

Tribunaalarchieven als digitale onderzoeksfaciliteit (TRIADO)



NIOD

NETWERK
OORLOGS
BRONNEN



na
nationaal archief

Projectvoorstel Onderzoeksfonds KNAW instituten, 26 september 2016, Amsterdam

Inhoud

1. Achtergrond	3
2. Doelstelling	4
3. Het Centraal Archief Bijzondere Rechtspleging (CABR)	5
4. Vraagstelling.....	8
5. Plan van aanpak	8
Uitwerking onderzoeksvraagstelling 1.....	8
Uitwerking onderzoeksvraagstelling 2.....	11
6. Eindproducten.....	12
7. Exploitatie	12
8. Projectorganisatie	13
9. Planning.....	14
10. Begroting	14
Bijlage 1: Planning.....	15
Bijlage 2: Begroting, dekking en specificatie salariskosten.....	<u>Bijlage 2 is niet opgenomen</u> 16
Bijlage 3: Projectpartners.....	18

Afbeelding voorzijde: Beeldbank WO2, Caféhouder-veehandelaar uit Purmerend en zijn echtgenote staan terecht voor het Bijzonder Gerechtshof te Amsterdam, 29 oktober 1946.

1. Achtergrond

De hoeveelheid 'data' of digital content is zodanig toegenomen dat we over Big Data zijn gaan spreken - hoe problematisch deze term ook is. Om deze rijkdom aan materiaal te ontginnen, moeten er nieuwe instrumenten ontwikkeld worden: instrumenten om de data te bevragen op betekenisvolle inhoud. Daarmee veranderen niet alleen de objecten van onderzoek, maar ook de werkwijzen van geesteswetenschappers. De laatste jaren spreken we over Digital Humanities (DH) als we het hebben over de digitalisering van bronnen en de aanpassing van onze onderzoeks- methoden op deze ontwikkelingen. De term DH omvat veel disciplines, is zowel breed als specialistisch en heeft zowel euforie als weerstand opgeroepen .¹

In de strategische agenda 2016-2020 van de KNAW wordt geschetst hoe binnen de geesteswetenschappelijke instituten samenwerking gestalte zal krijgen door de methodologie van Digital Humanities (DH) als speerpunt te kiezen. Een dergelijke benadering maakt het mogelijk grote corpora van gedigitaliseerde data multidisciplinair te bevragen en zodoende nieuwe onderzoeksperspectieven te creëren. Dit concept gaat wel uit van de aanname dat er relevant bronnenmateriaal reeds is gedigitaliseerd en gestandaardiseerd toegankelijk is gemaakt. Dit is vaak in de praktijk echter niet het geval.

Collecties liggen vaak verspreid over vele organisaties. Alleen rondom het thema Tweede Wereldoorlog zijn er naar schatting meer dan vierhonderd organisaties (archieven, bibliotheken, musea, wetenschappelijke instituten, etc.) die beschikken over origineel bronnenmateriaal. Slechts een fractie hiervan (ca. 8%)² is gedigitaliseerd, een kleiner deel digitaal doorzoekbaar en een nog kleiner deel geschikt voor onderzoek binnen de DH-gemeenschap.

De opbouw en ontsluiting van oorlogscollecties weerspiegelen de logica en het belang van de archiefvormer, maar sluiten daardoor niet noodzakelijk aan bij de behoeften van de gebruikers van het archief. Zo zijn justitiële archieven vaak uitsluitend ontsloten op de naam van de aangeklaagden. De inhoud van de dossiers blijft een raadsel. Dat maakt cumulatief zoeken naar in principe eenvoudige vragen als “waar”, “wanneer”, “waarom” of “wie nog meer?” (plaats en tijd delict, motief, getuigen, slachtoffers en medeplichtigen) tot nu toe een uitermate omslachtig en tijdrovend proces. Onderzoekers moeten zich nu handmatig door stapels papieren worstelen. Statistisch onderzoek is alleen maar mogelijk door middel van hele kleine, bewerkelijke steekproeven.

Onderzoekers die met digitaal instrumentarium archiefcollecties willen benutten, lopen tegen een optelsom van problemen aan: veel belangrijk bronnenmateriaal is überhaupt niet gedigitaliseerd, niet nader ontsloten en niet of nauwelijks gecontextualiseerd met bijvoorbeeld verbindingen naar andere collecties of datasets. Binnen de erfgoedsector is dit probleem inmiddels helder en algemeen onderkend. Op instigatie van het ministerie van OCW is in 2015 de uitwerking van de Nationale Strategie Digitaal Erfgoed gestart, dat uitgaat van verbinden van collecties en het toegankelijk maken daarvan. De KNAW is hierin aangewezen als het knooppunt voor de sector wetenschap. De KNAW ziet het als haar

¹José van Dijck, Ketelaarlezing 2014

²Zie Enumerate-enquête: www.den.nl/art/uploads/files/Publicaties/Enumerate-core-survey-NL2013-2014.pdf

taak actuele wetenschappelijke kennis in te brengen, met name op het gebied van Digital Humanities en wetenschappelijke infrastructuren.

