



Ministerie van Binnenlandse Zaken en
Koninkrijksrelaties

Rapport BODO-pilot: Overzicht en Inzicht



Inhoud

1	Inleiding	3
2	Aanleiding pilot	4
3	Doelstellingen en onderzoeksvragen	5
4	Geplande aanpak	6
5	Uitgevoerde aanpak	7
	5.1 Keuze bronapplicaties	7
	5.2 Keuze ICT-platform	7
	5.3 Configuratie IDOL Find	8
	5.4 Evaluatie en afronding	9
6	Resultaten	10
	6.1 Resultaatbeelden	10
7	Evaluatie pilot	17
	7.1 Antwoorden op onderzoeksvragen	17
	7.2 Bevindingen pilot	18
8	Conclusie	20
9	Aanbevelingen	21

1 Inleiding

Het Rijksprogramma voor Duurzaam Digitale Informatiehuishouding (RDDI) helpt Rijksorganisaties bij het op orde brengen én houden van hun informatiehuishouding (IHH). Informatie moet gemakkelijk terug te vinden zijn en niet verloren gaan, zolang het bewaard moet worden. In brede zin is de uitdaging om de IHH van de Rijksoverheid te verbeteren door deze duurzaam toegankelijk te maken.

RDDI heeft het project Bevordering Overbrenging Digitale Overheidsinformatie (BODO). Dit project richt zich op het opdoen van concrete kennis en ervaring met de randvoorwaarden waaraan Rijksorganisaties moeten voldoen in het overbrengingsproces van digitale archieven naar het Nationaal Archief (NA).

Binnen het project BODO lopen twee pilots in samenwerking met Doc-Direkt. Bij het departement van Financiën (Fin) is een pilot gericht op het zoeken en opruimen van dubbele documenten. Bij het departement van Sociale Zaken en Werkgelegenheid (SZW) is een pilot gericht op het creëren van overzicht en inzicht over bestaande informatie- en archiefbronnen. Dit document dient als rapportage van de uitvoering van de tweede pilot bij SZW.

2 Aanleiding pilot

Het departement van Sociale Zaken en Werkgelegenheid (SZW) was al bezig met een eigen actieplan “SZW Open op Orde” voor een betere informatiehuishouding (IHH). Hierdoor hebben zij zich als vrijwillige organisatie opgegeven waarbij RDDI zijn BODO-pilot “Overzicht & Inzicht” kan uitvoeren, vanwege de verschillende aanhakingspunten tussen de doelstellingen van beide organisaties.

In het actieplan “SZW Open Op Orde” wordt onder actielijn 3 “Informatiesystemen” de fiche “Duurzaam Toegankelijke Informatiesystemen” opgenomen. Deze fiche heeft als doel om tot concrete verbetervoorstellen te komen voor het bronapplicatieportfolio van SZW. Dit omvat zowel de bronapplicaties¹ afzonderlijk als in samenhang met elkaar. De eerste stap bij deze fiche is het creëren van overzicht en inzicht in de inhoud én samenhang van de bronapplicaties. Dit vertaalt zich uiteindelijk naar het creëren van overzicht en inzicht in de informatiehuishouding. Dit wil men graag doen met behulp van een tool die automatisch overzicht en inzicht kan genereren. Na een positieve demo van de tool “IDOL Find” van Micro Focus (gehouden op 16 maart 2022) is er besloten dat in deze pilot onderzocht wordt in hoeverre IDOL Find handige overzichten van een aantal bronapplicaties van SZW automatisch kan creëren.

IDOL Find (anders genoemd IDOL Unstructured Data Analytics) is een geavanceerd zoek-, kennisontdekking-, en analyseplatform. Het maakt gebruik van AI en machine learning (waaronder tekstanalyse, audio-analyse, videoanalyse en beeldanalyse) om belangrijke inzichten te verkrijgen uit ongestructureerde data. Daarnaast, biedt IDOL Find functionaliteiten om deze inzichten in verschillende vormen ofwel overzichten te visualiseren. Voorbeelden hiervan zijn verder in dit rapport te vinden.

N.B. De Rijksoverheid bezit al een licentie op IDOL Find via het softwareprogramma Zoek & Vind dat afgenomen wordt bij Knowledge Plaza. Zoek & Vind en IDOL Find draaien beide op dezelfde technologie onder de motorkap, maar ze zijn wel twee afzonderlijke applicaties van twee verschillende bezittende leveranciers. De applicaties verschillen zich dusdanig van elkaar in functionaliteiten dat ze als twee niet vergelijkbare applicaties worden gezien. Bijvoorbeeld IDOL Find kan overzichten visualiseren, maar Zoek & Vind niet. Zoek & Vind heeft een module om gevoelige informatie te monitoren, maar IDOL Find niet.

¹ Bronapplicaties zijn de applicaties die informatie en data van SZW bevatten. Applicaties zoals een knipprogramma, rekenmachine e.d. vallen niet hieronder.

3 Doelstellingen en onderzoeksvragen

De doelstellingen van de pilot zijn als volgt geformuleerd:

1. *Onderzoeken en beproeven van de mogelijkheden om overzichten en inzichten te genereren uit bronapplicaties met behulp van IDOL Find (met beperkte technische integratie).*
2. *Onderzoeken en beproeven met welke functionaliteiten van IDOL Find overzichten en inzichten van de inhoud van bronapplicaties gecreëerd kunnen worden.*

Dit vertaalt zich naar de volgende onderzoeksvragen voor de pilot:

Kan IDOL Find...

- De Rijksoverheid voorzien in (snel) overzicht en inzicht krijgen in zijn IHH?
- Automatisch een (begin-)overzicht en inzicht genereren?
- Van dezelfde data verschillende overzichten genereren afhankelijk van de analysebehoefte?
- Op de metadata van documenten zoeken? Dus bijvoorbeeld specifiek naar de naam van een auteur die voorkomt in de documenten in de bronapplicaties.
- Werken met meerdere typen bronapplicaties?
- Werken met verschillende bestandsformaten?
- Voldoen aan de informatiebeveiligingsvoorwaarden van de Rijksoverheid?
- Schaalbaar zijn op de grote hoeveelheid data van de Rijksoverheid?

De antwoorden op deze vragen zijn te vinden aan het einde van het hoofdstuk “Resultaten”.

