



# 1<sup>e</sup> Nota van Inlichtingen SBIR Historische transcripties met AI voor Nationaal Archief 2019

Versie 11 september 2019

Deze nota bevat de vragen die tijdens de informatiebijeenkomst zijn gesteld en via e-mail zijn binnengekomen met de antwoorden die daarop zijn gegeven of naderhand geformuleerd zijn.

## A. Vragen over de SBIR algemeen en het proces

	Vraag	Antwoord
1.	Waar vind ik de meest recente informatie over de SBIR?	De meest up-to-date informatie vindt u op <a href="https://mijn.rvo.nl/sbir-innovatie-in-opdracht">https://mijn.rvo.nl/sbir-innovatie-in-opdracht</a>
2.	Moet de offerte in het Nederlands worden ingediend of mag dit ook in het Engels?	De offerte moet in het Nederlands worden ingediend, met uitzondering van het business plan, dat mag ook in het Engels.
3.	Moet de managementsamenvatting apart als bijlage worden aangeleverd of als onderdeel van het projectplan?	Bij voorkeur de managementsamenvatting bijvoegen in een apart document.
4.	Mogen we binnen een consortium samenwerken met buitenlandse partijen?	Ja, dat mag. Zolang de indiener van de offerte maar aan de vereisten voldoet (zie SBIR-handleiding over eisen aan de partijen onder 10.1).
5.	Is het mogelijk om samen te werken met een kennisinstelling?	Ja, dit is mogelijk. Maak met elkaar wel goede afspraken over eigendom van kennis.
6.	Moeten we het format aanhouden van de SBIR?	In verband met het competitieve karakter van de SBIR, zijn het format en het maximaal aantal woorden bindend voorgeschreven om voorstellen eerlijk met elkaar te vergelijken.
7.	Blijft het IP van de ontwikkeling behouden voor de ondernemer die het ontwikkeld heeft met een SBIR-opdracht?	De opdrachtnemer behoudt het intellectueel eigendom behoudens de in paragraaf 4.2 van de handleiding genoemde situaties  NB: Dat is anders bij het traject dat op de SBIR volgt: Het (op basis van 1 prototype) te ontwikkelen eindresultaat is van het NA.
8.	Van wie is het prototype aan het eind van de ontwikkeling?	Het prototype is van de ondernemer die het gebouwd heeft.
9.	In hoeverre kan je het Intellectueel eigendom	Voor deze en andere vragen over intellectueel Eigendom gaat u naar de website <a href="http://www.rvo.nl/octrooien">www.rvo.nl/octrooien</a> of u belt met Yp Kroon op nummer 088-602 6425

	Vraag	Antwoord
	beschermen als je open-source eist?	
10.	Er worden 8 of 9 indieners uitgenodigd om een pitch te geven. Hoe vindt de selectie hiervoor plaats?	Anders dan in de Handleiding voor Ondernemers staat, kan er bij deze SBIR wel voorgeselecteerd worden in fase 2. (Fase 2 is hier immers de eerste fase) Voorselectie wordt gedaan door medewerkers van het Nationaal Archief (NA) en ook door medewerkers van de Rijksdienst voor ondernemend Nederland (RVO). Voorselectie kan gedaan worden op grond van de managementsamenvatting.
11.	Wie zitten er in de beoordelingscommissie?	De volgende personen nemen zitting in de beoordelingscommissie: 1. Ivo Zandhuis (Consultant Cultural Heritage & ICT) 2. Annelot Vijn (Utrechts Archief) 3. Liesbeth Keijser (Nationaal Archief) 4. Jan-Willem Hiddink (RVO/EZK) Het beïnvloeden van deze personen kan tot uitsluiting leiden.
12.	Waarom is er een criterium <b>'Economisch Perspectief'</b> ? En wat is het doel van het maken van een Business plan?	Binnen een SBIR-competitie is het van belang dat ondernemers zelf de ontwikkelde producten/diensten gaan vermarkten. Met SBIR creëert de overheid een aanbodzijde van de markt waarop zij zelf wil gaan inkopen. Het Nationaal Archief is echter niet de enig mogelijke (potentiele) inkoper, ook al heeft het NA aangegeven graag met een partij het prototype door te ontwikkelen.
13.	Waarom is het nodig om een businessplan op te stellen, het gaat toch om het ontwikkelen van een prototype voor een overheidsorganisatie?	Een SBIR offerte voor fase 2 betreft (nagenoeg) altijd het ontwikkelen van een prototype voor of ten behoeve van een overheidsorganisatie. Het enige dat in deze specifieke oproep anders is, is het gegeven dat het Nationaal Archief reeds in een vroeg stadium heeft aangegeven heel graag één van de afnemers van het eindproduct te willen zijn en daar een budget voor heeft gereserveerd. Dat maakt een businessplan niet overbodig. Indien u geen businessplan hebt ingediend, dan wordt uw voorstel niet in behandeling genomen.
14.	Een offerte moet op alle drie de criteria 60% of meer van het maximaal te behalen aantal punten scoren. Dus dat geldt dus ook voor het criterium Economisch Perspectief (minimaal 6 van de 10 punten). Wanneer scoort een voorstel een onvoldoende op het criterium economisch perspectief?	Aan de hand van het businessplan en de betreffende paragraaf uit het projectplan legt u onder andere uit waarom uw bedrijf (met eventuele partners) dit project tot een commercieel succes kan maken.  Indien u slechts inzet op het verkrijgen van de vervolgopdracht tot het verder doorontwikkelen van het prototype zoals genoemd in de oproep, dan is dat onvoldoende.  Indien u verdere ideeën voor het vermarkten van uw product/dienst enigszins heeft uitgewerkt scoort u al snel een voldoende op dit criterium
15.	Hoe verhoudt de wens zoveel mogelijk open source mogelijkheden toe te passen	Het is aan de ondernemer om daar een passend antwoord op te vinden.

