

KVAN

TOT OP DE BODEM!

Geautomatiseerd toegankelijk maken van archieven
een samenwerking tot versneld digitaliseren binnen de archiefsector

Versiegeschiedenis

Versie datum	Wijzigingen	Markering/wijzigingen
4 juli 2025	Eerste opzet (v0.9) - work in progress-, IO	
19 aug. 2025	Versie 1.0 met nog in te vullen details, IO	
20-08-2025	Review en omzetten naar huisstijl KVAN	

Inhoud

1. Missie, visie en doelstelling.....	5
2. Plan van aanpak	6
Sturing, team en proces	6
Transparantie	6
Kick-off opstarten	7
3. Deelnemende partijen	7
4. Minimum viable product.....	8
Technieken:	8
Outputvereisten	8
Afbakening	8
Toekomstperspectief (fase 2 en 3, buiten MVP):	8
5. Scope	9
Aanlevering archiefmateriaal.....	9
Selectiecriteria voor het archiefmateriaal	9
Toelichting.....	9
Verantwoordelijkheid voor online beschikbaarstelling	10
6. Realisatie	10
Governance	10
Deel 1 Selectie en digitalisering	11
1. Onderdelen en werkzaamheden.....	11
2. Plan van aanpak	12
3. Planning.....	12
4. Eindproducten.....	12
Deel 2 Transcriptie en verrijking	13
1. Onderdelen en werkzaamheden.....	13
2. Plan van aanpak	13
3. Planning.....	14
4. Eindproducten.....	15

DEEL 3 Digitale beschikbaarstelling	16
1. Onderdelen en werkzaamheden.....	16
2. Plan van aanpak	16
3. Planning.....	17
4. Eindproducten.....	17
DEEL 4 Evaluatie en vooruitblikken.....	18
1. Onderdelen en werkzaamheden.....	18
2. Plan van aanpak	18
3. Planning.....	18
4. Eindproducten.....	19
DEEL 5 Sturing, team en proces	20
1. Onderdelen en werkzaamheden.....	20
2. Plan van aanpak	20
3. Planning.....	21
BIJLAGEN	22
Risicomanagement.....	24

1. Missie, visie en doelstelling

De Koninklijke Vereniging Archiefsector Nederland (KVAN) staat voor duurzaam toegankelijke informatie, beschikbare informatie en openbaarheid. De archiefsector heeft als belangrijkste taak: het bieden van de mogelijkheid tot reconstructie, het basisbeginsel van een transparante overheid binnen de democratische rechtsstaat. KVAN ziet dat de archieffunctie en informatiebeheer andere betekenissen krijgen en zet zich in voor een goede positionering van de waarden, rollen en verantwoordelijkheden van de archieffunctie.

Onze missie is dan ook om de belangen te behartigen van de archiefsector en ondersteuning te bieden bij het anticiperen op en profiteren van veranderingen in de samenleving.

KVAN streeft naar duurzame toegang tot en het recht op informatie voor iedereen. En werkt in samenwerking met de archiefsector aan een inclusieve, digitale erfgoedtoekomst voor een breder én toekomstig publiek.

KVAN lanceerde in december 2023 de whitepaper 'Het archief als nutsvoorziening'. In deze whitepaper stellen we vast dat de rol van archiefinstellingen in het digitale domein verandert van een centrale (opslag en beheer) naar een verdeelstation (het verdelen en distribueren van informatie). In augustus 2024 is het programma 'nutsvoorziening' van start gegaan met onder andere een roadshow. Bijna organisch vloeide daar de samenwerkingswens uit voort dat nu in de vorm van dit projectplan voor ons ligt. Een stappenplan van digitale content, naar verbonden erfgoed naar integraal doorzoekbaar archiefmateriaal. In deze eerste pilotfase zetten archiefinstellingen zich samen in om archiefmateriaal geautomatiseerd toegankelijk te maken. Zo versnellen we het digitaal toegankelijk maken voor een breder publiek.

De toekomst van archiefbeheer is het archief als nutsvoorziening. Dat houdt in dat **archieven publiek toegankelijk, betrouwbaar en duurzaam beschikbaar moeten zijn voor iedereen**, net als water, elektriciteit of het internet. De weg hier naartoe ligt in het benutten van nieuwe technologische ontwikkelingen, automatisering, context, verbonden erfgoed, samenwerking en openbaarheid.

Waarbij te allen tijde de focus ligt op de eindgebruiker

De pilot voorziet in de eerste stap (digitale content) op weg naar het archief als nutsvoorziening. Fase 2 zal focussen op toegankelijkheid (LOD) en fase drie op verbinding en netwerk, waarbij de basis 'digitale content' is (fase 1).

De archiefinstellingen die bijdragen in dit pilotproject realiseren met elkaar een digitale werkwijze die handmatige, tijdsintensieve processen kan vervangen door de potentie van technologische innovatie optimaal te benutten.

In gezamenlijkheid ontdekken we wat al wél kan en wat nog meer aandacht nodig heeft en zo realiseren we een nieuwe methodiek en uniforme, eigentijdse manier van werken voor de archiefsector. Deze methodiek kan een leidende rol vervullen in de transitie naar een verbonden, digitaal erfgoedecosysteem waarin erfgoedinstellingen samen optrekken en elkaars kennis en technologie benutten.