4 Geplande aanpak

Het projectteam heeft de volgende samenstelling:

Naam	Rol/functie
RDDI + Doc-Direkt + SZW	Opdrachtgevers
René IJperlaan	Projectleider ENIGMA bij SZW
Rèdjinald Lichtenberg	Projectleider bij Doc-Direkt
Melior Hoekstra	Programma-architect en pilotregisseur bij RDDI
Frank de Pater	Senior technische consultant IDOL Find bij Micro Focus

Gezien dat het doel van deze pilot is om de mogelijkheden van IDOL Find te onderzoeken, is er gekozen om de scope beperkt te houden. Zo kan de pilot ook zo snel mogelijk van de grond komen en uitgevoerd worden. De details van de geplande aanpak zijn als volgt:

- *Keuze aantal te doorzoeken bronapplicaties.*
SZW heeft zelf een totaal aan ca. 20 bronapplicaties. Om het behapbaar te houden en omwille van de tijd is ervoor gekozen om 3 bronapplicaties van 1 directie te doorzoeken. N.B. de directie heeft al een idee wat er in de bronapplicaties te vinden zijn. Hiermee kunnen de gegenereerde overzichten en inzichten met IDOL Find gevalideerd worden door de experts van de bronapplicaties.
- *IDOL Find laten landen op het platform van een Rijks ICT-leverancier en de techniek neerzetten.*
De keuze hiervoor ligt tussen het platform van ODC-Noord, het Standaard Platform van Logius en het platform van SSC-ICT; de uitwerking van de uiteindelijke keuze volgt in het volgende hoofdstuk. Hierbij hoort dat de ICT-platformleverancier toegang en ondersteuning moet bieden aan de leverancier van IDOL Find om de bronapplicaties te koppelen en aansluiten met IDOL Find. IDOL Find wordt beschouwd aan de informatie-beveiligingsvoorwaarden van de Rijksoverheid te voldoen, omdat het op een ICT-platform van de Rijksoverheid zal landen die niet verbonden is met externe bronnen via het openbare internet o.i.d. Verder zal er een beperkt aantal gebruikers toegang naar hebben.
- *De leverancier zal IDOL Find configureren om bruikbare overzichten en inzichten te creëren.*
In samenwerking met experts van de bronapplicaties verkent de leverancier van IDOL Find de verwachte typen documenten in de gekozen bronapplicaties en op basis hiervan IDOL Find configureren om kwalitatieve overzichten en inzichten te realiseren. Het configureren van IDOL Find gebeurt iteratief zodat er een geoptimaliseerd eindresultaat wordt geleverd.
- *Evaluatie en afronding pilot.*
Het eindresultaat wordt besproken en afgestemd met de opdrachtgevers. Met hun goedkeuring wordt ter afronding van de pilot een demo van het eindresultaat gehouden aan de grotere BODO-klankbordgroep en levert het pilotteam hun bevindingen en advies op in een eindrapportage.

5 Uitgevoerde aanpak

De volgende uitwerking van de gevoerde aanpak voor de pilot is in chronologische volgorde geschreven. De koppen komen ook puntsgewijs overeen met uitgewerkte stappen in de geplande aanpak.

5.1 Keuze bronapplicaties

De keuze naar de top 3 bronapplicaties ging naar de bronapplicaties waarin de meeste medewerkers van de directie Organisatie, Bedrijfsvoering en Personeel (OBP) hun documenten opslaan. Dit zijn hun Digidoc-mappen, hun netwerkschijven en hun samenwerkruimte. Maar uiteindelijk is de samenwerkruimte niet meer meegenomen in de pilot, omdat deze wordt geleverd door een derde partij die geen ondersteuning biedt om de data toegankelijk te maken voor een tool zoals IDOL Find.

In eerste instantie was SZW ook geïnteresseerd om de bronapplicatie “Mozard” te doorzoeken. Echter, deze bronapplicatie blijkt zelf geen data te bewaren, maar eerder data uit verschillende bronnen te halen om daarmee een zaak af te handelen. Met zo’n applicatie zou IDOL Find aangesloten moeten worden met alle toeleverende bronnen om zo alle relevante data op te kunnen halen. Er moest dus eerst onderzocht worden welke onderliggende bronnen er allemaal zijn en vervolgens hoe deze bronnen aangesloten moeten worden met IDOL Find. Dit maakte het vraagstuk voor de pilot te ingewikkeld, waardoor Mozard als optie werd geweigerd.

5.2 Keuze ICT-platform

Zoals eerder vermeld, moest er een keuze worden gemaakt tussen de platforms van de drie Rijks ICT-leveranciers: ODC-Noord, Logius en SSC-ICT. Er was besloten om SSC-ICT als eerste keuze te omzeilen om allerlei verwarringen te vermijden rondom de al bestaande IDOL-motor die al op hun platform draait voor Zoek & Vind. Hierdoor ging de eerste keuze naar het platform van ODC-Noord, de tweede keuze naar het Standaard Platform van Logius en als laatste optie het platform van SSC-ICT.

5.2.1 Eerste keuze: ODC Noord

Het platform van ODC-Noord werd als eerste keuze gekozen vanwege de volgende redenen:

ODC-Noord...

- Werkt volgens moderne standaarden (Agile, Scrum, etc.),
- Is in staat tot snelle leveringen (hoge flexibiliteit),
- Levert schaalbare generieke diensten
- En is beschouwd als “in control” op gebied van de Baseline Informatiebeveiliging Overheid (BIO).

Het platform van ODC-Noord leek in eerste instantie dus de beste optie te zijn waarop IDOL Find kan landen. Maar, bij nader inzien bleek dat er geen ondersteuning vanuit ODC-Noord was voor het verbinden en aansluiten van IDOL Find op hun platform met bronapplicaties op platforms van andere ICT-leveranciers. Daarnaast, staan de meeste bronapplicaties van SZW op het platform van SSC-ICT. Bovendien, bleek dit verbinden en aansluitingsvraagstuk best complex te zijn én veel meer middelen (mankracht, expertise, budget en tijd) te vereisen. Dit vormde een bedreiging voor de focus van de pilot, want de focus werd meer over het uitpuzzelen van hoe verbinding en aansluiting gemaakt kan worden tussen de twee ICT-platforms in plaats van het onderzoeken van de mogelijkheden

van IDOL Find. Dit allemaal heeft samen ertoe geleid dat het platform van ODC-Noord niet meer een passende optie was voor de pilot.

5.2.2 Overstap naar tweede keuze: Standaard Platform van Logius

Het Standaard Platform van Logius leek in eerste instantie ook een goede optie te zijn waarop IDOL Find kon landen. Het platform biedt gebruikers kant-en-klare werk-omgevingen (Kubernetes Namespaces) waarmee overheden zelf hoogwaardige digitale diensten zoals applicaties of websites kunnen ontwikkelen, testen en beheren. Maar, het Standaard Platform werkt op basis van een docker-systeem. Bij een docker-systeem moet een applicatie met alle nodige instellingen en configuraties in één keer als één pakket worden aangeboden aan het platform. Dit betekent dat het ingewikkelder is om configuratiefouten in een programma op te sporen, bij te houden en verbeteringen hierop door te voeren. Bovendien, moet ook op zo'n manier verder rekening worden gehouden met het verbodings- en aansluitingsvraagstuk tussen de bronapplicaties op verschillende platforms. En alhoewel ondersteuning hiervoor beschikbaar was vanuit Logius, waren er meerdere expertises nodig om dit te regelen. Dit creëerde hetzelfde focusprobleem als beschreven bij de uitwerking van de eerste optie bij ODC-Noord. Dus het Standaard Platform was hierdoor ook geen passende optie meer voor de pilot.