Het Netwerk Oorlogsbronnen wordt vanuit het instituut NIOD-KNAW gecoördineerd met tot doel de collecties op het thema Tweede Wereldoorlog in het toenmalige Koninkrijk der Nederlanden met elkaar te verbinden, de digitale toegang te verbeteren, bronnen te verrijken, deze bruikbaar te maken voor wetenschappelijk onderzoek en voor organisaties die de collecties willen gebruiken voor publieksdoeleinden.

DH-onderzoek experimenteert met een veelheid aan nieuwe methodieken en technologieën om data te verwerken en te analyseren, vaak toegespitst op hele specifieke onderzoeksvragen. Research communities verenigen zich en wisselen tools en data uit binnen digitale onderzoeksinfrastructuren zoals CLARIAH, DARIAH, EHRI, PARTHENOS en andere initiatieven. De technologie die hier ontwikkeld wordt op bijvoorbeeld het terrein van Named Entity Recognition, Handwritten Text Recognition en Topic Modelling is in potentie zeer beloftevol om omvangrijke erfgoedcollecties nader te ontsluiten. De stap van experiment naar toepassing wordt echter zelden gezet. Daarnaast is er binnen de digitale humaniora een groot gebrek aan “bruikbare” digitale corpora, wat ertoe leidt dat het onderzoek in Nederland zich vooral toespitst op het krantencorpus van de KB en de Staten Generaal Digitaal. Als de erfgoedwereld al collecties heeft gedigitaliseerd, is het zelden gereed gemaakt voor digitaal onderzoek.

2. Doelstelling

Het TRIADO-project brengt de erfgoedsector en het wetenschappelijk onderzoek nader tot elkaar door het innovatief potentieel van de Digital Humanities aan te boren. Voor deze pilot wordt geëxperimenteerd met documenten uit het meest geraadpleegde justitiële archief in Nederland: het Centraal Archief Bijzondere Rechtspleging (CABR), beheerd door het Nationaal Archief.

Het project sluit aan bij de strategische agenda van de KNAW en de Nationale Wetenschapsagenda; het terrein van de Digital Humanities staat voor de enorme uitdaging om een groot aantal complexe bronnen met elkaar te verbinden en daarmee de meerwaarde van ons historisch-culturele erfgoed te vergroten.³ Tevens geeft de KNAW op deze wijze invulling aan haar rol als knooppunt binnen het Netwerk Digitaal Erfgoed, door enerzijds de wetenschappelijke kennis toe te voegen (op het terrein van DH-methodieken en inhoudelijke verrijking van de bronnen) en anderzijds de thematische, cross-sectorale ontsluiting via thematische netwerken te verstevigen. Uiteindelijk moet dit leiden tot een sprong voorwaarts in interdisciplinaire wetenschappelijke samenwerking, waarbij digitale

³ Wetenschapsagenda, clustervraag 111; Gaat digitalisering ons erfgoed redden? Hoe verbeteren en vergemakkelijken we de ontsluiting en toegang tot ons digitaal erfgoed? Ook van toepassing op dit project is clustervraag 056: Hoe kunnen vrede, veiligheid en coöperatie worden bevorderd – en bedreigingen en schendingen geadresseerd – binnen en tussen nationale rechtsstaten en binnen en tussen groepen en samenlevingen?

onderzoeksinfrastructuren en –technologie de vraagstellingen en benaderingen van de humaniora, de mens- en maatschappijwetenschappen en de rechtswetenschappen kunnen integreren.

Dit gebeurt door:

- het toepassen van bestaande ontsluitingsmethodes op een archief van substantiële omvang met als doel te kijken welke methode het meest effectief en efficiënt is;
- het onderzoeken of op deze wijze de collectie ook verrijkt kan worden zodat de bronnen bruikbaar zijn voor onderzoekers;
- het onderzoeken of het archief op deze wijze ontsloten kan worden, verbonden met andere archieven;
- het toetsen of er via een gekozen ontsluitingsmethode de mogelijkheid bestaat betrouwbare statistische patroonanalyses te genereren op basis van nieuwe onderzoeksvragen.

3. Het Centraal Archief Bijzondere Rechtspleging (CABR)

Het CABR, beheerd door het Nationaal Archief, is het meest geraadpleegde WO2-archief in Nederland. Het omvat 4,5 strekkende kilometer aan processtukken en appendices van ca. 300.000 personen die na de Tweede Wereldoorlog verdacht werden van collaboratie met de vijand. In het CABR zijn zowel de dossiers van de lokale opsporingsdiensten als de dossiers van het Directoraat voor de Bijzondere Rechtspleging, het openbaar ministerie, de tribunalen, de bijzondere gerechtshoven en de bijzondere raad van cassatie en de Stichting Toezicht Politieke Delinquenten terug te vinden. De strafdossiers, geordend op naam van de verdachte, zijn niet gedigitaliseerd. Deze dossiers kenmerken zich door een grote variëteit aan documenten: gedrukt materiaal, getypt materiaal, handgeschreven materiaal, voorgedrukte formulieren, foto's etc. (zie figuur 1). De kwaliteit van het papier (daterend van tijdens of direct na de Tweede Wereldoorlog) is slecht. Het CABR omvat een schat van informatie voor historici, maar ook bijv. rechtsgeleerden, sociologen, criminologen en (sociaal)-psychologen. In 2015 werd dit archief 3100 maal geraadpleegd en er werden ongeveer 10.000 dossiers ter inzage gegeven. De onderzoeker moet nu handmatig door stapels originele documenten bladeren. Het archief is slechts toegankelijk gemaakt op de naam van de verdachte en diens geboortedatum, niet op de inhoud.