2. Plan van aanpak

Centraal bij deze pilot en opzet staat samenwerken in een experimentele, leerzame vorm.

Bij de samenwerkingsaanpak delen we pilotfase 1 op in drie expertisedelen en een vierde, vijfde deel evalueert het proces en geeft toelichting op sturing en borging van de pilot:

- Selectie en digitalisering
- Transcriptie en verrijking
- Digitaal beschikbaar stellen
- Evaluatie en vooruitblikken

Sturing, team en proces

De pilot is experimenteel en lerend van aard. Om iedereen uit de archiefsector - én breder - te kunnen laten participeren én leren, is er gekozen voor drie vormen van participatie. Een 'actieve deelname'; de instelling draagt bij in expertise, uren en inzet. Een 'aanleverende participatie'; de instelling draagt financieel bij en draagt archiefmateriaal aan. En tot slot een 'supporters/ambassadeurs participatie'; deze instellingen dragen bij aan kennisdeling of faciliteren in bijvoorbeeld beschikbaarstellen van een locatie.

Participanten (vorm 2 en 3) krijgen toegang tot de eerste bevindingen en inzichten. Participanten vorm 1 onderzoeken, werken en realiseren met elkaar de uitkomst van de pilot.

Transparantie

Resultaten en proceservaringen worden gedeeld via:

- Kwartaalbijeenkomsten
- Tussentijdse kennisdelingssessies en webinars
- Evaluatierondes met inspraakmogelijkheid

Participanten zijn mede-pioniers in deze vernieuwende aanpak en dragen actief bij aan de ontwikkeling van duurzame digitaliseringsmethoden voor erfgoedinstellingen.

In het eerste jaar van het project vatten we grofweg de vier stappen van het project samen in deliverables gebaseerd op een afgestemd Minimum Viable Product in de Tweedaagse opstart sessie. De Teamcoördinator stemt telkens de te nemen stappen af met het team op weg naar de op te leveren 'producten' per kwartaal. Iedere week is er een korte stand-up waarin duidelijk wordt wie waaraan werkt.

Het kan zijn dat sommige verrijkingen bij het online beschikbaar stellen nóg niet tot zijn recht komen. Deze punten nemen we mee als aandachtspunten pilot fase 2 en 3.

Bij adviezen over de architectuur of benodigde aanpassingen zal rekening worden gehouden met de DERA (Digitale Erfgoed Referentie Architectuur). Functionele eisen: intuïtieve interface, publieksvriendelijk, responsive voor desktop, tablet en mobiel.

Kick-off | opstarten

In een Tweedaagse bijeenkomst komen we tot elkaar met een akkoord op het proces, (kwaliteits)afspraken, teamvorming en roadmap.

De tweedaagse is gericht op alignment en een kick-off voor de pilot en wordt professioneel begeleid. Documentatie en processtructuren worden al zoveel mogelijk een week voor de Tweedaagse opstart gedeeld zodat feedback kan worden verzameld en meegenomen in de sessies:

- **Dag 1** is gericht op: doel, verwachtingen, proces, governance en werkwijze.
 - **Dag 2** gericht op: verdere teamvorming, roadmap/planning en taakbeschrijving.
- Met aan het eind van dag twee een gezamenlijke, feestelijke borrel.

Ook hierin passen we een agile werkvorm toe en wisselen we intensieve sessies af met energizers, denkoefeningen en creativiteit om zo met elkaar tot een optimaal resultaat te komen.

We maken drie projectteams van 4 tot 5 expertises. Ieder projectteam wijst een teamcoördinator aan die telkens de te nemen stappen afstemt met het team op weg naar de op te leveren 'producten' per kwartaal. Iedere week is er een korte stand-up met alle drie de projectteams, waarin duidelijk wordt wie waaraan werkt en zorgt voor onderlinge afstemming.

3. Deelnemende partijen

De deelname aan pilotfase 1 ziet er als volgt uit. Aansluiten kan nog tot eind augustus 2025 in participatievorm 2 of 3:

Vorm 1 'actief'	Vorm 1 'aanleverend'	Vorm 3 'support/ambassadeur'
Gelders Archief	Waterschap Amstel, Gooi, Vecht	HTR-Experts
Regionaal Archief Nijmegen	Tresoar	Waterlands Archief
Stadsarchief Amsterdam	West-Brabants Archief	Groninger Archieven
Noord-Hollands Archief	IISG	Regionaal Archief Tilburg
Westfries Archief	Zeeuws Archief	Amsterdam Time Machine
Het Utrechts Archief	Regionaal Archief Rivierenland	Stadsarchief Breda
	BHIC	

4. Minimum viable product

Een Minimum Viable Product is het meest noodzakelijke resultaat van je project. Een MVP helpt je focussen op de essentie van het gezamenlijk product en maakt het mogelijk sneller bij te sturen. Daarmee beperken we risico, besparen we kosten, leren we sneller en scheppen bovenal duidelijkheid.

Doel van de MVP in projectfase 1 van de pilot is:

aantonen dat (on)gestructureerd archiefmateriaal automatisch gestructureerd en doorzoekbaar gemaakt kan worden, en archieven zo versneld over digitale content kunnen beschikken.

