

# Data van Dagboeken

## *Linguistics and Literature*

*Douwe A. Zeldenrust*

Meertens Instituut, Amsterdam, the Netherlands  
*douwe.zeldenrust@meertens.knaw.nl*

*Nina A. Wijsbek*

Meertens Instituut/Nederlands Dagboekarchief, Amsterdam,  
the Netherlands  
*nina.wijsbek@meertens.knaw.nl*

*Menzo A. Windhouwer*

Meertens Instituut, Amsterdam, the Netherlands  
*menzo.windhouwer@meertens.knaw.nl*

### Abstract

Since 2014, the Nederlands Dagboekarchief (Dutch Diary Archive) is a part of the Meertens Institute. In the DANS-financed project 'Data from Diaries' (2016), the paper metadata from the collection of ego documents have been made digitally accessible for researchers in a sustainable way. To this end, the Meertens Institute created a relational database in which the metadata can be stored. The database has been deposited at DANS. Additionally, CMDI-metadata have been created, which enable technical and content-level interchange between different European diary archives.

### Keywords

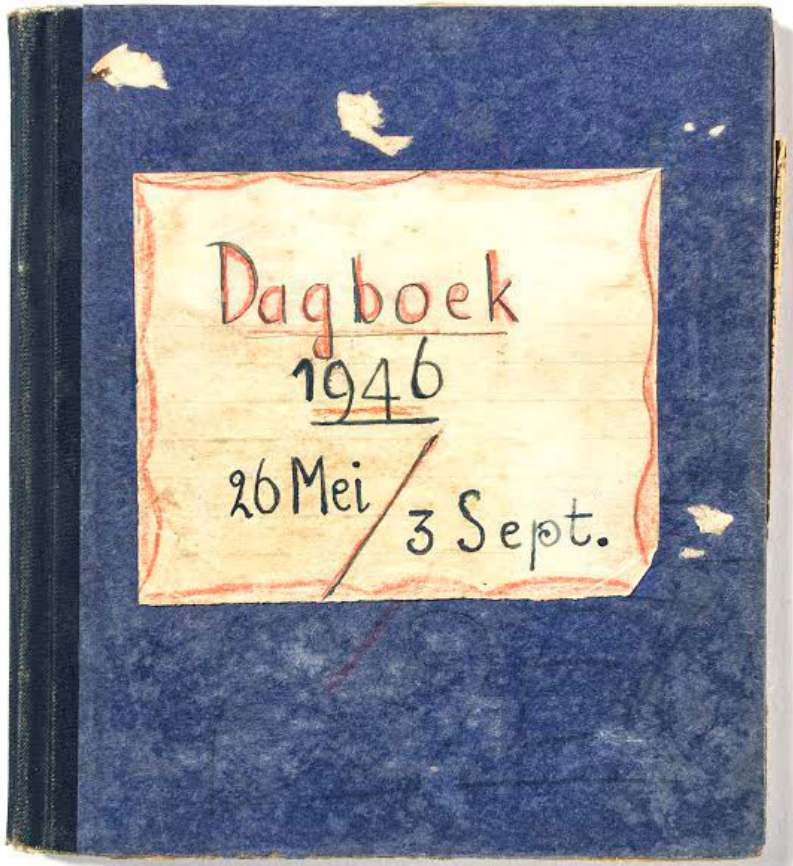
ego documents – Nederlands Dagboekarchief – database – CMDI – Meertens Instituut – diaries

- Gerelateerd aan de dataset “KDP Data van Dagboeken” met DOI <http://dx.doi.org/10.17026/dans-xcn-hy2q> gedeponeerd bij “DANS”.

## 1. Inleiding

1-4-'91

'Ik lig nu in het ziekenhuis terwijl ik dit schrijf. Ik heb namelijk een gebroken been. Nee hoor grapje 1 april. Hahahaha. We zijn vandaag (op 2e paasdag dus) naar Tineke + Jack gegaan. Hij had een tam parkietje en bijna had ik hem thuis, maar ons pa wilt hem niet. (Shit shit shit). 's Avonds hebben we nog Grieks gegeten, dat was heel lekker. Toen we thuiskwamen had Danny gebeld en ik moest hem terugbellen. Hij vroeg iets over Wilma en over hoe het weekend was. (Even uithoren dus.).'