	Vraag	Antwoord
	zich tot een goed verdienmodel.	
16.	In het begrotingsformulier wordt er ook gevraagd naar het Hypothetisch offertebedrag exclusief btw, wanneer alle rechten aan opdrachtgever overgedragen zouden worden. Wat betekent dat?	Het is een (geschat, fictief) bedrag dat aangeeft wat het project zou kosten, in het theoretische geval dat de partijen het uitvoeren zonder er zelf een belang bij te hebben. Het opgeven van het bedrag dient om te kunnen laten zien dat het gevraagde bedrag geen commercieel bedrag is aangezien het om een pré commercieel traject gaat.
17.	Over het algemeen, wat is het bedrijfsmodel waar jullie naar op zoek zijn?	Wij zijn niet op zoek naar een bepaald bedrijfsmodel, het is aan de ondernemer om zelf het meest passende model te beschrijven.
18.	Is het mogelijk om het project deels met eigen middelen of andere cofinanciering te financieren als de projectomvang groter is dan maximale bedrag van de SBIR?	Ja, dat kan. U krijgt van ons een opdracht met het maximale SBIR-bedrag. U bent dan verder zelf verantwoordelijk voor de overige financiering en de te bereiken projectresultaten. Geeft u wel duidelijk aan in uw voorstel wat de hoogte is van de ingediende SBIR-offerte (en dit bedrag dient dus niet hoger te zijn dan het maximale offertebedrag dat in de oproep is vermeld).
19.	Zijn er vergelijkbare projecten die jullie doen?	Het Nationaal Archief neemt ook deel aan het project TRIADO waarbij archief wordt ontsloten met onder andere optical character recognition. ( <a href="https://www.oorlogsbronnen.nl/tribunaalarchieven-als-digitale-onderzoeksfaciliteit">https://www.oorlogsbronnen.nl/tribunaalarchieven-als-digitale-onderzoeksfaciliteit</a> ),
20.	Hoe is de samenwerking met andere eigenaren/beheerders van grote archieven?	Er wordt nauw samengewerkt met andere archiefbeheerders van de Regionaal Historische Centra voor zowel het scannen als de opslag en het beheer van de digitale archieven.
21.	Waarom is er binnen dit project geen samenwerking met de Koninklijke bibliotheek, en het programma Delpher, gelegd?	Er is zeker gesproken met de Koninklijke Bibliotheek over de mogelijkheden aan te sluiten bij Delpher. Delpher sluit niet aan bij de eisen en wensen van het Nationaal Archief. We hebben nu meer de kans onze vragen te specificeren.
22.	Staat vast dat één van de drie partijen die de opdracht krijgen om een prototype te ontwikkelen ook de opdracht van EUR 385.000 krijgt voor het verder doorontwikkelen?	Het Nationaal Archief beoogt om 1 van de 3 partijen een opdracht te geven voor doorontwikkeling.  Dit staat echter niet vast. Als de kwaliteit van de drie prototypes te laag is, dan krijgt geen van de drie partijen deze opdracht tot doorontwikkelen.  Theoretisch gezien kan het zijn dat er in de ontwikkelperiode (fase 2) nieuwe marktpartijen met innovatieve prototypen op de markt komen. In voorkomend geval zou het kunnen dat deze marktpartijen ook meedingen naar een opdracht tot doorontwikkeling (fase 3).
23.	Indien de vervolgoopdracht openbaar aanbesteed zal worden: In hoeverre is het reëel dat deze 'gewonnen' zal	Dit valt niet in te schatten.

	Vraag	Antwoord
	worden door een partij die niet deel heeft genomen aan de SBIR?	
24.	Wordt fase 3 via een open aanbesteding uitgevoerd?	Fase 3 (inkoop) valt buiten de scope van de SBIR. De opdracht tot doorontwikkelen wordt verstrekt op basis van een aanbestedingsprocedure.  Welke aanbestedingsprocedure wordt gevolgd, is in dit stadium nog niet bekend. Naar verwachting de 3 partijen uitgenodigd voor deelname aan de aanbesteding.
25.	Op basis van welke criteria wordt de offerte voor doorontwikkeling (fase 3) beoordeeld?	De beoordelingscriteria voor fase 3 worden na het afronden van het SBIR traject (fase 2) vastgesteld.
26.	Hoe kan ik partners vinden voor samenwerking in deze SBIR?	De aanmeldingen voor de voorlichtingsbijeenkomst komen niet openbaar beschikbaar.  Wordt lid van de SBIR Linked-in groep: <a href="https://www.linkedin.com/groups/8675827/">https://www.linkedin.com/groups/8675827/</a>
27.	We hebben een samenwerking tussen drie partijen, waarvan ik de kleinste ben, maar wel de projectleider ben. Moet ik van alle drie partijen jaarcijfers etc aanleveren?	Nee, dat is niet nodig. Dit hoeft alleen voor de hoofdaannemer (indiener) van de offerte.
28.	In het model projectplan moeten een aantal dingen worden ingevuld mbt Haalbaarheidsonderzoek, maar er is geen fase 1, hoe zit dat?	Deze oproep kent inderdaad geen fase 1. Het Nationaal Archief heeft onlangs uitgebreid voorbereidend onderzoek gedaan. Deze vragen kunt u overslaan.
29.	wordt bij deze tender ook naar hoogste en laagste bod gekeken, want er wordt uitgegaan van een vast budget.	Nee, prijs is geen criterium. De prijs-kwaliteit verhouding wordt wel meegenomen onder het criterium impact.
30.	Er is maximaal 66.666 euro per prototype beschikbaar. Er staan inhoudelijke beoordelingscriteria op de website voor het prototype, maar niet hoe de betaling van het opgeleverde prototype hier vanaf hangt. Wanneer wordt er uitbetaald?	Paragraaf 7.1 van de handleiding geeft hier meer informatie over en in de bijlage van de handleiding is het model overeenkomst opgenomen. Voor deze SBIR geldt: 40% bij opdrachtverlening, 40% bij ontvangst voortgangsrapportage en 20% na demonstratie prototype en goedkeuring eindrapportage. Indien u zich onverhoopt niet aan de overeenkomst houdt, dan staan daar privaatrechtelijke rechtsmiddelen RVO ter beschikking.
31.	Wij zijn een net opgericht nieuw bedrijf. We hebben als organisatie dus nog niet echt een 'track-record' opgebouwd. In hoeverre wordt dit meegenomen in de	Ja, we moedigen start-ups aan om mee te doen met deze innovatie-competitie. Een goede oplossing is belangrijker dan een lang track-record. Maak duidelijk in het voorstel dat jullie de juiste partij zijn om dit te ontwikkelen. Let op alle uitleg van de drie criteria, dus