Technieken:

- ATR (OCR - Optical Character Recognition, HTR - Handwritten Text Recognition)
- AI (machine learning, artificial intelligence)
- Optioneel: Named Entity Recognition (NER)
- Termennetwerken / semantische verrijking

Outputvereisten

* (machine leesbare aanlevering):

- Image formaten: PDF, JPEG
- Data formaten: XML
- **Toegang formaten:** 'klassiek' EAD, RiC-O
 - * gespecificeerd en vastgelegd in 'kwaliteitseisen'-document

Afbakening

- De MVP beperkt zich tot de digitale ontsluiting van analoog archiefmateriaal en de demonstratie van automatische structurering en doorzoekbaarheid.
- Het toevoegen van Linked Open Data-entiteiten via NER-technologie wordt onderzocht maar is géén eindproduct.
- De integratie met ander erfgoedmateriaal (zoals bedoeld in fase 2 en 3) valt buiten de scope van deze pilotfase en nemen we alleen mee indien tijd, geld en inzet dit toelaat.

Toekomstperspectief (fase 2 en 3, buiten MVP):

- Fase 2: verbinden van archiefmateriaal onderling en met ander erfgoed (verbonden erfgoed). Gericht op innovatie en toegankelijkheid
- Fase 3: het integraal doorzoekbaar maken van deze verbonden collecties. Gericht op verbinding en netwerk

5. Scope

Aanlevering archiefmateriaal

Iedere deelnemer (participatievorm 1 of 2) levert maximaal 5 strekkende meter archiefmateriaal scanklaar aan.

- ® Minder mag, méér kan eventueel in overleg maar wordt naar rato in rekening gebracht én mag niet ten koste gaan van het gezamenlijk op te leveren eindproduct (€ 1000,- per extra meter of € 500,- per halve extra meter)
- ® De begroting is gemaakt op een aantal scans dat overeenkomt met een totaal van circa 300.000 scans
- ® Het aangeleverde materiaal mag bestaan uit één archief of meerdere kleinere archieven die samen de maximumomvang benaderen
- ® Het materiaal mag zowel analoog als reeds gedigitaliseerd aangeleverd worden.

Selectiecriteria voor het archiefmateriaal

- ® Herkomst: Alle typen archiefmateriaal zijn toegestaan
- ® Er is een voorkeur voor thematisch materiaal (slavernijverleden, koloniale collecties of andere thematische insteek), maar dit is geen vereiste
- ® Datumgrens: Bij voorkeur materiaal van vóór 1925 (in verband met openbaarheid)
- ® Selectie: Het materiaal moet voor 3 september 2025 geselecteerd zijn en bij voorkeur scanklaar zijn.

Het materiaal moet zodanig voorbereid zijn dat het zonder extra handelingen gescand kan worden. Dat houdt in:

- ☒ Geen nietjes, paperclips, plakband of andere bindmiddelen
- ☒ Gladde, onvouwde pagina's (geen gevouwen hoeken of gekrulde randen)
- ☒ Schoon (vrij van stof, vuil of schimmel)
- ☐ Hersteld waar nodig (gescheurde pagina's geplakt of gemarkeerd voor aparte behandeling)
- ☐ In logische volgorde (paginavolgorde of dossiervolgorde)
- ☐ [Aanbevolen] Uniform formaat, of duidelijk gescheiden formaten (niet groter dan A1)
- ☒ Goed leesbaar (geen ernstige verkleuring of vervaging)
- ☒ Verpakt in geschikte dozen of trays, voorzien van labels (naam/nummer archief én instelling)
- ☐ Inventaris (indien beschikbaar) of registratie aanwezig, voor tracking en controle

Toelichting

Logische volgorde en uniforme formaten zijn aanbevolen, maar geen vereiste.

Afwijkingen kunnen echter leiden tot extra verwerkingstijd en verminderde kwaliteit. Omdat alle deelnemende archieven gelijk worden behandeld binnen dezelfde werkwijze en tijdsverdeling, kan een grote variatie in complexiteit of structuur van het bronmateriaal leiden tot een minder consistente kwaliteit van de terug geleverde data.

Bij het aanleveren van het materiaal voor digitalisering is het van groot belang om de juiste informatie mee te geven.

- Geef te allen tijde een ISIL-code of andere unieke ID van de instelling mee;
- Indien een inventarislijst, inventaris nummers of plaatsingslijst aanwezig, lever deze dan aan;
- Indien de dozen geen inventarisnummer hebben, maar wel een barcode, label of andere UI. Geef dit dan s.v.p. mee.
- Dit ten behoeve van overzicht op te halen materiaal en naamgeving terug te leveren materiaal.

Verantwoordelijkheid voor online beschikbaarstelling

Aanleverende participanten (vorm 2) zijn verantwoordelijk voor:

- « Tijdige aanlevering van scanklaar materiaal
- « Online publicatie van het digitaal terug geleverde archiefmateriaal vóór 4 september 2026
(houd s.v.p. rekening met eventuele korte doorlooptijd van aanlevering naar publicatie)

Indien ondersteuning nodig is bij ingest of publicatie, dient dit tijdig gemeld te worden, zodat begeleiding en/of ondersteuning georganiseerd kan worden.

