

Marktconsultatievragen Vervolg Leveren Automatische Metadata (VLAM)

Onderdeel 1: Situatieschets eerste fase

In een eerste fase wil de KB zich richten op het registreren van de ontvangst van de stroom van boeken die via giften of op een andere manier nog niet “aangekondigd” zijn. Op hoofdlijnen volgt dat proces de volgende onderstaande stappen:

1. Voorbereiden

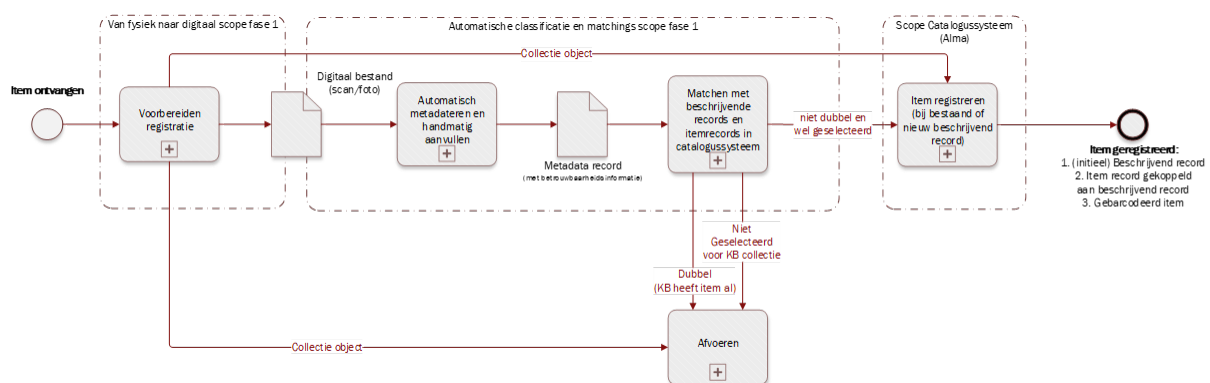
Het item moet worden ontvangen en voor een aantal kenmerkende pagina's (bv. colofon) gescand. De output is een digitaal bestand wat opgeslagen wordt.

2. Metadateren

De scan van het materiaal wordt automatisch geclassificeerd op meerdere standaard beschrijvende metadatavelden. Tevens wordt dit handmatig gecontroleerd en aangevuld (bv. op velden die niet af te leiden zijn uit een scan). De output is een bestand met beschrijvende metadata van het item.

3. Matchen

In deze stap wordt gekeken op basis van data uit het catalogussysteem of er een “match” is. Dit proces zal deels automatisch en waar nodig ook deels handmatig worden uitgevoerd. De uitkomst is driedelig: het item is al in het bezit of is niet geselecteerd om opgenomen te worden en kan dus worden afgevoerd. Als dit niet het geval is betreft het een nieuw item dat in het catalogussysteem opgenomen moet worden voor verdere verwerking. De output van dit proces is dus metadata aangeleverd aan het catalogussysteem voor verdere verwerking.



Het hele proces (en dus ook de bijbehorende workflow) zal in de eerste fase ondersteund moeten worden door de applicatie(componenten), waarbij in het bijzonder de stappen 2 en 3 zo goed mogelijk ondersteund moeten worden door kunstmatige intelligentie.

Nummer	Vraag
1.1	Is het proces van fase 1 helder genoeg om een goed beeld te krijgen van de benodigde applicatiefunctionaliteiten voor deze fase? Zo nee, wat zou verhelderd moeten worden om een goed beeld te krijgen?
1.2	Welke mogelijkheden voor welk(e) type(n) van kunstmatige intelligentie ziet u binnen de verschillende stappen?
1.3	Zou uw voorkeur uitgaan naar “supervised” of “unsupervised” learning algoritmes hiervoor in te zetten en waarom?

1.4	Ziet u standaard op de markt verkrijgbare software die hierbij ingezet zou kunnen worden, zo ja welke en waarvoor? Wat zijn daarvan de voor en/nadelen?
1.5	Is de aard van de oplossing in uw ogen meer integratie van verschillende software componenten of meer het ontwikkelen van maatwerk software? Waarom?
1.6	Is het uitvragen van deze gecombineerde functionaliteit (zie inleidende tekst) in uw ogen logisch? Waarom wel/niet?
1.7	Welke competenties van een leverancier zijn in uw ogen essentieel voor de uitvoering van deze fase?
1.8	Wat zijn aandachtspunten die u ziet bij het uitvragen van deze functionaliteit, wat is hierbij essentieel om te beschrijven?

Onderdeel 2: Doorontwikkeling

Na de eerste fase wil de KB de functionaliteiten van de oplossing uitbreiden en tevens ook voor andere itemstromen en processen gaan gebruiken. Daarnaast willen we ook meer gaan aansluiten bij initiatieven die de publieke waarden van de KB beter vertegenwoordigen.

Functionaliteiten

Te denken valt aan het metadateren niet alleen op basis van een scan, maar ook op basis van andere data, zoals al de reeds beschikbare data in het catalogussysteem en andere (externe) thesauri, registers en datasets/bronnen.

Itemstromen

Naast boeken zou de oplossing ook kunnen worden ingezet voor andere typen items zoals kranten, tijdschriften, E-boeken, enz.