	Vraag	Antwoord
	beoordeling (en maken we dus een reële kans)?	zowel die uit de Handleiding (hier met name 2b en 3b) als die uit de oproep.
32.	Dit is een offerte voor fase 2. Er moet een SBIR nummer offerteverzoek vermeld worden. Waar kan ik dat SB nummer vinden?	Die heb je niet, dus dat vakje mag je leeglaten.

## B. Vragen over de dataset

	Vraag	Antwoord
33.	Is er van de toegangen (beschrijvingen van archieven) ook een XML bestand beschikbaar?	ja, die zal worden toegevoegd aan het datapakket.
34.	Waar kan ik de dataset vinden van automatisch getranscribeerde scans?	Die kunt u hier vinden: <a href="http://sbir.datapakket.nl/Datapakket/">http://sbir.datapakket.nl/Datapakket/</a>
35.	Waar kan ik de transcripties vinden in de dataset?	Dat is per export van een inventarisnummer het <b>mapje met de name 'Page'</b> . Bijvoorbeeld <a href="http://sbir.datapakket.nl/Datapakket/scan-page.export.job_599574/168998/NL-HaNA_1_04_02_9511/page/">http://sbir.datapakket.nl/Datapakket/scan-page.export.job_599574/168998/NL-HaNA_1_04_02_9511/page/</a> . Er zal aan het datapakket ook een export worden toegevoegd in ALTO XML waarbij de coördinaten van woorden bij benadering zijn aangegeven.
36.	De dataset van de transcripties is een Page format waarbij geen coördinaten op woord niveau beschikbaar zijn. Om zoekresultaten te kunnen highlighten is dat wel nodig.	De Page export uit Transkribus geeft inderdaad geen coördinaten op woordniveau mee. We zullen van deze dataset eveneens een export maken in het format ALTO XML waarbij wel de coördinaten van woorden bij benadering zijn aangegeven. Tijdens de ontwikkeling van de prototypes moet worden bepaald welk format gebruikt zal worden. zie vraag 67 en 98
37.	Er is sprake van 6,15% fouten, maar er is ook een voorbeeld met 10% fouten (in de presentatie van het Nationaal Archief). Welk percentage is het?	De 10% is nog een oud percentage, inmiddels heeft er een verbetering plaatsgevonden naar 6,15% CER op de testset.
38.	Er wordt een precisiepercentage genoemd van 80%. Klopt dit nog (ook ten opzichte van de foutmarge van 6,15%)?	In het projectplan hebben we inderdaad gesteld dat we streven naar een accuraatheid van het herkennen van de karakters van 80%. We zijn daarmee heel voorzichtig geweest. Inmiddels behalen we dus 6,15% op de testset.
39.	Er worden 2 miljoen archiefstukken getranscribeerd. Geldt hiervoor die 6,15%? Zijn al deze stukken gelijk behandeld?	De 6,15% CER is bepaald over een beperkte dataset. Van de 2 miljoen scans is een sampleset samengesteld. Het huidige model behaalt een accuraatheid op deze sampleset van 8,77% CER.
40.	Het lijkt er op dat het percentage fouten groter is dan 6,15% (in een getoond voorbeeld in de presentatie). Wat is eigenlijk een fout?	Oude spelling is geen fout en tekst die van een pagina valt is ook geen fout. Deze onderwerpen worden dan ook niet meegenomen in de foutmarge. De fouten in de lay-out worden evenmin als fout aangemerkt door het model.
41.	Is het model met een character error rate (CER) van 6,15 op de testset ook gebruikt voor de automatisch getranscribeerde	Antwoord: Nee. Daarvoor hebben we een eerder model gebruikt met slechtere resultaten. We zullen het datapakket uitbreiden met

	scans die in het datapakket beschikbaar zijn gesteld?	transcripties die zijn gemaakt met het model met een CER van 6,15 op de testset.
<b>42.</b>	Vraag: Kunnen we ook beschikken over de door de mens gemaakte transcripties (ground truth).	Nee, deze zullen vooralsnog niet beschikbaar worden gesteld. Het maken van automatische transcripties is niet in scope van deze SBIR oproep.

