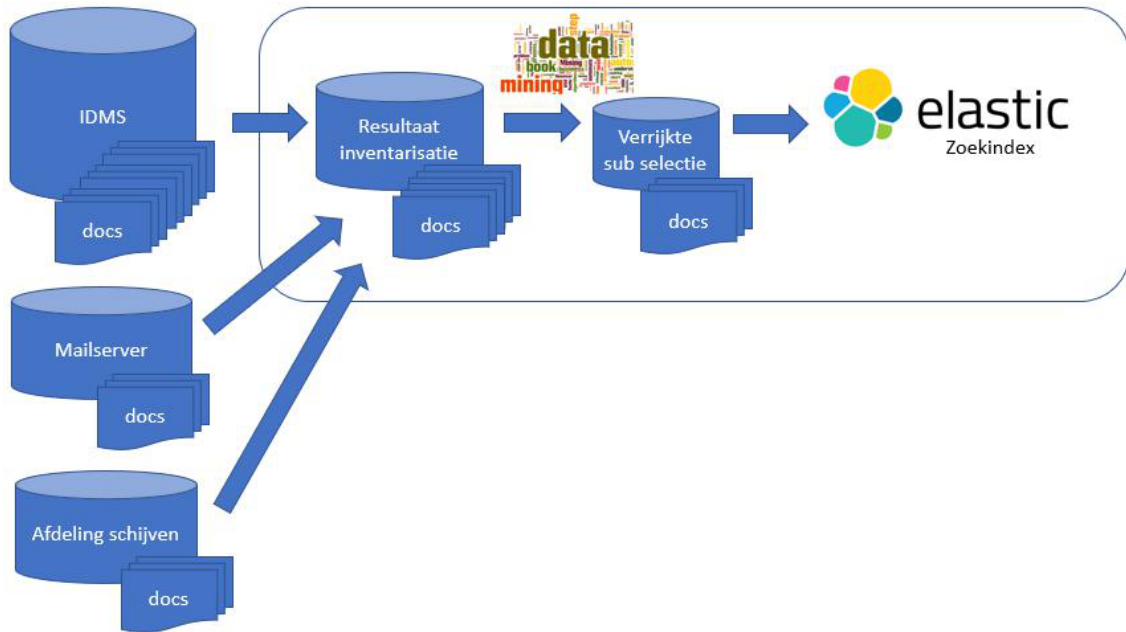


Consequentie:

- a. Financiële impact: de inzet van aanvullende professionals bij het vDWH is noodzakelijk om deze stap snel en aanvullend op lopende leveringsverplichtingen te kunnen doen. Daarnaast zorgt de (tijdelijke) inrichting van opslag- en analysecapaciteit in Azure voor aanvullende kosten op de Microsoft Azure subscription. Deze kosten kunnen apart worden toegerekend aan het uitvoeren van deze oplossingsrichting. Zie bijlage 1 voor een specificatie van de geraamde kosten.
  - b. Afstemming met functionaris gegevensbescherming (FG) en Record Manager (RM): de oplossing maakt gebruik van het tijdelijk dupliceren van documenten vanuit IDMS naar een locatie in de afgeschermden Microsoft Azure cloud-omgeving van de provincie. Op voorhand vraagt het vDWH team de instemming van zowel de FG als de RM voor het uitvoeren van deze oplossingsrichting, waarbij de volgende randvoorwaarden worden gehanteerd:
    - i. De data-uitwisseling vindt volledig plaats binnen het beveiligde netwerk van de provincie, waar de Azure-omgeving van de provincie naadloos op is aangesloten.
    - ii. Na afloop van het WOB-verzoek en na een nader overeen te komen bewaarperiode wordt de tijdelijke omgeving, inclusief alle verzamelde en geanalyseerde documenten, zodanig opgeheven dat de documentaire informatie als permanent vernietigd kan worden beschouwd. Alle stukken over de afhandeling van dit WOB-verzoek die worden opgemaakt of verzonden aan de indiener zullen worden gearchiveerd in het IDMS.
3. Na het uitvoeren van de Proof of Concept wordt er een Go/No Go beslissing genomen over de inzet van de oplossingsrichting voor de gehele provincie. Indien de Proof of Concept door het team van Milieu en Mobiliteit en door het WOB-team als succesvol wordt beoordeeld, dan wordt stap 2 herhaalt met als scope alle IDMS documenten van de provincie.



4. Na (of parallel aan) het uitvoeren van stap 3 voor alle IDMS documenten kan dezelfde exercitie worden uitgevoerd voor andere bronnen
  - a. Afdelingschijven
  - b. Mailserver



## Bijlage 1: Specificatie meerkosten van oplossingsrichting

De meerkosten bestaan uit zowel tijdelijke aanvullende inzet van specialistisch personeel en van out-of-pocket kosten voor diensten op het Microsoft Azure cloud-platform van de provincie.

Type inzet	Kosten bij PoC (uitvoeren stap 2)	Kosten bij inzet voor gehele provincie (uitvoeren stappen 3 en 4)
Ondersteuning Data Scientist met tekst-mining specialisatie	€5.000	€20.000
Ondersteuning Elasticsearch ontwikkelaar	€2.400	€4.000
Meerkosten Azure-platform*	€2.000	€10.000
<b>Totaal</b>	<b>€9.400</b>	<b>€34.000</b>

\*betreft een globale schatting, die na het uitvoeren van de PoC kan worden verfijnd.