AFBEELDING 1 *Dagboek uit de NDA-collectie.*  
FOTO: ADRIE MOUTHAN.

Dit fragment komt uit een dagboek van een 18-jarig meisje. Het dagboek is geschonken aan de Stichting Nederlands Dagboekarchief (NDA), die sinds 2014 onderdeel is van het Meertens Instituut (Koninklijke Nederlandse Akademie van Wetenschappen).

Dagboeken en andere egodocumenten geven een specifieke en persoonlijke kijk op de verschijnselen die het alledaagse leven in onze samenleving vormgeven. Voor taalkundigen, historici en etnologen vormen deze dagboeken een unieke bron. Vanuit taalkundig perspectief geven de dagboeken bijvoorbeeld de mogelijkheid om geschreven dialect te bestuderen. Voor historici geven de dagboeken de mogelijkheid om te onderzoeken hoe historische gebeurtenissen en tijdsvakken werden ervaren. De ongepubliceerde egodocumenten kunnen zo een waardevolle aanvulling zijn op de officiële geschiedschrijving. De onderzoeksmogelijkheden strekken zich echter ook uit naar onderzoek buiten de traditionele domeinen van taalkundigen, historici en etnologen. Ook voor bijvoorbeeld psychologen en sociologen zijn dagboeken een waardevolle bron. Zo wordt op dit moment door medewerkers van het NDA een publicatie voorbereid over puberdagboeken.

Zoals hierboven al staat komt het hier gebruikte fragment uit de collectie dagboeken van het Meertens Instituut. Het Meertens Instituut is een onderzoeksinstituut dat zich bezighoudt met de bestudering en documentatie van Nederlandse taal en cultuur (Bennis, 2012). In dat kader verzamelt het instituut dagboeken en werkt het samen met de Stichting Nederlands Dagboekarchief. Centraal in het onderzoek van het Meertens Instituut staan de verschijnselen die het alledaagse leven in onze samenleving vormgeven. Zo verschaft bovenstaand fragment informatie over schrijven in dialect ('ons pa'), wordt er een traditie genoemd (grappen maken op 1 april) en geeft het informatie over eetgewoontes (Grieks eten). In totaal bestaat het dagboekarchief uit meer dan 1500 dagboeken en de collectie groeit in aantal en omvang snel. In 2015 is een database ontwikkeld om de papieren registratie en de rijke metadatering te digitaliseren.

## 2. De data

In 2016 heeft DANS (Data Archiving and Networked Services) het Meertens Instituut een subsidie toegekend voor het project 'Data van Dagboeken' (Zeldenrust, 2016). Het doel van het project was om de metadata van de collectie op een duurzame manier voor onderzoek via het internet toegankelijk te maken. Tevens kunnen de metadata gebruikt worden voor onderzoek om er, los van de dagboeken zelf, analyses mee te maken. Zo kan de database door onderzoekers gebruikt worden om een onderzoekscorpus te selecteren,

bijvoorbeeld dagboeken waarin het thema ‘ouderschap’ naar voren komt, of dagboeken van 70-plussers. Dankzij de database is eenvoudig op te zoeken welke dagboeken in de collectie van het NDA aan deze criteria voldoen.

Dr. Sophie Elpers, onderzoeker aan het Meertens Instituut, zegt over de gebruiksmogelijkheden van de database voor onderzoek het volgende: ‘Een databank met metadata van een dergelijk groot dagboekarchief kan mij als etnoloog helpen om goede bronnen te vinden over het alledaagse leven in het verleden. Egodocumenten zijn in mijn vak bijzonder waardevol omdat zij vertellen over de belevingen van mensen, wat ze belangrijk vinden en wat niet, en waar ze welke betekenissen in zien. Maar een databank met de metadata kan ook als bron *an sich* waardevol zijn, in het bijzonder als er ook korte samenvattingen van de inhoud in zijn verwerkt. Dan kunnen we makkelijk te weten komen welke onderwerpen in welke tijd op welke plaats en door welke sociale groep bijzondere aandacht kregen. De kwalitatieve informatie die uit een enkel dagboek gegenereerd kan worden, kan daardoor aangevuld worden met kwantitatieve informatie. Kwesties van representativiteit (in kwantitatieve zin) kunnen daardoor makkelijker beantwoord worden.’