### C. Vragen over de studiezaalbezoeker

	Vraag	Antwoord
43.	Is er feedback bekend van bezoekers? Zijn er klachten?	Bezoekers komen vaak met de verwachting dat de archiefstukken die ze zoeken met een druk op de knop zijn te vinden. Bijvoorbeeld hebben ze een naam van een familielid en dan verwachten ze net als bij Google dat ze direct alle relevante resultaten te zien krijgen. Wij zijn er voor om de bezoekers te instrueren en ze uit te leggen hoe het zoeken werkt. Je moet je altijd eerst afvragen in welk archief en in welke map je moet zoeken. Welk Ministerie gaat het over bijvoorbeeld. En dan is er ook het gegeven dat de beschrijvingen heel kort zijn.
44.	Is het gewenste resultaat een soort Google? Dat wil zeggen semantisch zoeken?	Dit is afhankelijk van de bezoeker. Er zijn ook mensen die juist geïnteresseerd zijn in de structuur en in de context van een bepaald archiefstuk. Dan is het belangrijk om dat ook inzichtelijk te maken.
45.	Is er al een doelgroepanalyse gedaan?	Globaal gezien bestaat de doelgroep uit drie groepen: Diepgravers, Ondiepgravers, Grasduiners. Diepgravers hebben vaak kennis van hoe je moet zoeken in de archieven. Onderzoekers, journalisten, scholieren en genealogen zijn afhankelijk van de complexiteit van hun onderzoek en het te bestuderen archief Diepgraver of Ondiepgraver. De Grasduiner is het in geschiedenis geïnteresseerde publiek. Zie ook vraag 48 en 75
46.	In vervolg op de voorgaande vraag. Wat zijn de bijbehorende cijfers?	Exacte aantallen hebben we niet. De grasduiners en ondiepgravers zijn echter de grootste groep. Een deel van de bezoekers komt maar één keer naar het Nationaal Archief. Vaak omdat ze een dossier komen raadplegen van een familielid.
47.	Kunnen we in het voortraject vragen stellen aan eindgebruikers?	Niet rechtstreeks. Stel alle vragen over deze SBIR aan de mailbox <a href="mailto:SBIR@rvo.nl">SBIR@rvo.nl</a>
48.	Zijn er voorbeelden van vragen van bezoekers? Waar zoeken zij typisch naar?	Veel vragen gaan over familieleden een voorouders. We krijgen ook veel vragen in de wie, wat, waar en wanneer trant. Andere vragen kunnen overal over gaan. Een onderzoeker kan echter ook een vraag stellen over het graven van grachten in Batavia om te achterhalen waarom deze grachten uiteindelijk een mislukking werden voor de gezondheid van de inwoners van Batavia. Zo groot als de collectie is, zo divers zijn de vragen die je daarover kan stellen. Zie ook vraag 48 en 75
49.	Welke vragen hebben bezoekers in het kader van de VOC-archieven bijvoorbeeld?	Hierin zijn twee hoofdcategorieën: (a) familieleden die stamboomonderzoek doen en meer willen



	Vraag	Antwoord
		weten en (b) wetenschappers die bijvoorbeeld de besluitvorming binnen de VOC onderzoeken.
50.	Welke restricties zijn er?	Restricties hebben doorgaans te maken met de privacy van nog levende personen. Inzage kan alleen indien er toestemming is of men de persoonlijke levenssfeer beschermd in de verwerking van de gegevens. Er zijn twee soorten beperkingen om stukken in te zien. De A-beperking is een lichte beperking. Bezoekers kunnen aan de balie van de studiezaal een medewerker de stukken laten aanvragen. Stukken met een B-beperking worden afgehandeld in zes weken door de afdeling dienstverlening. Indien aan de voorwaarden (zie privacy) is voldaan kan inzage plaatsvinden. Archieven met oude handschriften zijn doorgaans niet beperkt openbaar. De stukken zijn dusdanig oud dat er geen personen voorkomen in de archieven die nog in leven kunnen zijn.
51.	Hoelang duurt het 'vinden' in het Nationaal Archief.	Dit varieert van 3 minuten tot een week. Soms wordt er ook niets gevonden. Dit heeft niet te maken met onze dienstverlening fysiek dan wel digitaal maar met de complexiteit van de archieven en de vraag die beantwoord moet worden.
52.	Wat is het belang van het zichtbaar maken van de structuur van het archief?	de structuur van een archief kan een onderzoeksonderwerp zijn. Daarnaast helpt de structuur bezoekers te begrijpen waar ze in zoeken. Het is door dit begrip ook makkelijker om verder te zoeken naar verwante archiefstukken.
53.	Zijn er beschrijvingen van foto's beschikbaar?	Ja, eigenlijk is dit de makkelijkste zoekfunctie aangezien onder iedere foto een tekst staat die <b>makkelijk doorzoekbaar is. (Foto's zijn geen onderdeel van de SBIR oproep.)</b>
54.	Wat ligt hier opgeslagen?	We zijn naast een rijksarchief ook een regionaal historisch centrum. Dat betekent dat regionale archieven, bijvoorbeeld rechtbank archieven uit Zuid-Holland ook onderdeel uitmaken van onze collectie. Daarnaast hebben we archieven die belangrijk worden geacht voor de Nederlandse geschiedenis. Daarbij kan je denken aan het archief van de VOC, het Hof van Holland en bijvoorbeeld het archief van de KNVB.
55.	Hoe gaan jullie om met privacy?	Dossiers ontsloten op persoonsnaam kunnen worden ingezien indien de aanvrager (bezoeker) voldoet aan een aantal voorwaarden. Met kan inzage krijgen door aan te tonen dat een persoon inmiddels is overleden, met toestemming en in enkele gevallen middels een onderzoekopzet waarin de aanvrager aantoont de persoonlijke levenssfeer te beschermen in het verwerken van de onderzoekresultaten.

	Vraag	Antwoord
56.	Moeten we ons richten op de VOC-archieven? En hoe verhoudt zich dat tegenover de beperkt openbare collecties?	Het prototype richt zich op oudere handschriften van de VOC van het Nationaal Archief en Notariële Archieven van de Regionaal Historische Centra. Maar het type vraag van bezoekers is hetzelfde, of je iets nu of over het verleden opzoekt, het gaat meestal over wie, wanneer en waar. En ook de getypte, gedrukte en beperkt openbare archieven moeten in de uiteindelijke functionaliteit kunnen worden doorzocht en getoond.
57.	Waarom is het Nationaal Archief opgericht?	Het is een fundamenteel onderdeel van de democratie. Het archief maakt het inzichtelijk hoe een besluit van een minister tot stand is gekomen. Echter tegenwoordig de vraag wat er wel en niet gearchiveerd wordt, denk aan whatsapp/e-mails. Het laatste woord is hier nog niet over gezegd.

