

# TRIADO brengt het CABR digitaal tot leven

**Het Centraal Archief Bijzondere Rechtspleging (CABR) telt ruim 670.000 dossiers van personen die na de Tweede Wereldoorlog verdacht werden van ‘collaboratie’ met de Duitse bezetter. In het project Tribunaalarchieven als Digitale Onderzoeksfaciliteit (TRIADO) zijn de mogelijkheden om dit archief digitaal te ontsluiten onderzocht.**

Historische archieven zoals het CABR bevatten een rijkdom aan informatie over personen, organisaties, gebeurtenissen, locaties en heel veel meer. De toegang tot het materiaal beperkt zich doorgaans tot algemene aanduidingen: ‘Correspondentie van de NSB in de periode van januari tot december 1934’ of ‘Ingekomen en uitgegane brieven van de Binnenlandse Strijdkrachten, 1945’. Iedereen die wil weten wat er aan informatie in een archief is te vinden, moet de stukken opvragen en er handmatig doorheen bladeren.

Er is allerlei computertechnologie voorhanden om grote hoeveelheden documenten – en vooral de informatie in deze stukken – digitaal doorzoekbaar te maken tot op bladzijdeniveau. Het gaat hierbij met name om software die de inhoud van tekstuele bronnen – getypte of gedrukte documenten, maar ook handgeschreven stukken – omzet naar **machineleesbare tekst**. In 2016 startten het Nationaal Archief, Huygens ING, NIOD Instituut voor Oorlogs-, Holocaust- en Genocidestudies en Netwerk Oorlogsbronnen (NOB) het project TRIADO. De bedoeling was om op basis van een steekproef uit het CABR de mogelijkheden van digitale ontsluiting te verkennen door te experimenteren met ‘proven technology’, ofwel bestaande software, zowel open source als commercieel.

## De digitale toekomst?

Als testcase werd een steekproef van 13,8 strekkende meter (167.197 bladzijden) geselecteerd uit het CABR. Dit archief bestaat uit processen-verbaal, besluiten, inlichtingenformulieren, lidmaatschapskaarten, foto’s en allerlei archiefmateriaal dat als bewijsstukken is toegevoegd aan de dossiers. Er werden twee onderzoeksvragen gesteld.

1. *Welke digitale methoden zijn het meest geschikt om grote hoeveelheden data, gebaseerd op analoge archiefcollecties, geschikt te maken voor onderzoek? De focus ligt hierbij vooral op de toegangen ‘Wie, Wat, Waar en Wanneer’.*
  - In TRIADO werd de steekproef met standaard OCR (Optical Character Recognition)-software **machineleesbaar** gemaakt. Deze ruwe data vormen de basis voor een full-tekst zoekfunctie. Vervolgens is er onderzocht in hoeverre er nadere toegangen door de computer kunnen worden gegenereerd. **‘Named entity recognition’**-software is gebruikt om namen van personen, organisaties, producten en locaties automatisch te herkennen. Het matchen aan bestaande databestanden bleek een valide strategie. Met behulp van het bestand van de [Nationale Database](#)

[Vervolgingslachtoffers](#) (NDVS) konden in de 13,8 meter testmateriaal al enkele slachtoffers geïdentificeerd worden.

- In TRIADO is ook geëxperimenteerd met het **automatisch herkennen van documenttype**. Voor 28 vooraf bepaalde categorieën (processen-verbaal, besluiten, etc.) zijn voorbeelden verzameld. De computer is op basis van deze data zichzelf gaan leren documenten te classificeren. De beste scores werden behaald met documenten met een vaste opmaak, zoals bijvoorbeeld NSB-lidmaatschapskaarten.
- Andere nieuwe vormen van archiefontsluiting deden zich voor bij de experimenten met het zogenaamde '**topic modelling**'. Deze technologie analyseert de machineleesbare tekst en geeft op basis daarvan een korte karakterisering van een of meer documenten in steekwoorden.
- Na afloop van alle experimenten is er een **demonstrator** gebouwd. De zoekopdracht 'Oostfront', levert daar in een paar seconden 619 treffers op, met verwijzingen naar documenten waar dit woord gebruikt wordt. Dit is een wereld van verschil met de huidige situatie waarin het CABR alleen op dossierniveau is ontsloten op naam van verdachte.

## 2. *Is het mogelijk om op basis van de gekozen ontsluitingsmethoden antwoord te geven op specifieke, vooral kwantitatief-statistische wetenschappelijke onderzoeksvragen?*

- Voor onderzoekers biedt de digitale ontsluiting nieuwe manieren van zoeken naar patronen. Door combinaties van steekwoorden te maken, kunnen mensen en daarmee groepen worden gevonden die anders verborgen zouden blijven - omdat we hun naam niet kennen of niet aan bepaalde groepen denken. Ook kan getest worden of een groep (bijvoorbeeld mensen veroordeeld wegens verraad of dienstname aan de Waffen SS) opvallende kenmerken vertoont.
- Bij het onderzoeken van kwantitatief-statistische wetenschappelijke onderzoeksvragen bleek dat bij analyse van de data een combinatie van kwalitatieve en kwantitatieve onderzoeksmethoden tot de meest bruikbare resultaten leidde. Er is nog ruimte voor verbetering van de digitale methodieken en reflectie op de aard van het materiaal, hier kan toekomstig onderzoek op voortborduren.

## Meer informatie over TRIADO:

- [TRIADO op Oorlogsbronnen.nl](#)
- [Videoserie 'De mensen van TRIADO'](#)
- ['Archieven en het semantisch web'](#). Artikel in het *Archievenblad*
- ['Enorme stap voorwaarts om archieven toegankelijk te maken'](#). Artikel in *IP | vakblad voor informatieprofessionals*

NETWERK  
OORLOGS  
BRONNEN

nationaal  
archief

huysens  
ing

NIOD  
instituut voor  
oorlogs-, holocaust-  
en genocidestudies

KNAW