Figuur 1: voorbeeld van CABR-dossier

Het CABR is op dit moment in zeer beperkte mate ontsloten. Dit terwijl er een schat aan informatie binnen de verschillende individuele dossiers voorhanden is. De geschiedschrijving van de Tweede Wereldoorlog heeft al vrij veel gebruik gemaakt van de informatie uit de in het CABR aanwezige strafdossiers, zowel wat betreft de geschiedenis van het verzet (en de vervolging daarvan door de Duitse bezetter) als de geschiedenis van collaboratie of lokale oorlogsgeschiedenis. Voorbeelden zijn de monografie van Madelon de Keizer over de geschiedenis van het illegale Parool, het onderzoek van Ismee Tames naar de naoorlogse re-integratie van NSB'ers, of het onderzoek van Bas von Benda-Beckmann over de Velsler-affaire. Het Nationaal Archief heeft recentelijk twee kleine pilots laten uitvoeren door Ad van Liempt.⁴ De monografieën maar ook deze pilots, gebaseerd op het grondig doorzoeken van enkele dossiers, tonen de grote onderzoekspotentie aan van de uitermate rijke en diverse informatie die in de dossiers zit. Die blijft bepaald niet beperkt tot de daders op wie de stukken betrekking hebben. Digitalisering maakt het mogelijk op veel grotere schaal en veel effectiever op zoek te gaan naar patronen van geweld, collaboratie, repressie en bevrijding. Dit leidt niet alleen tot diepere inzichten maar opent de deur tot het maken van nieuwe verbindingen en het maken van tot dusver onbereikbare kwantitatieve analyses.

⁴ Ad van Liempt/Jan H. Kompagnie (red), *Jodenjacht*, Amsterdam, 2011 en Ad van Liempt (red), *De Jacht op het verzet*, Amsterdam, 2013

Tegelijkertijd kan er na digitalisering een relatie worden gelegd met andere belangrijke oorlogsarchieven en – collecties, zoals het (eveneens nog te digitaliseren) Nederlands Beheersinstituut en het Nederlandse en Internationale Rode Kruis. Het project zal daarom een voorbeeldwerking en een stimulerende uitwerking hebben op de verdere ontwikkeling van de oorlogsarchieven als onderzoeksfaciliteiten, zowel in Nederland als internationaal (International Tracing Service, EHRI, etc.).

```
<TextLine HEIGHT="52" WIDTH="1379" VPOS="413" HPOS="832">
  <String WC="0.2349999994" CONTENT="De" HEIGHT="38" WIDTH="57" VPOS="416" HPOS="832"/>
  <SP WIDTH="32" VPOS="416" HPOS="890"/>
  <String WC="0.2240000069" CONTENT="Bijzonder*" HEIGHT="48" WIDTH="300" VPOS="415" HPOS="923"/>
  <SP WIDTH="28" VPOS="415" HPOS="1224"/>
  <String WC="0.1700000018" CONTENT="Raad" HEIGHT="41" WIDTH="118" VPOS="415" HPOS="1253"/>
  <SP WIDTH="32" VPOS="416" HPOS="1372"/>
  <String WC="0.1666666716" CONTENT="van" HEIGHT="31" WIDTH="87" VPOS="423" HPOS="1405"/>
  <SP WIDTH="28" VPOS="415" HPOS="1493"/>
  <String WC="0.3633333445" CONTENT="Cassatie," HEIGHT="49" WIDTH="267" VPOS="415" HPOS="1522"/>
  <SP WIDTH="34" VPOS="415" HPOS="1790"/>
  <String WC="0.3083333373" CONTENT="Lerata" HEIGHT="39" WIDTH="179" VPOS="415" HPOS="1825"/>
  <SP WIDTH="33" VPOS="415" HPOS="2005"/>
  <String WC="0.3033333421" CONTENT="Kamer." HEIGHT="40" WIDTH="172" VPOS="415" HPOS="2039"/>
</TextLine>
<TextLine HEIGHT="55" WIDTH="1558" VPOS="713" HPOS="834">
```

Figuur 2: voorbeeld van een machineleesbaar bestand na digitalisering (CABR)