5.2.3 Overstap naar derde keuze: SSC-ICT

Het platform van SSC-ICT was al bewezen een passende optie te zijn voor het landen van IDOL Find, want zoals al eerder vermeld, draait hier al een IDOL-motor voor Zoek & Vind. Bovendien, staan de bronapplicaties van SZW ook op dit platform, waardoor het verbodings- en aansluitingsvraagstuk hierbij minder complex zou zijn. Zoals verwacht, moest er echter in de voorgesprekken met SSC-ICT telkens verwarringen worden opgeklaard waarom een tweede instantie van een IDOL-motor gewenst was en hoe IDOL Find zich verschilt van Zoek & Vind. Deze voorgesprekken leidden tot een te lange doorlooptijd voordat de pilot kon starten, waardoor er besloten werd om naar een workaround te kijken.

5.2.4 Workaround

Gezien dat Zoek & Vind al aangesloten is op de te doorzoeken bronapplicaties van OBP én omdat het draait op een IDOL-motor is een dataexport uit Zoek & Vind compatibel met IDOL Find. Hierdoor kon het hele vraagstuk rondom het verbinden en aansluiten van IDOL Find met de bronapplicaties worden omzeild, terwijl de nodige data toch geraadpleegd kon worden voor het doel van de pilot. Deze workaround werd in samenwerking met KnowledgePlaza gedaan voor de dataexports en IDOL Find werd geland op een server-omgeving van Doc-Direkt.

5.3 Configuratie IDOL Find

Voor de configuratie van IDOL Find waren filters en stopwoorden gebruikt om de kwaliteit van de overzichten te optimaliseren.

5.3.1 Filters

De filters in IDOL Find waren ingedeeld volgens de mapindeling van Digidoc én de netwerkschijven. Dit is gekozen om zo het zoeken binnen IDOL Find te faciliteren via de al bestaande zoekpaden in de bronapplicaties. Hiermee werd ook de validatie van de gegenereerde overzichten gefaciliteerd, want IDOL Find kan ook overzichten genereren op mapniveau in een bronapplicaties.

5.3.2 Stopwoorden

Stopwoorden zijn woorden die uitgesloten worden van de IDOL-index. Deze woorden worden dus als niet relevant beschouwd en uiteindelijk niet meegenomen in de gegenereerde “topic map” weergaven in IDOL Find (zie hoofdstuk Resultaten). Daarnaast worden stopwoorden gebruikt om de IDOL-index schoon te houden van woorden die niets toevoegen bij de zoekresultaten. Denk hier aan bijvoorbeeld lidwoorden, persoonlijke voornaamwoorden, voorzetsels, e.d.

Er zijn twee lijsten met stopwoorden gebruikt: een korte en een lange lijst. De korte lijst is een basispakket van stopwoorden die standaard meegeleverd wordt door Micro Focus. De lange lijst is verder aangevuld met stopwoorden die toegespitst zijn op de context én inhoud van de gevonden documenten bij het hanteren van de korte stopwoordenlijst. Zo werden dus extra stopwoorden toegevoegd aan de korte stopwoordenlijst die voorkwamen in de clusters van de minder optimale topic map weergaven. Hierdoor komen o.a. Rijksadressen (zoals “Den Haag”) niet meer voor als clusters in de gegenereerde topic map weergaven.

Opmerkelijk was dat na het toepassen van de lange stopwoordenlijst is IDOL Find clusters gaan maken van tekststukken zoals uitdrukkingen. Dus in plaats van clusters met een duidelijk thema of woord waren het tekststukken zoals ‘in het geval van’ of ‘in relatie tot’ e.d. Dit kwam voor door de grootte van de dataset; van beide bronapplicaties waren in totaal ca. 2000 documenten gevonden. Voor IDOL Find is deze datasetgrootte aan de kleine kant. Met een grotere dataset (zoals in het geval dat IDOL Find verder geïmplementeerd wordt op een grotere schaal binnen de Rijksoverheid) worden zulke tekststukken niet meer als clusters verwacht in de topic map weergaven.

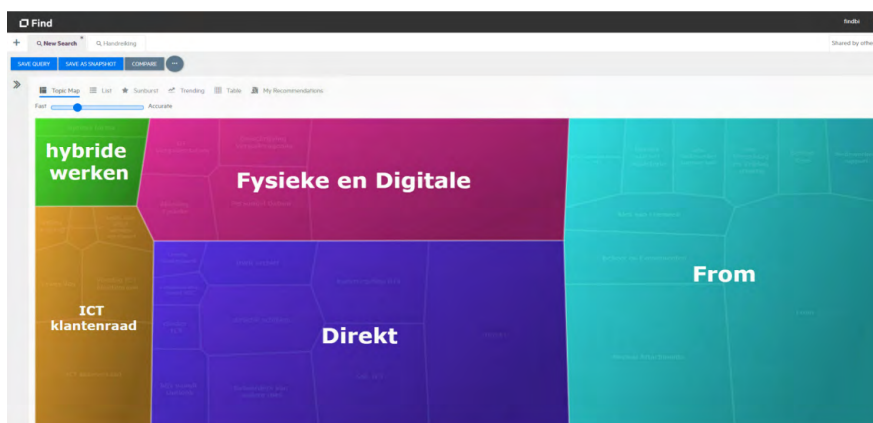
5.4 Evaluatie en afronding

De daadwerkelijke uitwerking van de evaluatie én afronding van de pilot verliep verder zoals beschreven in de geplande aanpak.