## D. Vragen over het archiefstuk

	Vraag	Antwoord
58.	Zijn er alleen teksten beschikbaar of ook bijvoorbeeld financiële documenten?	<p>VOC-archief dat nu wordt getranscribeerd bestaat uit Missiven, resoluties, dagregisters, uitgaande en inkomende stukken van de Gouverneur Generaal in Batavia, Een deel van de stukken zijn van de Raad van Justitie van Batavia. De hoeveelheid financiële informatie in deze stukken is beperkt. Wel is er een stuk boekhouding, Dit is circa 1,5 % van het VOC archief binnen dit project.</p> <p>Notariële teksten zijn heel divers. Een financieel aspect speelt vaak een rol, denk bijvoorbeeld aan een schuldbekentenis, boedelbeschrijving of de eigendomsoverdracht van een huis. En ook om een verklaring af te leggen of een testament op te laten tekenen ging je naar de notaris.</p>
59.	Wat gebeurt er bij transcriptie met symbolen (bijvoorbeeld het VOC-symbool, een handtekening).	Het VOC teken wordt bij het maken van mens-gemaakte transcripties niet getranscribeerd en deletet als testregio. Er is immers hiervoor geen teken beschikbaar op het toetsenbord. Een handtekening wordt indien er letters herkenbaar zijn (deels) getranscribeerd.
60.	Hoe werkt de nummering bij losbladige archiefstukken?	<b>De pagina's van archiefstukken, zowel gebonden als losbladig, zijn soms genummerd.</b> Bij digitalisering verwerken we deze nummering niet in de scannaam. De bestandsnaam is opgebouwd uit verschillende onderdelen: een kenmerk voor het Nationaal Archief, het archief, het inventarisnummer en een volgnummer van de scan. Er is dus geen relatie tussen het nummer op de pagina en het volgnummer van de scan. Bij een aantal archieven hebben we d.m.v. data-entry een relatie gelegd tussen de scan en de paginanummering en kan je eenvoudig naar de juiste scan springen. Als dat niet is gedaan dan moet je bij het zoeken naar een bepaalde pagina op onze website binnen een inventarisnummer handmatig scrollen. (Het toekennen van pagina nummering aan de scan is geen onderdeel van de SBIR oproep). Zie vraag 109
61.	Hoe zit het met de verschillen binnen de VOC-stukken?	Er zijn grote verschillen. Ouder stukken kennen vele diverse handschriften. Jongere stukken zijn door een klerk geschreven waardoor het aantal handschriften kleiner is. De meeste VOC-stukken zijn goed leesbaar met uitzondering van bijvoorbeeld dagboeken. De stukken die geschreven zijn voor de communicatie met Holland zijn vaak heel helder.
62.	Wat doet Transcribus met plaatjes en kaartjes?	Niets, want de meeste zijn er reeds uitgehaald en naar andere collecties verplaatst.
63.	Heeft de betreffende medewerker van het NA deze beschrijvingen gemaakt? Hoe maak je de afweging wat je	Nee, het opstellen van een inventaris, en dus het beschrijven van alle stukken, wordt gedaan door een archiefbewerker.

	Vraag	Antwoord
	meeneemt?	De huidige inventaris van de VOC is vervaardigd in de <b>jaren '70/'80 van de vorige eeuw. Voor dit archief is</b> van belang dat bij de beschrijving van de stukken gekeken is naar de vorm van het archiefstuk (dagboek, kasboek, reisjournaal, losse stukken etc.) en of het stuk deel uitmaakt van een serie stukken die gelijk zijn aan elkaar. In het laatste geval worden stukken onder een verzamelbeschrijving geplaatst en krijgen zij individueel een beknopte beschrijving (bv. alleen een jaartal).
64.	Hoeveel mensen betrokken bij het <b>tranciberen,?</b>	Dat is 3,5 fte. Zeven mensen a circa 20 uur.
65.	Weten jullie wie wat transcribeert?	Ja, in transkribus zie je de verschillende versies en wie die versie heeft gemaakt.
66.	Als test werken er geen twee medewerkers één document uit?	Nee niet in de regel. Wel controleert een coördinator een deel van de transcripties, In het begin intensiever nu minder intensief. (Gemiddeld circa 35% van de transcripties). Ook hebben we een transcriptieconventie om te zorgen dat iedereen op dezelfde manier werkt.
67.	Hebben we de data om woorden te laten oplichten? En Is dit Alto XML ipv page?	zie vraag 36 en 98
68.	Hoe vertalen we woorden in het oud Nederlands? Is daar een database van? woordenboeken, en oud Nederlandse woordenboeken is dit te vertalen.	Ja, de historische woordenboeken zijn digitaal beschikbaar, soms tegen betaling. Ook middels synoniemen woordenboeken, en oud Nederlandse woordenboeken is dit te vertalen. Zie bijvoorbeeld: <a href="http://gtb.inl.nl/search/?owner=WNT">http://gtb.inl.nl/search/?owner=WNT</a> <a href="http://resources.huygens.knaw.nl/vocglossarium">http://resources.huygens.knaw.nl/vocglossarium</a>  Zie vraag 110