6. Realisatie

Governance

Het project wordt aangestuurd via een stuurgroep die strategische kaders en besluitvorming bewaakt. De dagelijkse coördinatie ligt bij de programmamanager, die tevens de rol van scrummaster bekleedt voor procesbewaking en samenwerking.

Er zijn drie zelfsturende teams, elk verantwoordelijk voor een eigen werkpakket. De teams werken autonoom binnen hun opdracht en zijn gelijkwaardig in zeggenschap. Afstemming en voortgang vinden plaats via wekelijkse stand-ups en maandsessies, waarbij transparantie en gezamenlijke verantwoordelijkheid centraal staan.

Deel 1 Selectie en digitalisering

1. Onderdelen en werkzaamheden

- Selectie en scanklaar maken (Verantwoordelijk: bronhouder)
- Voorbereiding voor transport (Verantwoordelijk: bronhouder, Coördinatie: Projectgroep 1)

Werkzaamheden Projectgroep 1:

- Het controleren scanklaar hebben en tijdig klaarzetten van het materiaal voor het ophalen van alle archiefstukken voor het vervoer naar het scanbedrijf vanuit de verschillende aanleverende archiefinstellingen.
- Het voeren van transportadministratie / logboek.
- Scannen (Extern bedrijf, Coördinatie Projectgroep 1)
- Contact met scanbedrijf over het scannen van de archiefstukken conform de gestelde kwaliteitsnormen binnen de vastgestelde termijn. Binnen de pilot hanteren we de Metamorfoze light richtlijnen, zie ook 'Volmacht & kwaliteitseisen-document'.
- Kwaliteitscontrole (Extern bedrijf, Validatie: Projectgroep 1)
- Het controleren van de door het scanbedrijf aangeleverde batches (technisch en inhoudelijk).
- Het controleren van de door het scanbedrijf aangeleverde batches (technisch en inhoudelijk).
- Eventueel openbaarheid opsporen (Indien noodzakelijk coördinatie: Projectgroep 1 i.s.m. bronhouder)
- Opslag van digitale bestanden (Coördinatie: Projectgroep 1, i.s.m. projectgroep 2)
- Na goedkeuring aangeleverde kwaliteit, het opslaan van alle bestanden in de opslagomgeving voor samenwerking (bv. Transkribus)
- Terugplaatsing van originele stukken (Coördinatie: Projectgroep 1), inclusief het voeren van een productieadministratie rondom genoemde werkzaamheden. En het terug leveren van de originele digitale bestanden (zonder verdere bewerking) aan bronhouder.
-

Projectgroep 1	Teamleden	Materieel
Te selecteren Teamcoördinator	- Coördinator logistiek - Controleur scans - Adviseur digitalisering (archivistiek) - Operationeel manager productcontrole	- Scanning (uitbesteed) - Transporten (uitbesteed) - Opslag

2. Plan van aanpak

De digitalisering zal door externe partij(en) worden gedaan, maar de voorbereiding moet zijn gedaan door de aan te leveren bronhouder.

Er wordt uitgegaan van een gemiddelde verwerkingscapaciteit van 100.000 tot 152.000 pagina's per week (2 tot 3 productieweken). Daarnaast moet transport en planningen gemaakt worden, waarbij we uitgaan van een gefaseerde oplevering van gedigitaliseerd materiaal. Meteen in september starten met digitaliseren zodat in oktober gestart worden met de data-curatie van het eerste aangeleverde materiaal.

In dit project draait het bovenal om de toegang tot het archief door middel van digitalisering te verbeteren. Er is dus geen sprake van 'preservation imaging' en er zullen daarom alleen JPEG-bestanden worden opgeleverd. Voor de beeldkwaliteit gelden de kwaliteitsnormen vastgelegd in de Metamorfoze Light-richtlijnen.

Na digitalisering zullen de aangeleverde batches inhoudelijk en technisch door Projectgroep 1 worden gecontroleerd. Bij goedkeuring worden de bestanden overgezet naar een bewerkingsomgeving (mogelijk Transkribus EON). Het originele materiaal met de scans zullen worden teruggegeven aan de bronhouder (voor de duurzame opslag).

3. Planning

	Begindatum	Einddatum
Voorbereiding digitalisering	week van 8-9-25	19-9-25
Transport	18-9-25	23-9-25
Digitalisering	22-9-25	14-10-25
Kwaliteitscontrole	7-10-23	31-10-25
Terugleveren originelen en scans bronhouder	31-10-25	14-11-25
Ingest opslag voor verdere verwerking	31-10-25	5-11-25

Nog te verifiëren data tijdens Tweedaagse op 3 en 4 september.

4. Eindproducten

- Alle documenten zijn conform planning getransporteerd en weer teruggeplaatst.
- Bij benadering 300.000 JFIF jpeg-bestanden met een compressie 1:10 (zie 'volmacht & kwaliteitseisen-document').

Deel 2 Transcriptie en verrijking

1. Onderdelen en werkzaamheden

- Documentseparatie, lay-out herkenning, document classificatie
- Automatische transcriptie
- Samenvatting genereren, inventarisnummers genereren en datering extraheren
- Dataverrijkingen extraheren (persoonsnamen, organisaties, plaatsnamen, gebeurtenissen, NRO)
- Koppelen aan Termennetwerk.