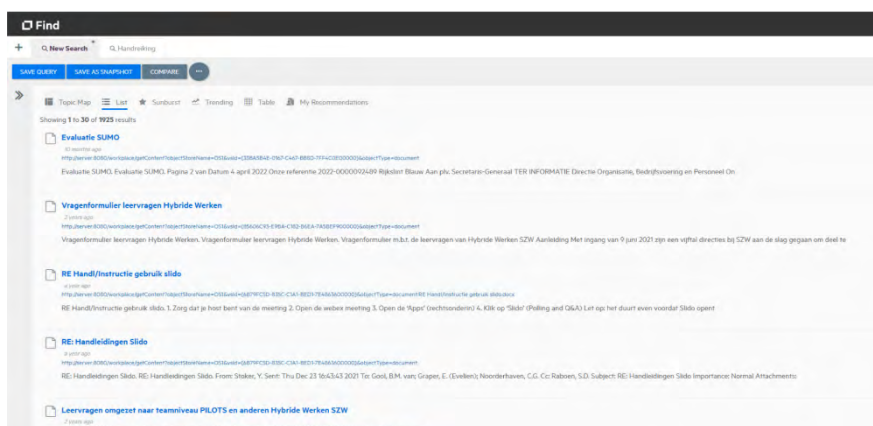
6 Resultaten

6.1 Resultaatbeelden

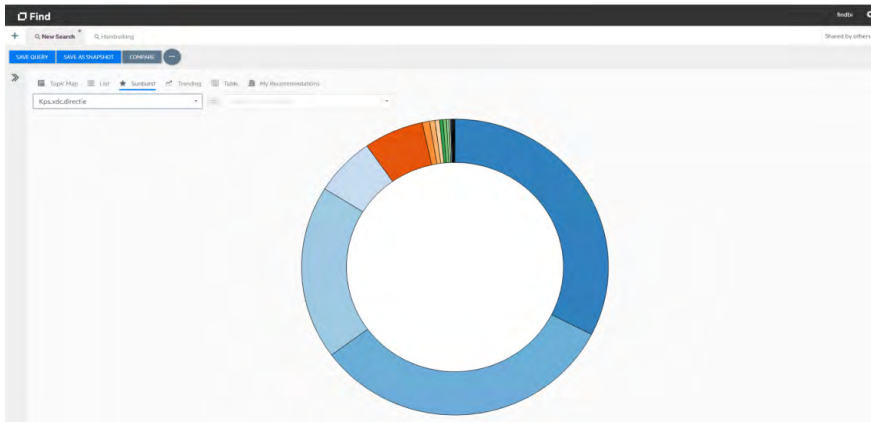
Hieronder volgen de resultaatbeelden van de pilot; elk met een korte omschrijving en verdere uitleg.



- **Hoofdpagina in topic map weergave.**
Dit is de standaard weergave van een dataset die IDOL als eerste laat zien. Hier kan gezien worden hoe de meest voorkomende onderwerpen in de dataset naar boven komen als clusters. Dit overzicht is overigens gegeneerd met twee iteraties van de lange stopwoordenlijst. Zo zijn er onhandige clusters zoals “Den Haag”, “Datum” e.d. uit het overzicht gehaald. Gezien het toepassen van een stopwoordenlijst een iteratief proces is, kan bij een volgende iteratie ook het woord “From” uit de bovenstaande topic map worden gehaald.



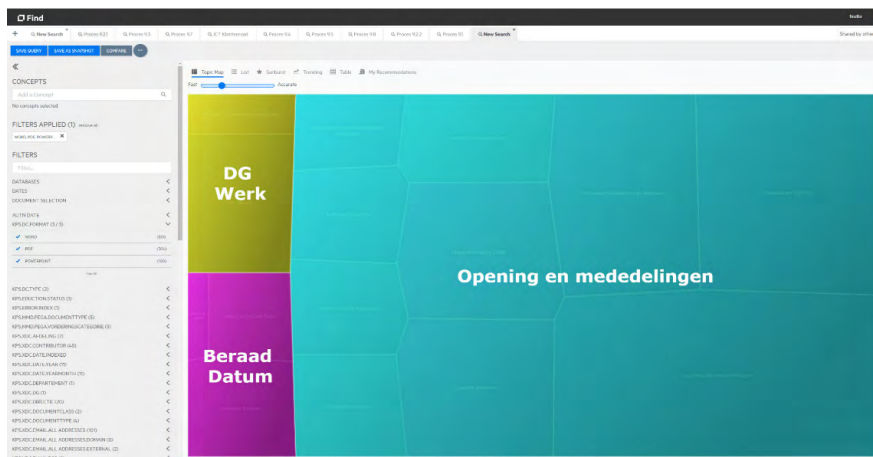
- **Hoofdpagina in een generieke lijstweergave.**
Met de lijstweergave zijn titels van de gevonden documenten te zien met hun opslaglocatie in de bronapplicaties. Door op de titel te klikken, wordt een preview van het document gegeven. En door op de opslaglocatie te klikken is het mogelijk om direct naar het document te gaan. Voor deze pilot is de toegang naar de preview én opslaglocatie van de documenten beperkt gehouden t.b.v. geheimhouding. Er zal hier in dit rapport dus geen voorbeeld van worden gegeven.



- **Hoofdpagina in een cirkeldiagramweergave (sunburst): gefilterd op de gevonden directies in de dataset.**
 Hier is te zien hoeveel documenten in de dataset behoren tot een bepaalde directie. Als er over een deel van het cirkeldiagram in IDOL wordt gehoverd, dan is te zien over welke directie dat deel gaat.

Kpaxdcirectie	Count
OSP_OV	500
OSP_FOV	533
OSP	436
OV	167
OSP_SCI	141
DSD	10
OSP_OBE	11
IZ	6
RO	7
BOB	7
POV	5
BV	2
OSP_HRM	2
ENGMA	1
FEZ	1
GOVK	1
POV	1
...	...

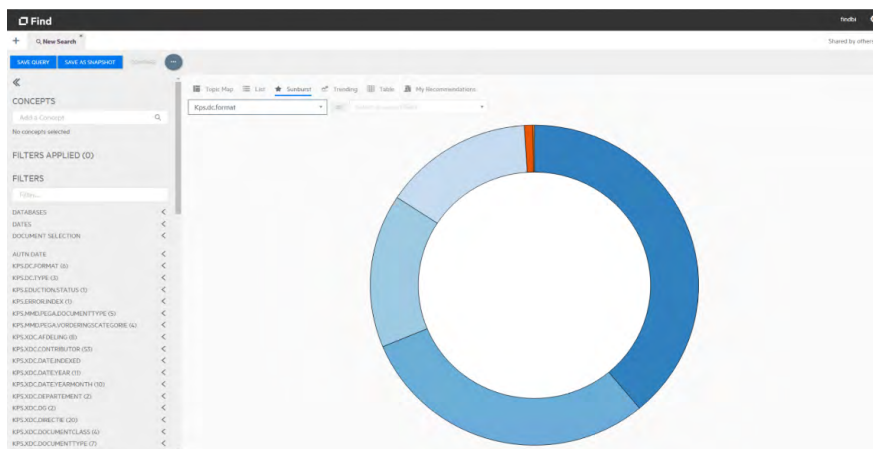
- **Hoofdpagina in een generieke tabelweergave: gefilterd op de gevonden directies in de dataset.**
 Hier is dezelfde data van het vorige cirkeldiagram te zien, maar in de generieke vorm van een tabel. Zo'n tabelweergave is handig, want tabellen worden vaak gebruikt in rapporten, verslagen, o.i.d.