## E. Vragen over de user journey website

	Vraag	Antwoord
69.	Kunnen onderzoekers die de archiefstructuur begrijpen de zoektocht die als voorbeeld wordt gegeven wel volgen (vinden zij dit een normale workflow)?	Er is een verschil tussen mensen die de archiefstructuur begrijpen en mensen die daar niet mee bekend zijn. Toch blijft het lastig zoeken, ook als je de structuur, indexen en metadata kent. Elk archief is weer anders gestructureerd en ook zijn zaken over de eeuwen heen veranderd.
70.	Voor een user is het wellicht handig om direct naar de bron te gaan, maar andere keren wil je juist graag de structuur inzichtelijk hebben. Is dit conflicterend?	Het is inderdaad vaak belangrijk om niet alleen een antwoord te krijgen op een directe zoekvraag, maar om daarbij de context (boomstructuur) van het archiefstuk ook inzichtelijk te hebben. Het probleem is echter <b>dat er veel "clicks" nodig zijn om bij het</b> gezochte stuk te komen. Onafhankelijk van wat je zoekt, zou dit verminderd (verbeterd) moeten worden. Het Nationaal Archief is hiermee bezig en hoopt dat dit eind van het jaar structureel beter zal zijn.
71.	Zijn er alleen transcripties of is de context (boomstructuur) van de transcripties ook bekend?	In de bestandsnaam is opgenomen uit welk archief een stuk afkomstig is en wat het inventarisnummer is. De beschrijving van het archief (de toegang) kan context geven aan de transcripties.
72.	Waar ligt binnen de SBIR-oproep de nadruk: bij het gericht zoeken of bij de user experience?	Het gaat om de combinatie. Bij de AI-toepassing zal de nadruk wellicht op het zoeken liggen maar er moet natuurlijk wel toegevoegde waarde zijn voor de gebruiker. Zie vraag 96
73.	Hoe zit het met revisies?	Nu kunnen klanten nog een revisie doen. Dat kan niet indien transcribus wordt toegepast. Dat systeem verbetert zichzelf.
74.	Kunnen bronnen wijzigen door revisie en transcripties?	Ja, dat kan gebeuren, maar dat gebeurt niet continu.
75.	Wat is jullie belangrijkste klant?	Alle klanten zijn even belangrijk. Een groot deel van onze klanten zijn gepensioneerd mannen. Maar we hebben ook veel onderzoekers en studenten die onderzoek doen in ons archief. Zie <b>ook de bijlage Persona's bij de SBIR oproep</b> . Zie ook vraag 45 en 48
76.	Moet het prototype voor alle collecties toepasbaar zijn?	Nee, het prototype moet nu toepasbaar zijn op de archieven van het VOC en het Notarieel archief. Uiteraard is wel het verzoek om de relatie tot de metadata en de brede zoekfunctionaliteiten in het oog te houden.
77.	Ik mis de digital humanities onderzoeker en de hergebruiker als personae! Suze is nog een erg klassieke onderzoeker. De digital humanities onderzoeker en zeker de hergebruiker wil de data / transcripties er	Die persona is zeker ook belangrijk, maar zoals ik het lees wil deze persoon alleen de ruwe data ontvangen om vervolgens zelf er intelligentie op los te laten. Dat is ook niet zozeer de uitdaging, dat is het aanbieden van open data. De SBIR

	Vraag	Antwoord
	graag als ruwe data uit kunnen halen/ grootschalig, los van de images (wel met uniek verwijzing naar) Om ze te kunnen vergelijken/ analyses op los te laten, te verbinden aan andere sets etc. Hebben jullie die personae bewust weggelaten?	oproep gaat om de wijze waarop we onze bezoekers van de website kunnen helpen met het zoeken in de data.
78.	Als er paginanummer staan op de scan, worden die dan ook getranscribeerd?	Niet altijd zijn deze paginanummers herkent als tekstregio of getranscribeerd.
79.	Heb je een beeld hoe je zou willen dat het systeem functioneert?	Het zou mooi zijn als we de beschrijving naast de scan kunnen zien. <b>Tevens leuk om daadwerkelijk te 'bladeren' door de scans, deze mogelijkheid zie je soms al bij digitale kranten.</b>
80.	De projectwebsite valt binnen de scope van het project, maar niet binnen het prototype?	Jawel, de website hoort bij het prototype want dit is belangrijk voor het doorzoeken en tonen van het archief.

## F. Vragen over de solution architectuur

	Vraag	Antwoord
82.	Zijn de in de Solution Architecture genoemde programmeertalen limitatief?	Nee, er worden sterke voorkeuren genoemd. Afwijkingen zijn mogelijk, want het gaat hier om innovatie. Uiteindelijk moet het wel inpasbaar zijn.
83.	Welke architectuur gebruikt NA nu?	Als de technische infrastructuur wordt bedoeld dan geldt: Er zijn nu 2 systemen namelijk Linux en Windows, men wil naar 1 containerplatform. De servers worden uitgebreid en zullen worden beheerd bij een extern overheidsdatacenter. Voor snelle implementatie heeft het gebruik van de technische infrastructuur de voorkeur, maar i.h.k.v. deze SBIR zijn er geen beperkingen en is het prototype free format.
84.	In de SBIR wordt gevraagd om een oplossing met behulp van AI, waarom specifiek AI?	Bij NA is nog weinig nagedacht over het doorzoeken van transcripties, AI lijkt voor het Nationaal Archief een goede oplossing voor de zoekfunctie in de data van het archief.
85.	Is er veel AI- kennis bij het NA?	Nee, nu niet, dit kan nog veranderen.
86.	Dienen networks de benodigde documenten te beoordelen?	Ja die liggen hieronder. Er is niet één algoritme wat breed gedragen wordt. Je hebt een netwerk nodig.
87.	Hoeveel ontwikkelaars heeft NA in dienst?	Binnen NA wordt nu agile gewerkt in 3 teams voor het ontwikkelen van de website, het e-Depot en de servicelaag daartussen
88.	Komt het onderhoud van de IT-oplossing bij het NA te liggen?	Ja, uiteindelijk wel.
89.	Wil het NA een oplossing die compleet in huis draait of mag het een service zijn op een andere locatie?	Het NA wil een oplossing die uiteindelijk in huis draait.
90.	Zoekt NA een kant-en-klare oplossing of mag er samenwerking worden gezocht met de interne IT dienst?	NA zoekt de samenwerking voor integratie van bestaande en nieuwe IT architectuur.
91.	Als het NA het uiteindelijke product wil integreren in huis, is er dan voldoende kennis aanwezig voor het beheer en doorontwikkeling van componenten met A.I.. Als dat niet het geval is, wordt er dan verwacht dat er een SLA wordt afgesloten hiervoor?	NA heeft zelf geen ontwikkelaars in dienst – NA heeft een contract voor het technische applicatiebeheer. Na oplevering is het de intentie dat de ontwikkelde software door deze partij in beheer wordt genomen.
92.	In de solution architecture NFE's wordt gevraagd om een oplossing die gebouwd is met Java en PHP. Wij vragen ons af in hoeverre Java een harde eis is aangezien wij meer waarde zien in talen als Python. Zeker gezien jullie visie op herbruikbaarheid, zien wij voor dit project	het gebruik van JAVA en/of PHP door ons wordt niet gezien als een harde eis voor het SBIR-traject. Van onze kant is dit bedoeld als voorkeur - de reden is dat er bij het Nationaal Archief een aantal applicaties aanwezig is die gebruik maken van Java en/of PHP.