In totaal zijn de metadata van meer dan 1000 dagboeken en ca. 350 brieven in de database ingevoerd. De database is bij DANS gedeponeerd als CSV-files (Comma Separated Value). Om de data in het originele formaat te exploreren is Filemaker software nodig. Bij de CSV-data is documentatie aangeleverd in de vorm van een structuurdiagram om de database zo goed mogelijk te kunnen reconstrueren. De metadata is tevens omgezet naar CMDI (Component MetaData Infrastructure). CMDI is een metadatasysteem dat gebruik maakt van een gecontroleerd vocabulaire om metadata vorm te geven, te hergebruiken en om interoperabiliteit tussen verschillende metadataprofielen mogelijk te maken. CMDI is geïnitieerd door CLARIN (Common Language Resources and Technology Infrastructure), zie: <https://www.clarin.eu/content/component-metadata>.

De metadata van het NDA bevat privacy-gevoelige informatie in de vorm van bijzondere persoonsgegevens. Bijvoorbeeld religieuze informatie en gegevens over ziekte en gezondheid. Daarnaast zijn veel van de dagboekschrijvers niet overleden. Dat betekent dat de data volgens de Wet Bescherming Persoonsgegevens alleen ten behoeve van wetenschappelijk onderzoek gebruikt mag worden indien:

- (a) het onderzoek een algemeen belang dient,
- (b) de verwerking voor het betreffende onderzoek of de betreffende statistiek noodzakelijk is,

- (c) het vragen van uitdrukkelijke toestemming onmogelijk blijkt of een onevenredige inspanning kost en
- (d) bij de uitvoering is voorzien in zodanige waarborgen dat de persoonlijke levenssfeer van de betrokkene niet onevenredig wordt geschaad.

De database (en de afgeleide csv bestanden) zijn bij DANS vanwege de privacy-gevoelige informatie onder ‘restricted access’ beschikbaar. Het Meertens Instituut zal aanvragen voor het gebruik van de database of de csv bestanden aan de vier bovengenoemde voorwaarden toetsen. De CMDI-metadata die is gegenereerd is geanonimiseerd. Die data zijn daarom wel open access bij DANS beschikbaar en worden tevens aangeboden via het OAI-PMH protocol (Open Archives Initiative Protocol for Metadata Harvesting). De gegevens zijn zo ook beschikbaar gemaakt via de Europese onderzoeksinfrastructuur CLARIN (Common Language Resources and Technology Infrastructure) zodat ze vergeleken kunnen worden met andere, bijvoorbeeld taalkundige, bronnen binnen Europa.

### 3. De database

De database is, zoals hierboven al staat, van origine een relationele FileMaker Pro database. De database bestaat uit diverse tabellen die met elkaar verbonden zijn. De eerste tabel is genaamd het ‘registratieformulier’; hierin wordt metadata op collectieniveau genoteerd. De voornaamste metadata-velden van het registratieformulier betreffen inventarisnummer, naam auteur(s), type document (memoire, dagboek, brief, etc.), aantal delen en de vorm van de ingezonden tekst (bijvoorbeeld ‘handgeschreven origineel’). Ook is er een tekstveld waarin een korte beschrijving van de collectie wordt genoteerd. De tweede tabel bevat informatie over de auteur (naam, geboortedatum, opleiding, beroep, religie, burgerlijke staat, bijzonderheden, overige biografische gegevens). De derde tabel bevat informatie over de schenker van de collectie (naam en adresgegevens). De vierde tabel bevat informatie over de vrijwilliger die de collectie heeft gelezen (naam en adresgegevens). De vijfde tabel bevat informatie over de vrijwilliger die de database heeft ingevuld. De auteurs, als die nog in leven zijn en niet de schenker zijn (tabel 2), de schenkers (tabel 3) en de vrijwilligers (tabel 4 en 5) geven toestemming voor het verwerken van hun persoonsgegevens.

Voor onderzoekers die inhoudelijk onderzoek naar egodocumenten doen is de zesde tabel, de ‘Leesrapportage’ van belang. In deze tabel worden collecties inhoudelijk ontsloten door middel van trefwoorden. Deze metadata

worden gegenereerd door vrijwilligers van het NDA, die de dagboeken lezen. De belangrijkste metadatavelden van het onderdeel leesrapportage betreffen periode en plaats van schrijven, motief om te schrijven, en een tekstveld waarin een samenvatting van de inhoud wordt genoteerd. Daarnaast kan hier een groot aantal trefwoorden worden aangeklikt of door de lezer zelf worden toegevoegd. Deze zijn op thema gesorteerd zoals maatschappelijk cultureel (politiek, school, etc.), persoonlijk (vriendschap, seksualiteit, etc.) etc. De laatste, zevende tabel, zorgt voor de koppeling van de relationele tabellen.