- **Hoofdpagina in een topic map weergave: gefilterd op de bestandsformaten word, pdf en powerpoint.**

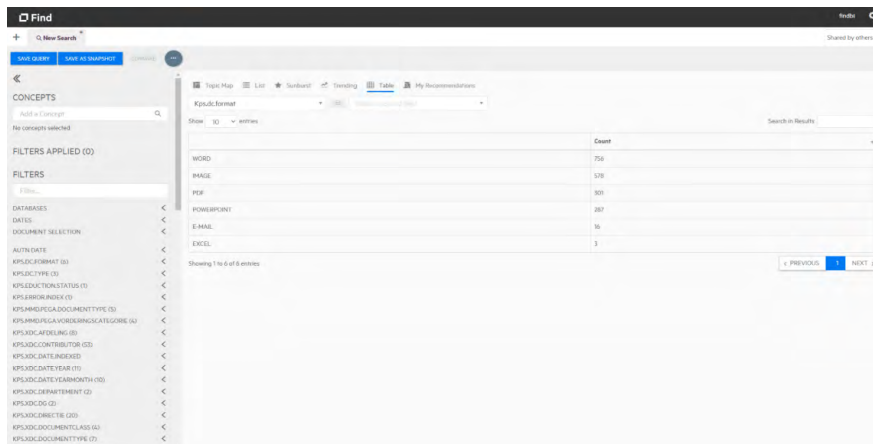
Hier is te zien dat het mogelijk is om de dataset te filteren o.b.v. bestandstypen.

Hiermee kan gezien worden welke onderwerpen het meest voorkomen in de geselecteerde bestandsformaten a.d.h.v. topic map weergaven.



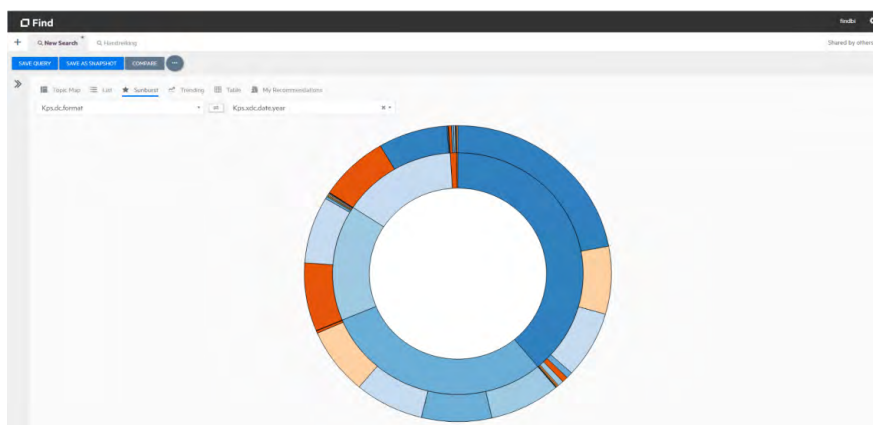
- **Hoofdpagina in een cirkeldiagramweergave: gefilterd op de gevonden type documenten in de dataset.**

In een cirkeldiagramweergave is het ook mogelijk om de dataset te filteren. Zoals te zien in de bovenstaande afbeelding is de dataset gefilterd op bestandsformaten via het drop-down-menu linksboven het cirkeldiagram. In vergelijking met de gekozen selectie van word, pdf en powerpoint bij de vorige topic map weergave behelst dit cirkeldiagram dus alle gevonden bestandsformaten in de dataset. En als er over een deel van het cirkeldiagram met de muis wordt gehoverd, is te zien naar welk bestandsformaat dat deel refereert.



- **Hoofdpagina in tabelweergave: gefilterd op de gevonden type documenten in de dataset.**

Hier is dezelfde data van het vorige resultaatbeeld (het cirkeldiagram) te zien, maar in een generieke tabelvorm.



- **Hoofdpagina in een cirkeldiagramweergave: gefilterd op de gevonden type documenten in combinatie met het jaar waarin ze gemaakt zijn.**

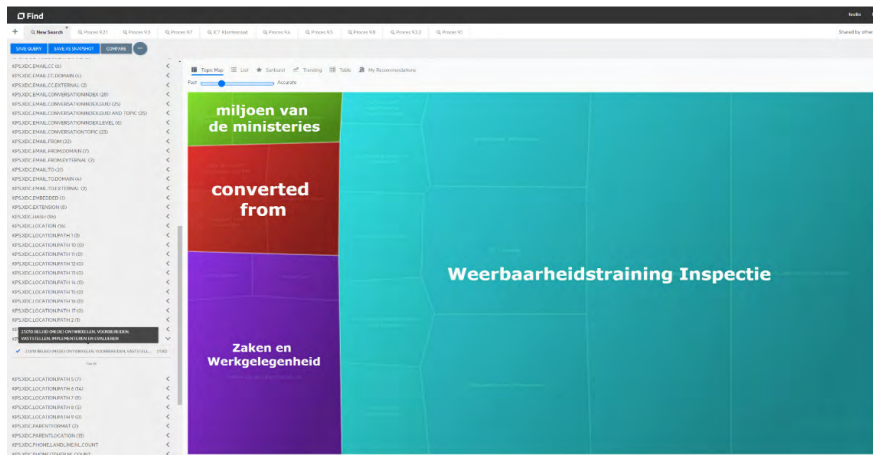
Hier is te zien dat in IDOL ook mogelijk is om cirkeldiagrammen te genereren van combinatie van kenmerken. Als er over een deel van het cirkeldiagram in IDOL met de muis wordt gehoverd, dan is te zien naar welk bestandsformaat of jaar dat deel refereert. Ook is hiermee te zien welke bestandstypen allemaal in een bepaald jaar zijn gemaakt.