De werkzaamheden omvatten het genereren van machine leesbare tekst op basis van de gedigitaliseerde archiefstukken, evenals het ontwikkelen van een automatische transcriptiepipeline. Optioneel wordt (verrijkte) data geproduceerd die als basis kan dienen voor het inrichten van filters (facets) in de zoekfunctionaliteit, zodat de bronhouder de eigen raadpleegomgeving hierop kan afstemmen. Dit vereist tevens de realisatie van een verrijkingspipeline. Onder dataverrijking verstaan we het extraheren van relevante gegevens, zonder deze direct als Linked Open Data (LOD) aan te bieden. Waar mogelijk wordt een koppeling gelegd met het Termennetwerk. Daarnaast wordt een beschrijving gegenereerd die de inventarisnummers structureel koppelt aan de bijbehorende digitale representaties [werkwijze nog te verifiëren].

Projectteam 2	Personeel	Materieel
Te selecteren Teamcoördinator	- Software engineer - Datacurator - Medewerker thesaurus - ICT informatiespecialist - Python engineer	- Transkribus - Licentie en toegangbeheer - Opslag

2. Plan van aanpak

De onderdelen **Automatische transcriptie** en **Verrijking** benaderen we hieronder los van elkaar. Indien relevant zullen specificaties worden meegenomen.

Automatische transcriptie

Een kwalitatief hoogwaardige machine leesbare tekst vormt de basis voor alle vormen van dataverrijking, met in het achterhoofd de eindgebruiker en het doel van een toekomstgerichte full-tekst zoekfunctionaliteit. Data-curatoren zullen worden ingezet om door middel van machine-learning de kwaliteit van de machine leesbare tekst te optimaliseren.

Voor de pilot gebruiken we de nieuwe EON end-to-end van Transkribus die in augustus 2025 beschikbaar zal zijn.

Verrijking

In pilotfase 1 zet de Teamcoördinator de te zetten kwartaal stappen uit. Indien relevant zullen specificaties worden meegenomen. Het doel is dat archiefmateriaal dusdanig verrijkt (met het oog op fase 2) wordt dat in ieder geval de volgende toegangen bij digitaal beschikbaarstelling kunnen worden getoond:

- **Archief- en/of dossier- en/of inventarisnummer:** verwijzing naar het fysieke dossier waarin het gedigitaliseerde document zich bevindt.
- **Persoonsnamen:** door middel van 'slimme' matchingscripts worden namen uit het archiefmateriaal gekoppeld aan namen aanwezig in het Termennetwerk en/of Wikidata.
- **Organisatienamen:** 'simple name matching' met de organisaties uit thesauri.
- **Plaatsnamen:** 'simple name matching' met alle plaatsnamen uit Wikidata en/of Basisadministratie Gebouwen.
- **Data:** met een datumherkenner worden datums in de tekst geïdentificeerd.
- **Trefwoorden:** matching met een selectie van concepten uit thesauri.
- **Gebeurtenissen:** matching met een selectie van gebeurtenissen uit thesauri.
- **Type document:** Identificeren van een select aantal documenttypen op basis van machine leesbare tekst en image-kenmerken.

Binnen dit projectteam werken software-engineers en datacuratoren (gericht op ground-truth en machine learning) samen om de beoogde verrijkingen te realiseren. Deze verrijking bestaat uit twee stappen: identificatie en contextualisering. De precieze aanpak zal zich gaandeweg, in samenwerking, verder uitkristalliseren.

Ter kennisgeving: in het project *Oorlog voor de rechter* zijn de verrijkingen via een 'export as annotation'-API beschikbaar gesteld aan de eigen webomgeving. Dit illustreert dat het terugleveren van verrijkte data mogelijk is, maar nog geen standaard of vanzelfsprekende route kent.

3. Planning

Automatische transcriptie

Ter illustratie te nemen stappen:

- inrichten van infrastructuur bij en onder Transkribus licentie EON
- inkloppen van ground-truth om automatische transcriptie te verbeteren
- tests en metingen met Transkribus
- aanscherpen en opleveren van autotranscriptiepipeline
- doorvoer van productiescans in auto-transcriptiepipeline
- inkloppen van ground-truth om automatische transcriptie en machinelearning te verbeteren.

	Begindatum	Einddatum
Vorbereiding	8-9-25	11-11-25
Eerste experimenten met productiedata	1-12-25	12-6-25
Oplevering methodiek pipeline autotranscriptie		11-5-25
Verwerken productiedata		

Nog te verifiëren / in te vullen data tijdens Tweedaagse op 3 en 4 september.

Verrijking

- inrichten van verrijkingspipeline waarbinnen machine leesbare tekst verrijkt kan worden.
- inrichten van omgeving om ground-truth aan te maken.
- inkloppen van ground-truth
- machinelearning door datacuratie

	Begindatum	Einddatum
Vorbereiding	1-11-25	12-6-25
Oplevering methodiek pipeline dataverrijking (en API voor productie)		11-5-25
Verwerken productiedata		

Nog te verifiëren / in te vullen data tijdens Tweedaagse op 3 en 4 september.