Het CABR is van groot wetenschappelijk belang, feitelijk een goudmijn voor een ieder die onderzoek doet naar samenlevingen onder ongewone spanning van oorlog, politieke repressie en gewelddadige vervolging. Het is ook interessant omdat het de weerslag vormt van een unieke vorm van rechtspleging, in het leven geroepen na een ontwrichtende periode van grootschalig geweld. Juist in Nederland, met de internationale fora op het terrein van internationaal recht en de Tribunalen die in Nederland gevestigd zijn, is onderzoek op dit terrein van belang. Met name het onderzoek naar Transitional Justice, rechtsherstel, internationale rechtspraak en de gevolgen daarvan op de samenleving (zie interdict Tilburg) kent specifieke vragen over ontwikkeling en effect die tot dusver alleen kwalitatief, op grond van doelgerichte steekproeven (Belifante, Romijn, Meihuizen) zijn beantwoord. Een alomvattende analyse van de Nederlandse politieke transitierechtspraak op basis van het hele corpus is tot nu toe niet mogelijk gebleken.

Voor de digitale, geautomatiseerde ontsluiting van archiefmateriaal zijn er weliswaar enkele verkenningen en pilotstudies uitgevoerd, maar een duidelijk toepasbare werkmethode heeft dit nog niet opgeleverd. In dit project zal deze overgang van “laboratorium naar *reality check*” worden beproefd. Naast het bevorderen en aanjagen van nieuw onderzoek is dit dus een bijkomende innovatie van dit project. Extra uitdaging bij het ontsluiten van archiefmateriaal is het gegeven dat de kwaliteit van de machineleesbare tekst vaak minder is voor handgeschreven of getypte tekstbronnen dan bij gepubliceerd materiaal. (zie figuur 2) De omgang met grote hoeveelheden “imperfecte”, ongestructureerde data ten behoeve van specifieke onderzoeksvragen, vormt een substantieel onderdeel van dit project.

4. Vraagstelling

Het project kenmerkt zich door een pilot karakter. Onderzocht wordt of er een mogelijkheid is om het CABR zodanig digitaal te ontsluiten dat het geschikt is als basis om uiteenlopende wetenschappelijke vraagstellingen te kunnen beantwoorden. Bij een positief oordeel zal op basis van deze uitkomst verdere financiering worden gezocht bij diverse subsidieverstrekkingen, met name NWO. Metingen gedaan in dit project worden gebruikt als uitgangspunt bij een vervolgaanvraag.

Het project heeft twee onderzoeksvragen, die feitelijk in elkaars verlengde liggen:

1. Welke digitale methoden zijn het meest geschikt (in termen van kwaliteit, efficiency, etc.) om grote corpora van ongestructureerd, imperfecte data, gebaseerd op analoge archiefcollecties, geschikt te maken als onderzoeksfaciliteit? Nadruk ligt hierbij op het toepasbaar maken van al beschikbare tools uit het digitale domein (van laboratorium naar *reality check*). De focus hierbij ligt vooral op het verbeteren van de toegangen “wie, wat, waar en wanneer”. [GENERIEK DEEL]
2. Is het mogelijk op basis van de gekozen ontsluitingsystematiek antwoord te geven op specifieke, vooral kwantitatief-statistische wetenschappelijke onderzoeksvragen? Een projectbegeleidingsgroep onder voorzitterschap van prof.dr. P. Romijn formuleert een aantal vraagstellingen vanuit verschillende vakgebieden, waarvan de beantwoording afhankelijk is van toepassing van de onder 1 ontwikkelde toegangen en evt. aanvullende data-curatie. [SPECIFIEK DEEL]

Dit project heeft in twee opzichten internationale potentie:

- a. Het is een *reality check* van tools en methoden die worden ontwikkeld binnen CLARIAH en het READ-project. Er zijn op dit gebied ook contacten met DARIAH-DE (Duitsland).
- b. Een beter toegankelijk CABR maakt de faciliteit aantrekkelijk voor aansluiting en samenwerking met EHRI, internationaal gerenommeerde archieven zoals bijvoorbeeld het International Tracing Service, het Bundesarchiv en andere tribunaalarchieven zoals bijvoorbeeld Neurenberg, Tokio Tribunaal, Joegoslavië-Tribunaal, Gacaca-archief (Rwanda), etc.

5. Plan van aanpak

Uitwerking onderzoeksvraagstelling 1

In 2015-2016 heeft Netwerk Oorlogsbronnen samen met het Nationaal Archief, het IMPACT Centre of Competence en het Centre for Language and Speech Technology (Radboud Universiteit Nijmegen) een verkennend onderzoek gedaan naar de kwaliteit van Optical Character Recognition- en named entity software, toegepast op een kleine testset van verschillende documenten uit het CABR (project Volautomatische Archiefontsluiting). Naast aanbevelingen voor de optimalisering van scans en

software, wees het onderzoek uit dat met standaard software (Abby FRS 11 SDK) gemiddeld ongeveer 4/5 van de tekst correct machineleesbaar kan worden gemaakt.⁵