	2011	2015	2016	2018	2019	2020	2021	2022
E-MAIL	0	0	0	1	0	0	1	4
EXCEL	0	0	3	0	0	0	0	0
IMAGE	0	0	0	1	0	161	0	163
PDF	0	2	1	2	0	161	1	5
POWERPOINT	0	0	0	0	0	161	2	0
WORD	1	0	2	3	1	161	426	18

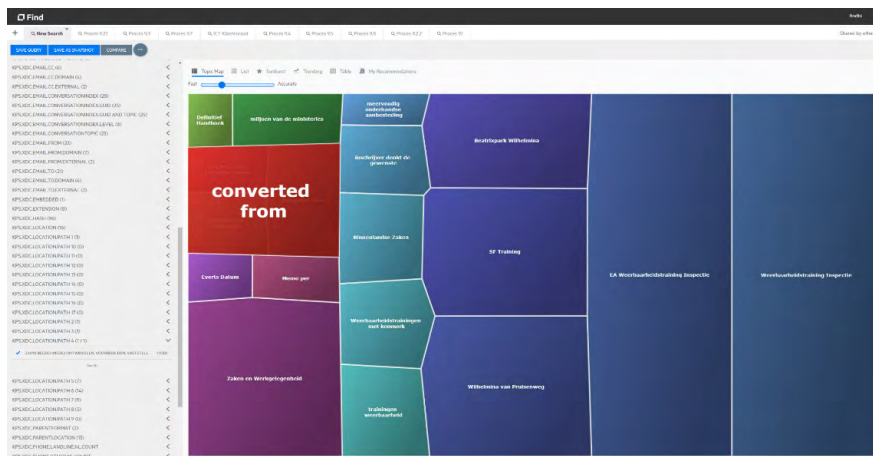
- **Hoofdpagina in tabelweergave: gefilterd op de gevonden type documenten in combinatie met het jaar waarin ze gemaakt zijn.**
 Hier is dezelfde data van het vorige cirkeldiagram te zien, maar in een generieke tabelvorm. In IDOL is het dus ook mogelijk zulke weergaven van combinatie van kenmerken in tabellen te genereren.

	E-MAIL	EXCEL	IMAGE	PDF	POWERPOINT	WORD
BIO	2	0	2	1	0	2
BSD	4	0	2	1	0	0
BV	0	0	0	1	0	1
CV	1	0	1	0	162	3
DSU	1	3	1	4	0	6
ENIGMA	0	0	0	0	0	1
FEZ	0	0	0	1	0	0
GAVV	0	0	0	1	0	0
IVV	0	0	0	1	0	0
IZ	3	0	3	0	0	3

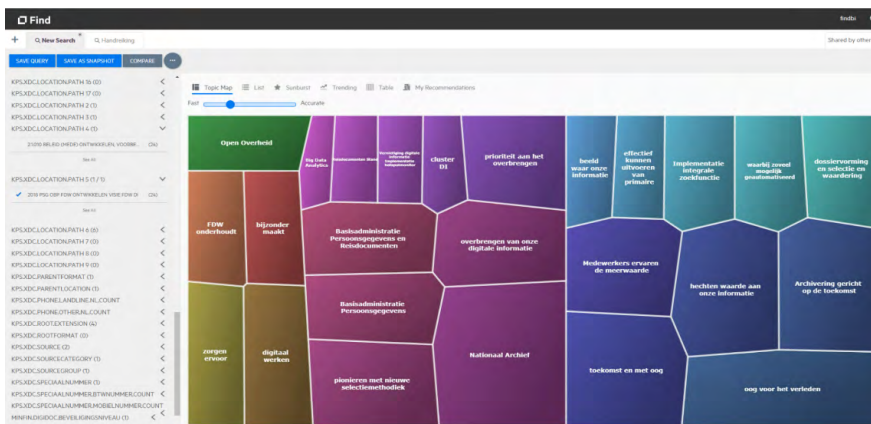
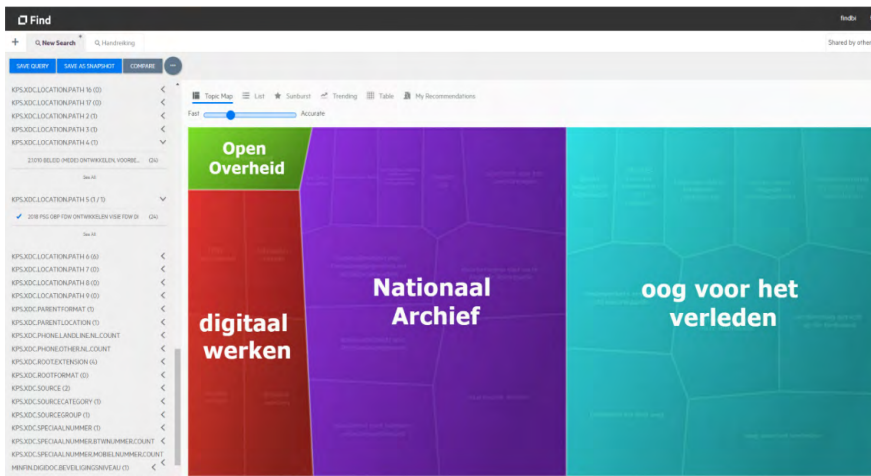
- Hier is te zien dat het genereren van weergaven met een kenmerkcombinatie in een tabelvorm los staat van die in cirkeldiagrammen. Er hoeft dus niet eerst een cirkeldiagram te worden gemaakt van de kenmerkcombinatie om deze vervolgens in een tabelvorm weer te geven.



- **Topic map weergave van documenten op mapniveau in Digidoc.**
 Hier is een willekeurige Digidoc-map in de linker filterlijst gekozen. IDOL Find kan dus ook van een map in een bronapplicatie direct een topic map weergave maken met de meest voorkomende onderwerpen in die specifieke map. Bovendien kan op mapniveau ook gebruik worden gemaakt van de lijst-, cirkeldiagram- en tabelweergavefunctionaliteiten van IDOL Find.



- **Topic map weergave van subclusters op mapniveau in Digidoc.**
 Als vervolg op het vorige resultaatbeeld, kan een gebruiker door te klikken op de hoofdclusters inzicht krijgen in de subclusters eronder. Door vervolgens op een subclusters zelf te klikken kan een gebruiker naar de topic map weergave gaan van die subcluster met zijn eigen onderliggende subclusters. Zo kan een gebruiker op een exploratieve manier door de documenten in een bronapplicatie zoeken a.d.h.v. topic map weergaven in IDOL. De gebruiker kan ook het zoekpad naar een gekozen subcluster opslaan als een vaste zoekvraag in IDOL, zoals te zien in de tabjes onder de zwarte balk bovenin. Een gebruiker kan uiteraard ook zo'n vaste zoekvraag maken voor een zoekopdracht met een vaak gezochte term (zoals "nota"). Bij zulke zoekopdrachten kunnen ook topic map weergaven worden gegenereerd waarin de gebruiker verder exploratief of juist gericht kan blijven zoeken en deze vervolgens weer als een vaste zoekvraag opslaan. N.B. vaste zoekvragen zijn niet permanent en kunnen verwijderd worden als een gebruiker ze niet meer nodig heeft.