#### 4. Context

De collectie dagboeken staat niet op zichzelf. In 2015, op een bijeenkomst van het Meertens Instituut met de vertegenwoordigers van andere Europese dagboekarchieven (onder andere uit Duitsland, Frankrijk, Engeland en Italië), is een Europese koepel van dagboekarchieven opgericht. De koepel is genaamd EDAC (European Diary Archives and Collections). Gerenommeerde onderzoekers op het gebied van (auto)biografieën zoals de Fransman Philippe Lejeune hebben zich daarbij aangesloten. Het secretariaat ligt in handen van het NDA en wordt gevoerd vanuit het Meertens Instituut. Zoals hierboven staat beschreven zijn de metadata van het NDA ook beschikbaar in CMDI formaat. Dat laatste geeft een springplank naar mogelijke technische en inhoudelijke samenwerking tussen de verschillende Europese dagboekarchieven, al dan niet in het kader van CLARIN. Daarnaast beheert het Meertens Instituut ook nog andere egodocumenten zoals de collectie 'Brieven aan de Toekomst'. In die brieven hebben meer dan 50.000 Nederlanders een dag uit hun leven (15 mei 1998) beschreven en gestuurd "aan de toekomst".

#### 5. Tot slot

Zoals hierboven al staat beschreven groeit de collectie gestaag. De papieren registratie van de dagboeken heeft plaatsgemaakt voor directe registratie in de FileMaker Pro database. Nieuwe versies van de database zullen regelmatig bij DANS worden gedeponereerd.

#### 6. Data

- Data gedeponereerd bij DANS – DOI:<http://dx.doi.org/10.17026/dans-xcn-hy2q>

NederlandsDagboekArchief (FM21)

Records: 1 Total (Unsorted)

Layout: leesrapport...\_schrijving View All New Record Delete Record Find Sort

perioode van schrijven: 2009

land van schrijven: Nederland

plaats van schrijven: Den Haag

taal: Nederlands

staat materiaal: ☒ goed ☐ matig ☐ slecht

conservatie advies:

motief om te schrijven: 5. heeft behoefte aan overzicht over woelig

manier van schrijven: ☒ microniveau (heel persoonlijk) ☒ mesoniveau (afstandelijk) ☐ macroniveau (van buitenaf beschouwd)

levensfase: ☒ 0-10: kindertijd ☒ 11-20: tienerjaren ☒ 21-30: jong volwassenen ☒ 31-40: dertigers ☒ 41-50: midlife ☒ 51-60: middelbare leeftijd ☒ 61-70: vroege ouderdom ☐ 71+: ouderdom ☐ Other...

0-10: kindertijd  
11-20: tienerjaren

leventhemas: ☒ dood ☐ ouderschap ☐ relaties ☐ familie ☒ religie ☒ Other... ☐ handicap ☐ rouw verlies ☒ huwelijk ☐ scheiding ☐ ongeval ☐ verlieftheid ☐ ouders ☐ ziekte

huwelijk  
familie

persoonlijk: ☐ dromen ☒ seksualiteit ☐ generatieconflict ☐ spiritualiteit ☐ geweld misbruik ☐ vakanties reizen ☐ hobby interesses ☒ vriendschap

100% Browse

AFBEELDING 2 Een ingevulde tabel 'leesrapportage' in de NDA-database.

## Literatuur

- Bennis, H. J. (2012). *Crossing Boundaries – Research plan Meertens Institute 2013–2018*. Amsterdam, Meertens Institute. <http://www.persistent-identifier.nl/?identifier=urn:nbn:nl:ui:17-826bb48a-2abc-4480-b50f-e80e1c19b5f3>.
- Zeldenrust, D. A. (2016). Data van Dagboeken. *Nieuwsbrief van het Meertens Instituut*, 6, Februari 2016 (Amsterdam). <http://hdl.handle.net/20.500.11755/b790e01b-g984-44f6-8a6a-7f09fa6c2efd>.

## Website

<https://www.clarin.eu/cmdl> (geraadpleegd op 21 november 2016).