Naam en voornamen: **GROEN,HENDRIK** Geboorteplaats en datum: **Zaandam, 29 October 1897**
Echtgenoot van / Beroep: voorheen **agent van Politie**, t Laatste woonplaats en adres: **Kanaalstraat 25 II Amsterdam** **Persoonsbewijs**-no.:z 2 01239 afgegeven te **Leeuwarden** Nationaliteit (evt. vroeger) **Nederlander** die ervan **verdacht** wordt: **joodsche personen** in macht van den **vijand** te hebben gebracht,terwijl hij in dienst was van de **S.D.** Terzake gehoord, verklaarde **verdachte** mij het volgende: dat hij **in dienst was getreden** van **Lippmann en Rosenthal** voor de inventarisatie van **joodsche goederen**,vervolgens overgegaan naar de **S.D.,afdeeling joodsche zaken** te **Amsterdam** Verdachte heb ik, optastvan den Chef **Opsporingsdienst D.P.M.** op **28 Mei 1945** **bewaring** , toegesteid, in het **Huis van Bewaring I** te **Amsterdam P.O.D. Amsterdam. Model A**

Domeinspecifieke woorden

Personen

Organisaties

Geografische lokatie

Datum

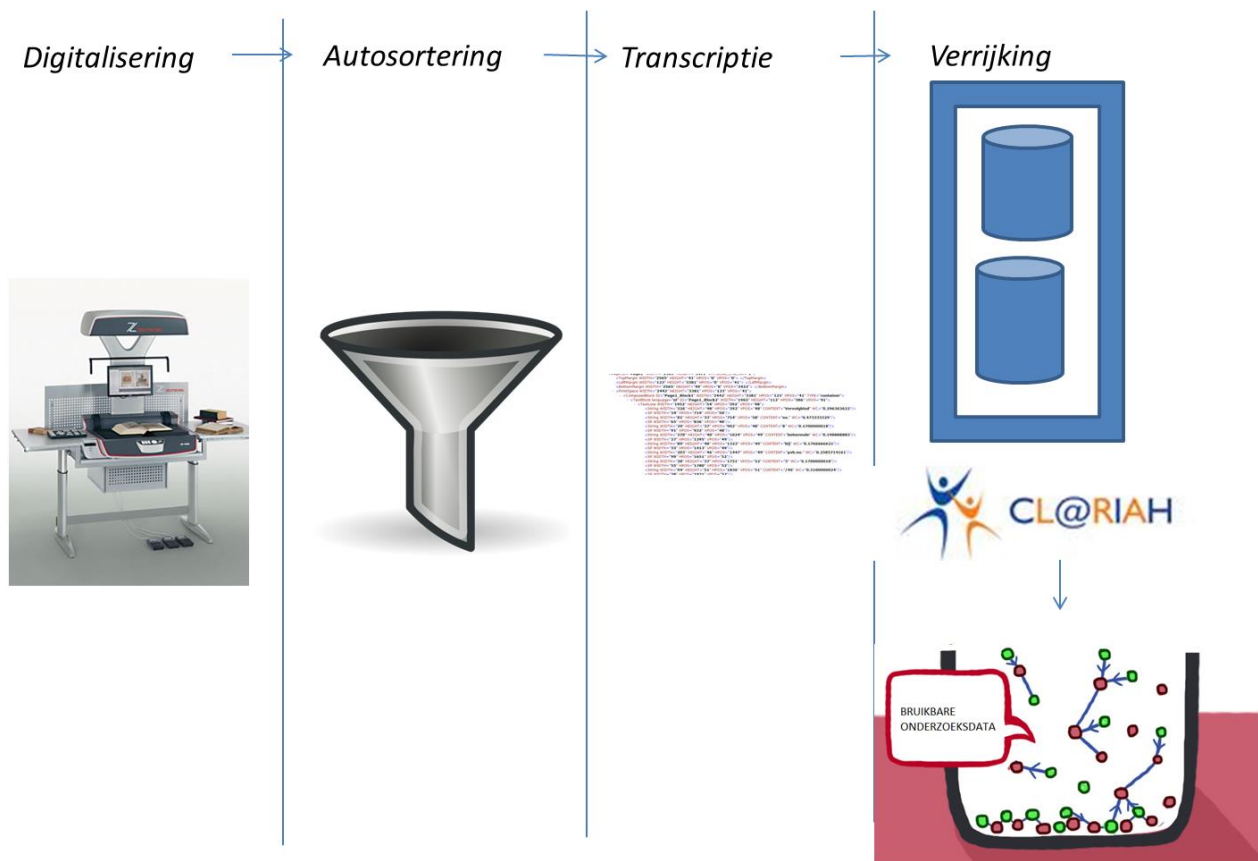
Figuur 3: fragment uit CABR-document met word error rate (= % incorrect omgezette woorden) van 19,45%. Het CABR is een rijke informatiebron vanwege de talloze, grotendeels eenvormige vermeldingen van plaats, tijd/datum, persoonsnamen, organisatienamen en domeinspecifieke termen en uitdrukkingen.⁶

In het TRIADO-project onderscheiden we vier stappen in de workflow (zie figuur 4):

1. Digitalisering
2. Autosortering
3. Transcriptie
4. Verrijking

⁵ Link naar rapport of project, zie <https://archieff2020.nl/projecten/volautomatische-archieffontsluiting>.