4. Eindproducten

- Machineleesbare tekstdocument in generieke XML voor elke pagina voor alle ATR, OCR- en HTR-varianten (zie 'volmacht & kwaliteitseisen'-document).
- Methodiek voor auto-transcriptiepipeline voor de verwerking van scans tot en met de opslag van de machine leesbare bestanden toe te passen in de 'eigen infrastructuur'
- Optioneel: (verrijkte) data die de basis leveren voor filters ('facets') in de zoekfunctie.
- Methodiek voor een verrijkingspipeline (optioneel)

DEEL 3 Digitale beschikbaarstelling

1. Onderdelen en werkzaamheden

- Klassieke EAD of RiC-O structuren als output, overleggen/creëren met technische partij
 - >> schema's aanleveren en valideren output kwaliteit
 - >> in samenwerking en afstemming met Projectteam 1 voor genereren van de velden
- Terugplaatsen van digitale (verrijkte) bestanden aan archief
 - >> Het aanleveren van de (verrijkte) data aan de bronhouder, overleg en documentatie over compatibiliteit 'eigen infrastructuur'
- Ingest in eigen infrastructuur/collectiebeheersysteem
 - >> Begeleiden, waar nodig, ingest in eigen infrastructuur 'bronhouder' zodat binnen looptijd pilot het materiaal online beschikbaar komt (deadline: 15 september 2026)
- Digitaal beschikbaar stellen archief voor publiek
 - >> opstellen documentatie methodiek transcriptie- en verrijkingspipelines, ingest in collectiebeheerssysteem.

Projectteam 3	Personeel	Materieel
Te selecteren Teamcoördinator	- Meta data specialist (EAD, RiC-O)	- Inhuur controle methodiek - Inhuur ingest opgeleverde data in eigen infrastructuur

2. Plan van aanpak

In goed overleg met de bronhouder worden plannings gemaakt voor aanleveren data, ingest en online beschikbaarstelling van het materiaal. De bronhouder heeft tijdig aangegeven hierbij wel of geen hulp nodig te hebben.

Het projectteam houdt toezicht op voortgang, uitlevering en resultaat.

Daarnaast is het projectteam verantwoordelijk voor de daadwerkelijke structurering van de data in EAD 2002 of RiC-O. Dat zal uiteraard in nauwe samenwerking gebeuren met projectteam 2 en Transkribus.

3. Planning

	Begindatum	Einddatum
Vorbereiding		
Data-entry en thesauruswerk		
Methodiek autotranscriptie-pipeline productieklaar		
Methodiek verrijking- pipeline productieklaar		
Digitaal beschikbaar bij bronhouder		13-6-25

Nog te verifiëren / in te vullen data tijdens Tweedaagse op 3 en 4 september

4. Eindproducten

- Een output in EAD en/of RiC-O
- Methodiek voor bredere toepassing in archiefsector
- Bronhouder heeft verrijkte data ontvangen
- Bronhouder maakt de data digitaal beschikbaar in eigen infrastructuur

DEEL 4 Evaluatie en vooruitblikken

1. Onderdelen en werkzaamheden

- Afsluitende presentatie pilot fase 1 (Coördinatie: programmamanager)
- Evaluatie verloop, voortgang, samenwerking pilot fase 1 (Coördinatie: programmamanager)
- Evaluatie deliverables pilot fase 1 (Coördinatie: programmamanager)
- Vooruitblikken pilot 2 en 3 (Coördinatie: programmamanager)

Projectteams 1,2,3 en Thuisgroep 4	Deelnemers	Materieel
Programmamanager	- Externe inhuur - Alle medewerkers - Toeleveranciers (optioneel) - Stuurgroep (optioneel)	

2. Plan van aanpak

In de maandsessies worden met elkaar de deliverables getoond, de voortgang besproken en de samenwerking geëvalueerd. Deze 3 kwartalen worden gebundeld en gepresenteerd bij de eindpresentatie.

De projectteams vullen ieder kwartaal het rapportage-template in (bijlage).

Aanwezigheid en betrokkenheid van alle teamcoördinatoren bij de maand en kwartaalsessies noodzakelijk voor een goed verloop. Aanwezigheid en betrokkenheid teamleden projectteams is aanbevolen en gewenst. Tevens zal er een eindevaluatie worden afgenomen en eventuele vervolgstappen besproken.

3. Planning

	Begindatum	Einddatum
Vorbereiding		
Afname eindevaluatie		
Fysieke bijeenkomst (hybride?)		
Presentatie deliverables		
Vooruitblik fase 2 en 3		3-9-26

Nog te verifiëren / in te vullen data tijdens Tweedaagse op 3 en 4 september

4. Eindproducten

- Ca. 300.000 scans conform kwaliteitseisen (zie document)
- Ca. 12 gestructureerde schema's (EAD/RiC-O)
- Methodiek voor toepasbaarheid in eigen organisatie/infrastructuur
- In kaart gebracht 'knelpunten' voor verdere optimalisatie
- Toepasbare evaluatie op sturing en proces, resultaat, samenwerking, leercurve, aanpak.