	Vraag	Antwoord
	componenten die out-of-the-box in Python verder ontwikkeld zijn en beter onderhouden worden.	Dientengevolge is er daarom kennis van deze technieken aanwezig.  Indien een andere techniek wordt gebruikt – i.e. Python – heeft dit tot gevolg dat in de toekomst daarvoor eventueel kennis aangetrokken dient te worden.
93.	In de presentatie op de eerste slide staan een aantal blokken in blauw. Vallen al deze blauwe blokken binnen de scope?	Nee, Alleen het Transcriptiebeheersysteem, de Annotation server, Elastic Search en de Projectwebsite transcripties
94.	Ik zie in datzelfde plaatje (vraag 93) geen pijl tussen het beheer en de eindgebruiker. Klopt dat wel?	Klopt. Die is er niet. In dit plaatje wordt met de gebruiker de klant in de studiezaal en op internet bedoeld. Deze klant heeft geen rol in het beheer.
95.	Hoeveel vrijheid hebben we om een innovatieve oplossing te bedenken?	We willen zeker een innovatieve oplossing. De hamvraag is hoe we het beste uit onze archieven (verschillende soorten) kunnen halen. Daarnaast zijn er randvoorwaarden. De architectuur is opgezet met als voornaamste kader de inpasbaarheid in onze huidige en voorziene infrastructuur. Als een leverancier andere c.q. betere ideeën heeft dan is hij van harte uitgenodigd om die in een prototype uit te werken.
96.	Wat is belangrijker vernieuwende AI of voldoen aan de user journey?	Zie vraag 72.
97.	Hebben we toegang tot domeinkennis?	Ja. Er zijn documenten beschikbaar en partijen die een opdracht krijgen voor fase 2 zullen geregeld contact hebben met het Nationaal Archief. Zie ook vraag 122
98.	De voorbeelddataset is in Page XML-formaat. Is dit een standaard waar we ons aan moeten houden?	Bij voorkeur gebruiken we Page XML. Maar de Page export uit Transkribus bevat geen coördinaten op woord niveau. Mogelijk moet Alto XML worden gebruikt om de zoekresultaten te kunnen highlighten in de scan. Tijdens de ontwikkeling van het prototype moet hierover een besluit worden genomen. <a href="#">Zie vraag 36, 67</a>
99.	Is er een voorkeur voor aanleveren in bijvoorbeeld API microservices?	Nee er is geen voorkeur. Tot nu werken wij met <b>API's. Als microservices een goed alternatief</b> hiervoor kunnen bieden, prima
100.	Is het voor het prototype belangrijk dat dit wordt opgeleverd in Linux Redhat packages?	Nee, voor het prototype is dit niet van belang. Voor het uiteindelijk op te leveren product wel.
101.	Is er een lijstje met KPI's voor het prototype?	Nee. Het gaat er om dat u voldoet aan de in de SBIR-oproep beschreven criteria.
102.	Hoe vindt de beoordeling plaats?	De beoordeling van de ingediende offertes vindt plaats op basis van in de SBIR-oproep en de <a href="#">handleiding</a> beschreven criteria.
103.	Kunt u iets meer zeggen over de huidige infrastructuur bij het Nationaal Archief?	Het grootste deel van het systeem draait op Linux-servers. Alles wat we doen, ook storage gebeurt in eigen beheer. We hebben verder een voorkeur voor open source software.
104.	Is er al een repository voor alle digitale assets en worden deze daarin opgeslagen?	De scans staan in het e-Depot. Er worden zichtexemplaren gemaakt die op de IIF-



	Vraag	Antwoord
		omgeving. Voor de transcripties is nog geen repository. Daar moet een beheersysteem worden gerealiseerd, dat is onderdeel van de opdracht.
105.	Moet de annotation server voor het beschikbaar stellen van transcripties ook worden ontwikkeld in deze SBIR?	Ja, de Annotation Server is een beoogd doel voor het eindresultaat, voor het prototype is die <u>niet</u> verplicht. Het is belangrijk om aan te tonen dat het web annotation format gebruikt kan worden.
106.	Is Elastic search een onvoorwaardelijke toepassing bij een nieuwe ontwikkeling?	Niet per se. Elastic search is nu de standaard en sluit aan bij de website, en het heeft onze voorkeur om deze zoekmachine te gebruiken. Vooral met het oog op kennis en integratie van zoekresultaten uit andere bronnen. Als daar goed beredeneerde alternatieven voor zijn, horen we het graag.
107.	De informatie in het projectplan De ijsberg zichtbaar maken sluit niet helemaal aan bij de informatie in de SBIR oproep. Hoe zit dat?	Het projectplan is een jaar geleden geschreven. Niet alle informatie in het plan is actueel. Voor deze SBIR zijn de oproep, de architectuur en de Toelichting op de SBIR oproep leidend boven het projectplan.
108.	Welke rol speelt modereren in de beheermodule (F6 uit MSCW lijst); welke functionaliteiten zouden jullie hiervoor graag zien?	Voor moderatie is in dit project in deze fase nog geen rol. Er is in de toekomst de mogelijkheid voorzien. Dat transcripties worden verbeterd – bijvoorbeeld door gebruikers of medewerkers. Na die wijzigingen moeten ze gemodereerd kunnen – dus door anderen verbeterd/geautoriseerd. Met deze functionaliteit willen we wel rekening houden, maar dat behoort niet tot de <b>MUST HAVEs of de SHOULD HAVE's in deze fase.</b>
109.	Wordt informatie over de coördinaten van woorden beschikbaar gesteld door de OCR oplossing, zodat woorden uit transcripties te linken zijn aan locaties in de scan?	Zie vraag 60
110.	Zijn er in het kader van onderdeel B11 uit de MSCW lijst reeds woordenlijsten en woordenboeken Oud Nederlands beschikbaar? als bestanden of worden deze via een CDN of iets dergelijks beschikbaar gesteld	Zie vraag 68
111.	Zijn er in de dataset t.b.v. de PoC reeds transcripties in meerdere versies aanwezig? Zitten zaken als de hiërarchie en archief categorisering dan in een dataset? Vergelijkbaar: worden de scans aangeleverd	Naar aanleiding van de vragen hebben we de dataset opnieuw getraind met het jongste trainingsmodel. De export van deze dataset komt beschikbaar in Alto XML word, Page en TXT Hiermee heeft u dus meerdere versies. In de eerste dataset zit echter geen ALTO XML. Hiërarchie is te halen uit de toegang (=beschrijving) van het archief. Deze is ook beschikbaar in het datapakket, zowel in PDF als XML.