⁶ In verband met privacy van betrokken personen zijn de namen, datums en plaatsnamen gefingeerd in dit voorbeeld.



Figuur 4: workflow TRIADO-project

Stap 1: Digitalisering

Op basis van een door de begeleidingscommissie vastgestelde steekproef zal circa veertien meter aan archiefmateriaal gescand en gemetadateerd worden conform de standaard kwaliteitsnormen van het Nationaal Archief en de aanbevelingen, voortgekomen uit de bevindingen van het project Volautomatische Archiefontsluiting (zie hierboven).

Stap 2: Autosortering

Het CABR is een heel divers archief: het omvat foto's, getuigenverklaringen, originele correspondentie (bewijsstukken), vragenformulieren, lidmaatschapskaarten, etc. Softwarematig, op basis van fysieke gelijkens (vormherkenning), zullen documenten worden gesorteerd. Dit stelt ons in staat de transcriptiesoftware meer gericht in te zetten voor specifieke soorten documenten. Er wordt ook geëxperimenteerd met autosortering na transcriptie, om bijvoorbeeld op basis van inhoudelijke formuleringen in de tekst, documenten te groeperen. Bijvoorbeeld: vonnissen en formulieren bevatten

standaardformuleringen waarop geselecteerd kan worden. De begeleidingscommissie geeft aanwijzingen voor de gewenste filteringen.

Stap 3: Transcriptie

De naar schatting circa 100.000 scans (veertien strekkende meter archief) worden softwarematig met Abby Finereader 11 SDK omgezet in ALTO-bestanden met machineleesbare tekst. De aanbevelingen uit het project Volautomatische Archiefontsluiting worden meegenomen in het proces van de voorbereiding van de scans en de instellingen van de software. Voor het handgeschreven materiaal wordt geëxperimenteerd met een afwijkende workflow via de Transkribus-tool.⁷

Stap 4: Verrijking

De ALTO-bestanden worden verrijkt met tools die in het kader van het CLARIAH-project zijn ontwikkeld. Met behulp van externe ontologieën (bestaande namenlijsten van personen en organisaties uit de Tweede Wereldoorlog, de WO2-thesaurus van Netwerk Oorlogsbronnen, etc.) wordt er een post-correctieslag gemaakt bovenop de bestaande machineleesbare tekst. De verrijkingen worden op gestandaardiseerde wijze in de bestanden verwerkt, zodat er aan het einde van de workflow een verrijkte set van digitale onderzoeksdata overblijft die kan worden aangewend voor nader onderzoek.

Binnen de omschreven workflow wordt ruimte genomen om te experimenteren. Ook worden er in elke slag kwaliteitsmetingen gedaan. De bevindingen en aanbevelingen worden opgenomen in een onderzoeksrapport, dat bedoeld is om het proces van analog archief naar bruikbaar digitale onderzoeksdata verder te specificeren. Uitgangspunt is dat het onderzoeksrapport ook voor andere collectiehoudende instellingen als adviesrapport en handleiding fungeert. De resultaten worden daarnaast ook middels in ieder geval twee wetenschappelijke artikelen en deelname aan relevante conferenties wereldkundig gemaakt.

Uitwerking onderzoeksvraagstelling 2

Het CABR is een goudmijn voor een ieder die onderzoek doet naar samenlevingen onder ongewone spanning van oorlog, politieke repressie en gewelddadige vervolging. Het CABR heeft betrekking op Nederland tijdens en na de Tweede Wereldoorlog, maar de bijzondere rechtspleging – een unieke juridische constructie zowel vanuit nationaal als internationaal perspectief – is ook exemplarisch voor andere moderne samenlevingen. Het CABR is veel meer dan louter een archief met strafdossiers; in feite gaat het om een amalgaam van politieke misdrijven en commune-criminaliteit. De dossiers weerspiegelen hoe individuele Nederlanders en de samenleving omgingen met de spanning van oorlog en bezetting. Het CABR kan daarom worden beschouwd als een spiegel van de samenleving tijdens en na de oorlog: het biedt een caleidoscopisch beeld van met de oorlog verbonden individuele en

⁷ <https://transkribus.eu/Transkribus/>

collectieve acties, vervolging en verzet, onrecht en rechtspraak, maar toont ook politieke en sociale transformaties, *coping strategies* en botsende normen, waarden en mentaliteiten.

Deze thema's behoren niet toe aan één specifiek wetenschappelijk domein. Het CABR geldt bij uitstek als een archief dat voor een breed scala aan disciplines (geschiedwetenschap, rechtswetenschap, mens- en maatschappijwetenschappen, criminologie, sociologie en humaniora in brede zin zoals bijvoorbeeld culturele studies, filosofie en religiewetenschappen) aantrekkelijk is. Het TRIADO-project biedt een uitstekend platform om samenwerking tussen deze disciplines verder te concretiseren door gezamenlijk digitale technologieën toe te passen en zo nieuwe inhoudelijke verbanden bloot te leggen.