- Hierboven volgen nog een set van twee resultaatbeelden van een topic map weergave op mapniveau met hun onderliggende subclusters. Hier is te zien dat zelfs als IDOL een hoofdcluster maakt met een algemeen onderwerp, zoals “oog op het verleden”, kunnen er toch handige subclusters worden gevonden zoals “Archivering gericht op de toekomst”, “dossiersvorming en selectie en waardering” en “Implementatie integrale zoekfunctie”. Aan de hand hiervan kan een gebruiker gaandeweg zelfs nieuwe verbanden in de onderwerpen van de documenten in een bronapplicatie opsporen.

7 Evaluatie pilot

7.1 Antwoorden op onderzoeksvragen

De antwoorden op de onderzoeksvragen voor de pilot zijn als volgt:

Kan IDOL Find:

- De Rijksoverheid voorzien in (snel) overzicht en inzicht krijgen in zijn IHH?
 - Ja, aan de hand van de verschillende weergaven die IDOL Find kan genereren in een oogopslag.
 - N.B. in IDOL Find is het ook mogelijk om trendanalyses van de documenten te doen. Zo kan er o.a. op een tijdlijn worden weergegeven wanneer documenten met een bepaalde term gecreëerd zijn en het meest terugkomen. Deze functionaliteit werd voor deze pilot buiten beschouwing gelaten, want het levert weinig nieuwe inzicht op wat de andere weergavevormen al leveren.
 - Ook kan IDOL Find in multimedia zoeken, maar in de dataset voor deze pilot kwamen alleen foto's naar voren. Bovendien heeft IDOL Find een actieve aansluiting met de bronapplicaties nodig om de multimediate bestanden te openen. Vanwege de gekozen workaround voor deze pilot was dit dus niet mogelijk, omdat IDOL Find zelf niet aangesloten was met de bronapplicaties.
 - IDOL Find blijkt ook vergelijkbare documenten te kunnen opsporen. In een lijstweergave kan een gebruiker met de knop "Similar document" suggesties krijgen voor vergelijkbaar gevonden documenten in de bronapplicaties. Deze functionaliteit werd niet verder onderzocht, want het overlapt met de andere BODO-pilot bij Fin "Zoeken & Opruimen".
- Automatisch een (begin-)overzicht en inzicht genereren?
 - Ja, dit kan direct worden gegenereerd in een topic map weergave bij het zoeken naar informatie.
- Van dezelfde data verschillende overzichten genereren afhankelijk van de analysebehoefte?
 - Ja, IDOL Find kan weergaven genereren in de vorm van topic maps, tabellen, lijsten en sunbursts (cirkeldiagrammen).
- Op de metadata van documenten zoeken?
 - Ja, IDOL Find kan zowel in de metadata als de tekst van documenten zoeken. In IDOL Find kunnen metadata ook als filters worden geconfigureerd. Dus er kan bijvoorbeeld gefilterd worden op een metadata zoals auteur en vervolgens gezocht worden naar een specifieke auteursnaam zoals "Jan Jansen".
- Werken met meerdere typen bronapplicaties?
 - Ja, dit was bewezen met de Digidoc en de netwerkschijven van SZW.
- Werken met verschillende bestandsformaten?
 - Ja, IDOL Find kan verschillende bestandsformaten herkennen en full tekst erin zoeken.

- Voldoet aan de informatiebeveiligingsvoorwaarden van de Rijksoverheid?
 - Ja, als IDOL Find op een ICT-platform van de Rijksoverheid wordt geland:
 - Die geen verbinding of aansluiting heeft met externe bronnen of het openbare internet.
 - Waarop de toegang van gebruikers wordt gecontroleerd.
- Op de grote hoeveelheid data van Rijksoverheid worden geschaald?
 - Onbekend. Vanwege de gekozen workaroud was dit niet mogelijk te onderzoeken.

7.2 Bevindingen pilot

1. Op 14 december 2022 was een demo van de resultaten van de pilot gehouden in de BODO-klankbordgroep. Daaruit kwamen positieve reacties op IDOL Find. Men deelde de indruk dat IDOL Find ervoor lijkt te zorgen dat een organisatie geen aanvullende derde partij applicatie meer nodig heeft om de informatie in zijn bronapplicaties in verschillende vormen te visualiseren. Bovendien hoeft de data niet meer gekopieerd te worden naar een derde partij applicatie wat meerdere instanties van eenzelfde data creëert. Dit zal bijdragen aan het doel van de Rijksoverheid om beter grip en overzicht te krijgen van zijn informatiehuishouding.
2. Uit de klankbordgroep is de behoefte geuit of de submappen van een Digidoc-map als de subclusters van een hoofdcluster in IDOL Find kunnen worden getoond. Want, dan zou een gebruiker ook zo gaandeweg door de mappenstructuur van Digidoc kunnen zoeken a.d.h.v. een visuele indeling hiervan met de topic map weergavenfunctionaliteit. In theorie is dit mogelijk, want IDOL Find kan in de metadata van documenten zoeken en de mapnaam als attribuut ophalen. Echter, het genereren van de clusters in topic map weergaven wordt gedaan o.b.v. de meest voorkomende termen of tekststukken in zowel de tekst als de metadata van de documenten. Hierdoor, is het in de praktijk hoogst onwaarschijnlijk dat de clusters automatisch worden gegeneerd o.b.v. mapnamen. Want, termen in de tekst van documenten in een willekeurige map zouden statistisch altijd vaker voorkomen dan de mapnaam zelf.
Een manier om toch door de mappenstructuur van een bronapplicatie te kunnen bewandelen in IDOL Find is via de filteropties. Nieuwe filters kunnen geconfigureerd worden o.b.v. de mappenstructuur van een bronapplicatie; net zoals gedaan is bij deze pilot. Hiervoor moet de mappenstructuur van de bronapplicatie wel eerst in kaart worden gebracht door de experts en/of leverancier zelf. Met deze kan de IDOL Find leverancier de nieuwe filters configureren.
3. Het is nog de vraag hoe de meerwaarde van deze pilot ingebed kan worden in de bestaande IHH-processen van de departementen en IHH-diensten van uitvoeringsorganisaties zoals Doc-Direkt. Uit de klankbordgroep is o.a. de mening gedeeld dat het toepassingsgebied van IDOL Find eerder lijkt te liggen op het vlak van WOO-verzoeken en parlementaire enquêtes dan bij overbrenging; wat initieel de aanleiding van alle BODO-pilots is. Ook is de mening gedeeld dat IDOL Find mogelijk een goede aanvulling zou kunnen zijn op de bestaande Zoek & Vind functionaliteiten zoals de “virtuele dossiers”. De opgehaalde ongestructureerde data uit bronapplicaties én de opgeslagen zoekopdrachten in de virtuele dossiers in Zoek & Vind zouden met IDOL Find visueel in kaart kunnen worden gebracht om zo verder geanalyseerd te worden. Het BODO-eindrapport die RDDI zal opleveren zal als een knip hierop kunnen fungeren om de resultaten van deze pilot beter te kunnen positioneren in relatie tot de andere BODO-pilots, bestaande IHH-processen en IHH-diensten binnen het Rijk.