Processen

Fase 1 bestaat alleen het proces van de nieuwe instroom van boeken. Delen van de functionaliteit zouden ook goed bruikbaar kunnen zijn voor andere processen, zoals het verrijken van de al bestaande metadata in onze catalogus of het bijhouden van selectielijsten van items die we wel/niet willen opnemen.

Publieke waarden en digitale soevereiniteit

De KB hecht als publieke organisatie groot belang aan het onderschrijven van publieke waarden in onze dienstverlening, zie <https://communities.surf.nl/publieke-waarden/artikel/digitale-soevereiniteit-autonomie-en-publieke-waarden-wat-bedoelen-we>. Hieronder valt ook digitale soevereiniteit en autonomie. De KB wil dan ook graag gebruik gaan maken van (publieke) initiatieven die deze waarden beter borgen dan de gangbare Cloudleveranciers wanneer deze productierijp beschikbaar komen (bv. GPT-NL, Dutch AI Factory).

Nummer	Vraag
2.1	Ziet u eventueel op fase 1 aanvullende software(componenten) die bij deze geschetste doorontwikkeling van toepassing zouden kunnen zijn?
2.2	Vraagt deze geschetste doorontwikkeling andere en/of aanvullende competenties van een leverancier dan voor fase 1?

2.3	Hoe ziet u de markt zich verder ontwikkelen w.b. AI en andere automatiseringen voor dergelijke toepassingen in de komende jaren? Wat zijn daarvan de gevolgen voor de markt waarin u opereert?
2.4	Hoe kijkt u aan tegen het inzetten op publieke waarden en digitale soevereiniteit en wat zou hiervoor in het kader van deze aanbesteding van belang kunnen zijn?
2.5	Heeft in uw ogen het inzetten op publieke waarden en digitale soevereiniteit ook nog consequenties op de benodigde competenties van de leverancier?
2.6	Welke aanpak en wat voor proces zou u de KB willen adviseren met betrekking tot de doorontwikkeling van de oplossing?

Onderdeel 3: Infrastructuur

De KB is voornemens bij aanvang van dit traject de applicatiecomponenten binnen de huidige Cloudinfrastructuur van de KB te laten implementeren en beheren door de leverancier. Op dit moment is deze Cloudinfrastructuur Microsoft Azure-omgeving wat binnen het SURFcumulus (onderwijs)contract wordt afgenomen. Hierbij wordt bij voorkeur van een zo hoog mogelijk PaaS niveau gebruik gemaakt. Zoals in het vorige onderdeel geschetst, zou de KB graag na deze eerste fase stapsgewijs overgaan naar (Cloud)infrastructuur die de publieke waarden en digitale soevereiniteit beter onderschrijft.

Nummer	Vraag
3.1	Zou u het gebruik van de Cloudinfrastructuur, en specifiek Microsoft Azure, voor deze oplossing aan/ontvragen en waarom?
3.2	Zou u het gebruik van de Azure-omgeving van de KB aan/ontvragen en waarom?
3.3	Bent u ook in staat om de dienstverlening op de Azure Cloudomgeving van de KB te implementeren en te onderhouden? Zo nee, wat is hiervoor de belemmering? Zo ja, welke aandachtspunten ziet u daarbij nog?
3.4	Ziet u meerwaarde in de Azure PaaS componenten (zoals de Microsoft Azure OpenAI Service) of ziet u hierbij ook belemmeringen? Welke Azure-componenten zou u aanraden te gebruiken?
3.5	Welke programmeertaal zou u aanbevelen voor de maatwerksoftware?
3.6	Wat zijn in uw ogen de belangrijkste aandachtspunten waar de uitvraag en de gevraagde oplossing aan moeten voldoen om in te kunnen zetten op alternatieve dienstverlening (ten opzichte van Azure) ten behoeve van de publieke waarden en digitale soevereiniteit?
3.7	Welke relevante initiatieven voor dergelijke alternatieve dienstverlening (die de publieke waarden en digitale soevereiniteit beter onderschrijven) ziet u op dit moment en wat zouden die op termijn bij kunnen dragen aan de gevraagde oplossing?

Onderdeel 4: Algemene vragen

Nummer	Vraag
4.1	Welke (overige) aandachtspunten wilt u de KB meegeven met betrekking tot een Programma van Eisen voor de toekomstige aanbestedingsprocedure?
4.2	Met betrekking tot de gunningscriteria van de aanbesteding: op welke wijze zou u zich als inschrijvers het liefst willen onderscheiden? Oftewel, welke onderwerpen vindt u passend voor de gunningscriteria van een dergelijke aanbesteding?
4.3	Welke (overige) aandachtspunten wilt u de KB meegeven met betrekking tot de Overeenkomst?
4.4	De KB is voornemens om de ARBIT 2022 van toepassing te verklaren op de te sluiten Overeenkomst. Zijn er voor u knelpunten met betrekking tot het toepassen van de ARBIT 2022?
4.5	De KB heeft het voornemen om bij de toekomstige aanbesteding de volgende prijssystematiek toe te passen: 1) Fase 1: prijsaanbieding op basis van een vaste aanneemsom voor de ontwikkeling van de functionaliteit; 2) Jaarlijkse kosten voor (technisch) beheer van de oplossing; 3) Doorontwikkeling: uitvraag op basis uurtarieven (op basis van fictieve aantallen). Vindt u deze systematiek logisch? Zo nee, wat zou u dan willen voorstellen?