DEEL 5 Sturing, team en proces

1. Onderdelen en werkzaamheden

- Projectcoördinatie
- Contact met alle stakeholders
- Communicatie over het project
- Secretariële ondersteuning van de teamcoördinatoren
- Organiseren van bijeenkomsten en vergaderingen
- Penvoerder subsidie

Projectleider	Deelnemers	Materieel
Programmamanager	- Projectgroep - Teamcoördinatoren - Stuurgroep - Communicatie-adviseur (HBB/KVAN) - Adviseur (KNAW)	- Bijeenkomsten - Communicatiekosten

2. Plan van aanpak

Hoe wordt de coördinatie van dit project gerealiseerd? Gezien het groot aantal deelnemende partijen is de verwachting dat het project, mede vanwege zijn experimentele aard en lerend vermogen, goed gemonitord moet worden. Het streven is iedereen in een vroeg stadium zoveel mogelijk autonomie en handvatten te geven om de stappen gedurende het project zelfsturend te kunnen zetten en af te wegen.

Daarnaast is het van **groot belang om telkens te kijken naar wat al wél kan, en eventuele tekortkomingen te noteren én te parkeren als punt van aandacht voor een later stadium** (fase 2, 3 of buiten de pilot om).

KVAN is de opdrachtgever van het project en heeft een stuurgroep ingesteld, onder voorzitterschap van Puck Huitsing. De stuurgroep bestaat uit afgevaardigden op directieniveau van Stichting WO2Net, KNAW Humanities cluster/NDE, CBG, VNG, HUA, en het NA.

De programmamanager is belast met de coördinatie van het project en stuurt op het resultaat. De programmamanager rapporteert aan de stuurgroep over de voortgang van het project op basis van voortgangrapportages (zie Rapportage-template). De coördinatie van dit project behelst:

- a) De interne afstemming binnen de verschillende onderdelen en projectteams.
- b) Het bewaken van het overzicht en resultaat op de verschillende onderdelen (in samenhang op kwaliteit, financiën, doorlooptijd en resultaat.
- c) Het organiseren van de betrokkenheid.

- d) Het organiseren voor kennisdeling en kwartaalsessies voor Supporters/Ambassadeurs en breder geïnteresseerden.
- e) Het organiseren van de input van deze relaties (NDE, Digital Humanities, e.a.) en de mogelijke consequenties voor het project.

Gezien de breedte van samenwerking wordt er gewerkt in drie projectteams. Met wekelijkse stand-ups worden de lijnen kort gehouden en de taken en voortgang transparant. We passen een experimentele, agile werkwijze toe en bespreken de voortgang, samenwerking tijdens de maandsessie. Ieder kwartaal tonen we de resultaten op basis van het rapportage.

De programmamanager stuurt binnen de vastgestelde scope (van het tijdens de Tweedaagse vast te stellen projectplan) op de verschillende projectactiviteiten. De stuurgroep houdt zich vooral bezig met “wat en waarom”, de programmamanager met “hoe”. De projectteams zijn compacte teams van circa 4 of 5 professionals. Het bestaat in ieder geval uit een aangestelde Teamcoördinator aangevuld met een aantal specialisten voor verschillende onderdelen (datacuratie, machinelearning, etc.). Elke Teamcoördinator heeft mandaat op de werkzaamheden vastgelegd in dit werkplan. Wanneer er kosten gemaakt worden die belast worden op dit project, dient de Teamcoördinator dit vooraf en vroegtijdig af te stemmen met de programmamanager.

Er is een wekelijkse stand-up, maandoverleg en kwartaalsessies met de projectteams en betrokkenen. Telkens in volledige transparantie en op voet van gelijkheid. De stuurgroep komt eens in de twee maanden bijeen. **Ieder teamlid, teamcoördinator, de stuurgroep, iedere deelnemer, de programmamanager, iedere supporter maakt onderdeel uit van het gehele lerende netwerk en haar/zijn/diens participatie of inbreng is van essentieel belang voor de voortgang.**

De programmamanager is belast met het ophalen en bundelen van de rapportages over de voortgang in de gehele pilot fase 1.

De betrokken adviseur(s) vanuit KNAW humanities cluster, staat het project bij met vakinhoudelijke bijdrages ter ondersteuning **[nog niet definitief, volgt later meer over]**.

3. Planning

	Begindatum	Einddatum
Vorbereiding	1-3-25	1-9-25
Uitvoering	3-9-25	3-9-26
Tussentapen - Evaluatie?		