	Vraag	Antwoord
112.	In hoeverre is het 'vertalen' en verbeteren van de characters naar correct Nederlands onderdeel van de scope? In de MSCW lijst geeft men wel aan dat een correctheidsscore gewenst is (C3, C4), dus lijkt het in de scope te vallen. Klopt dit? Wat is dan leidend bij weergave- de originele transcriptie of de verbeterde?	De originele transcripties is leidend. Vertalen en <b>verbeteren naar hedendaags "correct"</b> Nederlands is niet de bedoeling.  Met de weergave van zoekresultaten zou rekening gehouden kunnen worden met de correctheidsscore om de probabilliteit te berekenen dat een zoekresultaat dat gene is wat wordt gezocht.
113.	In de solution architectuur lijkt alles een 'stand alone' oplossing te zijn. Wordt er dan ook in de PoC geen enkele koppeling met een API	In de uiteindelijke Solution architectuur zijn <b>minimaal 4 API's voorzien. Die zouden ook gerealiseerd en gebruikt moeten worden binnen het uiteindelijke project.</b>
114.	Mag het prototype ook met een ander sturingssysteem worden gemaakt?	Als de systemen maar met elkaar kunnen praten.
115.	GAF toegestaan?	GAF maakt geen onderdeel uit van het prototype.
116.	Moet er ook een beheerssysteem worden gemaakt? en Hoe denken jullie dat deze extra benodigdheden passen binnen het beschikbare budget?	Ja, want er is een systeem nodig om de data te beheren. Het importeren en exporteren van data is van belang. Wij denken dat dit past, je hoeft geen heel systeem te maken. Het prototype moet kunnen laten zien wat er mogelijk is. We zijn dus op zoek naar een innovatieve manier om te zoeken en resultaten te laten zien. Als het product niet een volwaardig beheerssysteem heeft, is dat minder belangrijk.
117.	Mag er naast te schrijven software ook standaard software worden ingezet waarvoor licentiekosten moeten betaald? Deze software wordt alleen op de achtergrond aangeroepen maar is wel noodzakelijk om het idee van de oplossing te realiseren.	Ja, dat mag. Geef wel duidelijk aan wat de licentiekosten zijn.
118.	Is er een internationale vergelijking gemaakt van leveranciers van een functionaliteit voor zoeken en tonen in transcripties?	Er zijn ontwikkelingen gaande. Maar we hebben hierin nog niet de oplossing gezien die voldoet aan de eisen en wensen van het Nationaal Archief.
119.	Kunnen wij gebruik maken van jullie B-to-C portal of dienen wij dit aan te bieden?.	Degene die een prototype bouwt doet dit met en op eigen middelen. Het NA stelt hiervoor geen faciliteiten beschikbaar
120.	Binnen de documenten wordt er gesproken over gebruik van "open source en Linux (Redhat)". Onze oplossing is een SaaS oplossingen en geen open source. Zijn jullie specifiek opzoek naar een open source oplossing of is een SaaS oplossing ook mogelijk?	Dit Verwijst naar SAAS. Dat is geen optie die het NA wil – met name indien dit gaat over beperkt openbaar materiaal (CABR bijvoorbeeld) mag dat niet naar buiten.
121.	Onze oplossing maakt gebruik van Google cloud. Past dit binnen jullie verwachting?	Hiervoor geldt hetzelfde als bij de vraag hiervoor. NA wil alles in eigen beheer/op eigen servers houden

	Vraag	Antwoord
<b>122.</b>	Met name voor het inrichten van de user interface is het voor de ontwikkelaar van belang om een goed beeld te hebben van de wensen en eisen van de gebruiker, de praktijk leert dat dit zelden volledig te ondervangen is in een plan van eisen. Is er gedurende de pilotfase (na toekenning van de SBIR) de mogelijkheid om contact te hebben met het NA om de ontwikkeling van de pilotsoftware met hen af te stemmen?	Ja, indien u een opdracht krijgt zijn er regelmatig overleggen tussen u en het NA. Uw idee hierbij neemt u op in uw projectplan. Zie ook vraag 97
<b>123.</b>	Indien het antwoord op bovenstaande vraag ja is, wordt de uitgewisselde informatie tussen het NA en één van de aanvragers ook gedeeld met de overige twee aanvragers?	Ja, van wege de eisen van openheid, eerlijkheid en transparantie. Uiteraard wordt bedrijfsgevoelige informatie niet gedeeld.

	Vraag	Antwoord
--	-------	----------

Ten slotte

	Vraag	Antwoord
124.	Volgt er nog een 2 <sup>e</sup> Nota van Inlichtingen?	Ja.