Een multidisciplinaire begeleidingscommissie zal zich buigen over de formulering van enkele wetenschappelijke vraagstellingen. Focus hierbij zal liggen op vraagstellingen naar frequentie en spreiding (geografisch, leeftijd, geslacht, etc.), in combinatie met andere relevante gegevens die geëxtraheerd kunnen worden (bijvoorbeeld beroepen, stafmaat, opleidingsniveau, lidmaatschap van organisaties). Hiermee kunnen nieuwe wetenschappelijke inzichten ontstaan over bijvoorbeeld de verschillen tussen regionale tribunaal en individuele rechters, sociale stratificatie van politieke delinquenten, de geografische spreiding van collaboratie (stad versus platteland), profielen van slachtoffers, het sociale netwerk van collaborateurs, verschillen in gepleegde delicten en strafmaat tussen man en vrouw, etc..

De begeleidingscommissie maakt keuze uit een beperkt aantal vraagstellingen. Deze vraagstellingen worden verder uitgewerkt door een onderzoeker, in combinatie met een data engineer, die gezamenlijk de onder 1. opgeleverde dataset, gaan bevragen. De bevindingen, zowel inhoudelijk als ook met betrekking tot de methodiek, worden vastgelegd in een onderzoeksrapport. Daarnaast worden minimaal twee wetenschappelijke artikelen over dit onderdeel gepubliceerd en zullen de resultaten ook op relevante conferenties en symposia publiekelijk worden gemaakt. Het doel van de vraagstelling is te toetsen of de onder 1. geproduceerde dataset ook daadwerkelijk bruikbaar is.

6. Eindproducten

- Semantisch gestructureerde, voor onderzoek bruikbare set van onderzoeksdata, op basis van circa veertien meter gedigitaliseerd archief.
- Tools die onderzoekers in staat stellen specifiek op WIE, WAT, WANNEER en WAAR te selecteren voor nader onderzoek.
- Twee onderzoeksrapporten.
- Tenminste vier wetenschappelijke publicaties in relatie tot doelstelling 1 en 2.

7. Exploitatie

De ambitie is om de opgedane kennis, de onderzoeksdata en ontwikkelde workflow als basis te nemen voor een gezamenlijke NWO-groot aanvraag. Op langere termijn, met het oog op het wetenschappelijk belang van het CABR voor een brede groep onderzoekers, willen we het CABR nomineren voor opname in de Nationale Roadmap Grootschalige Onderzoeksfaciliteiten. Het streven is om in 2025, wanneer het gehele CABR zonder restricties openbaar beschikbaar mag worden gesteld, middels een digitale onderzoeksfaciliteit vrijelijk het volledige archief ter beschikking te stellen aan onderzoekers en andere geïnteresseerden.

Het CABR biedt allerlei inhoudelijke aanknopingspunten tot verbindingen met internationale dataverzamelingen zoals de International Tracing Service, de NSDAP-collectie van het Bundesarchiv, de collecties van Yad Vashem en US Holocaust Memorial Museum, EHRI, etc. In bredere zin – als weerslag van een uniek rechtssysteem dat is ingesteld na een periode van grootschalig geweld – biedt een digitaal bruikbaar CABR ook een uitstekend uitgangspunt om vergelijkingen te maken met bijvoorbeeld de tijdelijke rechtssystemen die zijn ingesteld in Rwanda en Zuid-Afrika.

8. Projectorganisatie

Het NIOD – in de rol van penvoerder van programma Netwerk Oorlogsbronnen – fungeert als hoofdaanvrager van het project. Het NOB levert de projectleider en ICT-projectleider en neemt de rol van “opdrachtgever” op zich.

Het Huygens ING vervult binnen het project de rol van “opdrachtnemer” en zal in die hoedanigheid het grootste deel van de werkzaamheden uitvoeren middels een data-manager en data-engineer.

Het Nationaal Archief stelt een deel van het CABR ter beschikking aan dit project en levert een coördinator, een collectiespecialist. Het Nationaal Archief organiseert de digitalisering van het materiaal op basis van projectspecificaties.

Het NIOD stelt een onderzoeker ter beschikking voor de nadere uitwerkingen van onderdeel 2 van de vraagstelling.

Een multidisciplinaire begeleidingscommissie onder voorzitterschap van prof. dr. P. Romijn (NIOD) fungeert als stuurgroep van het project. Vertegenwoordigers van de projectpartners, aangevuld met onderzoekers uit verschillende disciplines, nemen zitting in de commissie.

Het project TRIADO vormt organisatorisch een onderdeel van het programma Netwerk Oorlogsbronnen.⁸ Binnen NOB wordt gewerkt volgens de managementmethodiek Programmatisch Creëren. Er wordt gewerkt met maandelijkse voortgangsrapportages.

Er wordt in dit project gewerkt met beperkt openbare collectiedata. Het Nationaal Archief, als eigenaar van de collectie, voert de regie over het beheer van de data in het project conform de door haar bepaalde voorschriften.