Nog te verifiëren / in te vullen data tijdens Tweedaagse op 3 en 4 september

BIJLAGEN

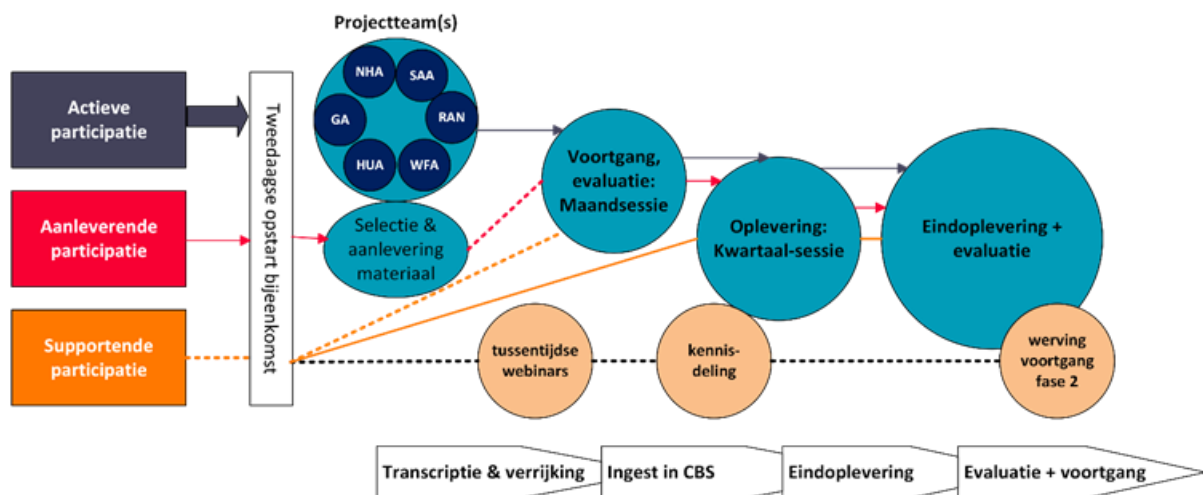


Diagram1 : samenwerking



Diagram2 : communicatie lijnen

De data in diagram 2 en 3 dienen nog geverifieerd te worden tijdens Tweedaagse en vervolgens in de diagrammen gecorrigeerd te worden

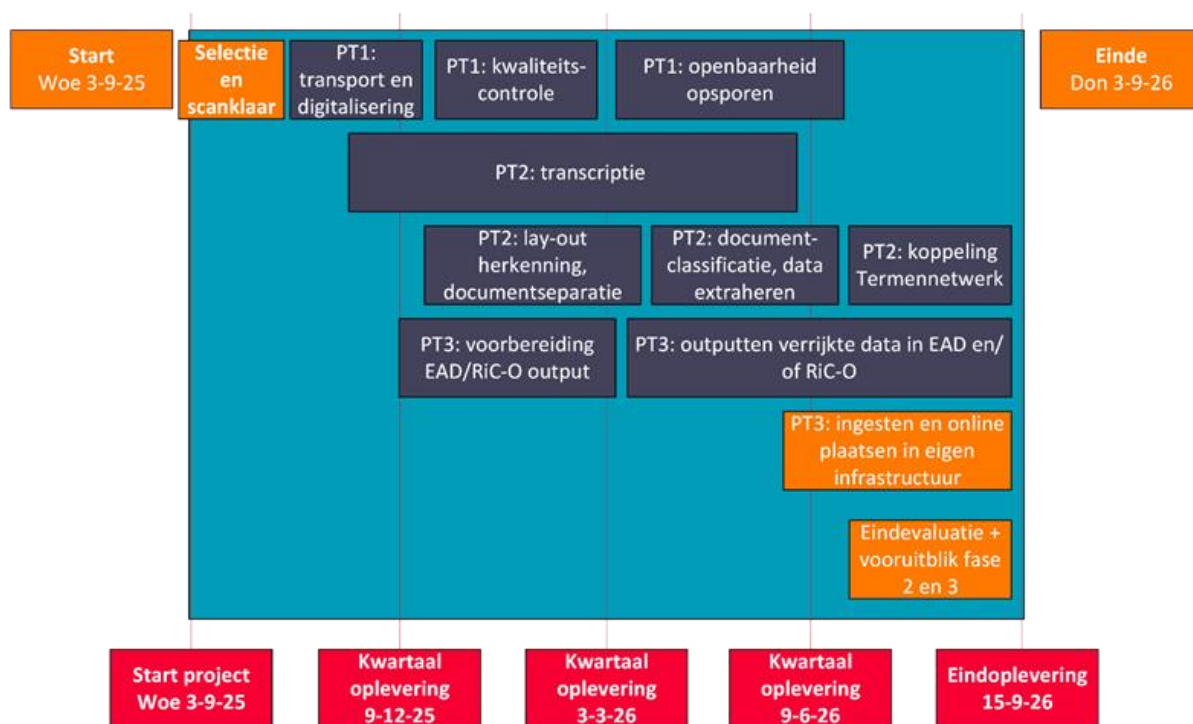


Diagram 3: jaaroverzicht

De data in diagram 2 en 3 dienen nog geverifieerd te worden tijdens Tweedaagse en vervolgens in de diagrammen gecorrigeerd te worden.

Risicomanagement

Periodiek worden de risico's in het project in kaart gebracht om tijdig te kunnen nemen aan de hand van onderstaande criteria.

Nr	Omschrijving	Risico	Impact	Beheersmaatregel
	Niet alles kan beschikbaar worden gesteld wegens privacy beperkingen			
	Kwaliteit en consistentie is zo divers dat verwerking materiaal langer duurt doorlooptijd project			Geen extra tijd beschikbaar, lagere kwaliteit opleveren
	Financiering			Terugschalen
	(On)mogelijkheid het klein te houden			Aansturen op MVP en PT's MVP
	Bestendigen adoptie nieuwe werkwijze			
	Te zware belasting (tijd, bezetting) van projectteam 2			Opschalen vanuit meewerkende archieven of extern aantrekken indien mogelijk