4. Verder is lichtelijk uitgestippeld dat een eventueel implementatietraject van IDOL Find mogelijk is als er aan bepaalde voorwaarden wordt voldaan zoals een toegewijde omgeving voor IDOL Find en een doorontwikkeling ervan onder de vlag van Doc-Direct. Er zou ook rekening gehouden moeten worden met de technische implicaties voor het verbinden en aansluiten van de bronapplicaties (die op meerdere ICT-platforms staan) met IDOL Find. Er moet ook een onderzoek worden gedaan naar de schaalbaarheid en prestatie van deze oplossing op een gekozen ICT-platform; impactanalyses en performancetests. Daarnaast moet er ondersteuning zijn van experts van de bronapplicaties (beheerders, eigenaren of gebruikers) om mee te helpen bij de configuratie van IDOL Find. Want, hoe beter de configuratie van IDOL Find, hoe beter de gegenereerde topic map weergaven.

8 Conclusie

Terugkijkend naar de doelstellingen van deze pilot is te concluderen dat het genereren van automatische overzichten en inzichten met behulp van IDOL Find een waardevolle tool kan zijn voor de Rijksoverheid t.b.v. IHH. IDOL Find biedt functionaliteiten om in een oogopslag overzichten in verschillende vormen weer te geven, waardoor de Rijksoverheid sneller aan haar IHH-verplichtingen en taken kan voldoen. Overzichten kunnen weergegeven worden in de vorm van topic maps, tabellen, sunbursts (cirkeldiagrammen) en lijsten. Bovendien kan IDOL Find werken met verschillende bronapplicaties én kan IDOL Find full tekst zoeken in (zowel de tekst als metadata van) documenten van verschillende bestandsformaten.

9 Aanbevelingen

Voortbordurend op de gedeelde mening uit de BODO-klankbordgroep, zou Doc-Direkt aan de hand van de resultaten van deze pilot de toekomst van de implementatie van IDOL Find binnen de Rijksoverheid kunnen bepalen. Doc-Direkt zou (eventueel samen met o.a. zijn klanten en RDDI) terug naar de tekentafel moeten gaan om te onderzoeken en bepalen waar in de bestaande diensten en werkprocessen IDOL Find van toegevoegde waarde kan zijn. Uit de BODO-klankbordgroep is al het idee gedeeld dat IDOL Find van toegevoegde waarde kan zijn bij WOO-verzoeken, parlementaire enquêtes en de virtuele dossiers in Zoek & Vind. Daarnaast, is goed te bedenken dat de trendanalysefunctionaliteit van IDOL Find van toegevoegde waarde kan zijn bij andere specifieke IHH-doelstellingen volgens "het generiek actieplan Open op Orde". Hoe dit in de bestaande en/of nieuwe IHH-diensten en processen zou kunnen passen, zou verder onderzocht moeten worden.

Een mogelijke implementatietraject van IDOL Find zou er als volgt uit kunnen zien:

- *Inschatting complexiteit implementatietraject IDOL Find.*

Voor zowel op Rijksbreed als departementenniveau zou er vóór het beginnen met een implementatietraject de slagkans ervan kunnen worden ingeschat aan de hand van:

 - Een analyse of IDOL Find (meta)data uit de gekozen bronapplicaties kan extraheren. Sommige bronapplicaties bevatten zelf niet alle te extraheren data die in IDOL Find geïndexeerd moet worden. De data wordt geroepen uit andere bronnen. Voor zulke bronapplicaties zou een maatwerkoplossing moeten worden gebouwd zodat IDOL Find met elke onderliggende gekoppelde bron ook aangesloten kan zijn.
 - Een verkenning naar de bereidheid van de leveranciers van de gekozen bronapplicaties om ondersteuning te bieden bij mogelijke ingewikkelde aansluiting, verbinding- en toegangsvraagstukken.
 - Op basis van de eerdergenoemde punten én prioriteit een selectie maken van de bronapplicaties die meegenomen worden in het implementatietraject. Er wordt gesuggereerd om klein te beginnen met een beperkt aantal bronapplicaties (bijvoorbeeld 3, 5 of 10), zodat de doorlooptijd van het implementatietraject ook niet te lang wordt. In de loop der tijd kunnen ook nieuwe of aanvullende bronapplicaties sneller worden aangesloten aan IDOL Find a.d.h.v. de eerder opgedane ervaring bij de eerst implementatietraject.
- *Keuze ICT-platform en landen IDOL Find.*

Er moet een ICT-platform worden gekozen:

 - Waarop IDOL Find kan landen.
 - Die een toegewijde omgeving voor IDOL Find kan bieden.
 - Die aan te sluiten is met de gekozen bronapplicaties.
 - Die kan voldoen aan de minimale kritische technische randvoorwaarden van alle gekozen bronapplicaties.
 - Die de schaalbaarheid van IDOL Find oplossing voornamelijk kan vergroten tot de behoefte van alle betrokken organisaties.
 - Die een leverancier heeft die ondersteuning biedt:
 - Voor het landen van IDOL Find,
 - Voor het uitvoeren van testen (zoals performancetesten, penetratietesten, e.d.)
 - Eventueel bij het aansluiten en verbinden met bronapplicaties die op andere ICT-platforms staan.

- *Het verbinden en aansluiten van IDOL Find met alle gekozen bronapplicaties.*
Hier zou de leveranciers van het ICT-platform, de bronapplicaties en IDOL Find nauw moeten samenwerken om dit voor elkaar te krijgen.
- *Het configureren van IDOL aan de gebruikerswensen voor alle gekozen bronapplicaties.*
Hierbij zou de leverancier van IDOL Find de stopwoordenlijst en filteropties voor elke gekozen bronapplicaties moeten configureren in samenwerking met experts van de bronapplicaties. Deze experts kunnen de IDOL Find leverancier helpen bij het bepalen van irrelevante woorden die niet meegenomen moeten worden in de clusters van de topic map weergaven én de juiste metadata instellen als filters.

Dit is een uitgave van:

Rijksprogramma Duurzaam Digitale
Informatiehuishouding (RDDI)

Rijnstraat 50
Postbus 16375
2500 BJ Den Haag

dashboardcoachihh@minocw.nl
www.informatiehuishouding.nl

Maart 2023 - Versie 1.0