⁸ <http://www.netwerkoorlogsbronnen.nl/nieuws/samen-naar-collectie-wo2-nederland>

9. Planning

Het project loopt van 11 januari 2017 tot 1 juli 2019. Zie bijlage 1 voor gedetailleerd overzicht.

Bijlage 1: Planning

Onderdeel	Periode	Eindproducten
Definitiefase	11-1-17	Start project
	17-1-17	Expert meeting met onderzoekers, collectiespecialisten en data engineers
	28-2-17	Projectcontract met gedetailleerd plan van aanpak en planning
Uitvoeringsfase	1-3-17 tot 1-1-18	Digitalisering: -circa 100.000 tiff-bestanden (kleur), -circa 100.000 ALTO-bestanden met ge-OCR-de tekst -metadata (inhoudelijk, technisch, structureel)
	2-1-18 tot 28-9-18	Autosortering, transcriptie en post-correctie: - verrijkte versies van de 100.000 ALTO-bestanden - onderzoeksrapport vraagstelling 1 - minimaal twee wetenschappelijke artikelen, disseminatie op conferenties
	1-10-18 tot 28-6-19	Toetsing wetenschappelijke vraagstellingen: -op maat verrijkte datasets op basis van vastgestelde wetenschappelijke vraagstelling - onderzoeksrapport vraagstelling 2 - minimaal twee wetenschappelijke artikelen, disseminatie op conferenties
Afsluitingsfase	1-7-19	Evaluatie van bevindingen en aanbevelingen voor vervolgonderzoek ten behoeve van NWO groot-aanvraag en nominatie CABR voor Roadmap Grootschalige Onderzoeksfaciliteiten.

Bijlage 3: Projectpartners

Het Netwerk Oorlogsbronnen

Het Netwerk Oorlogsbronnen (NOB), gecoördineerd door het NIOD, is een netwerkorganisatie van WO2-erfgoedinstellingen, dat zich ten doel heeft gesteld de gezamenlijke, digitale toegang tot de collectie WO2 Nederland te verbeteren. Het programma houdt zich onder andere bezig met linked (open) data, automatische tekstherkenning, crowdsourcing en andere methodes om collecties beter toegankelijk te maken. Het NOB onderhoudt nauwe contacten met onder meer het door OCW gestimuleerde Netwerk Digitaal Erfgoed (KNAW is hier knooppunt voor de wetenschap). Gezien het eminente belang van het CABR voor de gehele oorlogserfgoedsector is het Netwerk Oorlogsbronnen met een Nationaal Archief een samenwerking aangegaan om de toegang tot het CABR te verbeteren. Binnen NOB is er al een proefproject gedaan rondom automatische tekstherkenning, het CABR en named entity recognition (project Volautomatische Archiefontsluiting).

Zie: programmaplan Netwerk Oorlogsbronnen, www.oorlogsbronnen.nl/programmaplan2015

Huygens ING

Het Huygens ING heeft in verschillende projecten (CLARIAH, Transkribus, READ) ruime ervaring opgedaan met automatische tekstherkenning en metadata-extractie. Het geschikt maken van ongestructureerde archiefcollecties voor onderzoek is een van de kernactiviteiten van het Huygens ING (bijv. project rondom Resoluties van de Staten Generaal, Geleerdenbrieven, Golden Agents: creatieve industrie en de creatie van de Nederlandse Gouden Eeuw).

Zie: www.huygens.knaw.nl

Nationaal Archief

Het Nationaal Archief is door de minister van OCW in de Archiefvisie aangewezen als trekker voor de versterking van de kennisfunctie binnen de archiefsector. Het Nationaal Archief is middels het DTR-programma (Digitale Taken Rijksarchieven) eind 2013 gestart met een grootschalig digitaal innovatietraject. De bedoeling is om in 10 jaar 10% van de volledige rijkscollectie digitaal toegankelijk te maken (de rijkscollectie wordt niet alleen door het Nationaal Archief beheerd, maar grote delen van de collectie worden ook beheerd door Regionaal Historische Centra). Het Nationaal Archief geldt binnen de archiefsector als autoriteit op het gebied van digitale collecties en vervult in deze hoedanigheid als vraagbaken en expertisecentrum voor de sector.

Zie: www.gahetna.nl/vraagbaak/onderzoeksgids/ik-zoek-veroordeelden-centraal-archief-bijzondere-rechtspleging

NIOD Instituut voor Oorlogs-, Holocaust en Genocidestudies (NIOD-KNAW)

Het NIOD Instituut voor Oorlogs-, Holocaust- en Genocidestudies houdt zich sinds 1945 bezig met studie naar de uitwerking van oorlogen, de Holocaust en andere genociden op individu en samenleving. Het bundelt sinds 1945 de inhoudelijke expertise op het gebied van de Tweede Wereldoorlog en zijn nasleep, waaronder collaboratie, verzet en de bijzondere rechtspleging.

Zie: <http://www.niod.knaw.nl>