

# Stand van zaken DCS eind juni 2012

---

## DBD

Zie voortgangsverslag, opgenomen in curatieplan op Google Docs: DBD curation plan

(<https://docs.google.com/document/d/1RO7ySmyFlpEoNMJdrZ0Q-GoaubcbTdxLiOX4VMpmkM/edit#>)

## Juridisch Corpus

Dit corpus bestaat uit 4 onderdelen:

- (1) 31 geluidsopnames van strafzittingen (meervoudige kamer, in 3 arrondissementen), begin jaren '90.
- (2) 20 geluidsopnames van politieverhoren (bureaus Nieuwmarkt en Warmoesstraat, Amsterdam), rond de eeuwwisseling.
- (3) 14 geluidsopnames van politieverhoren (roofteam in de Bijlmer), 2007-2009.
- (4) 5 video-opnames van zittingen van dezelfde zaken (jeugdzaken), 2007-2008.

Alle opnames zijn digitaal opgeslagen.

(1) en (2) zijn bij Martha Komter van het Nederlands Studiecentrum Criminaliteit en Rechtshandhaving (NSCR). De transcripten zijn aanwezig en geanonimiseerd. Ook heeft Martha ons van allerlei metadata voorzien (vnl. over de zittingen en verhoren, niet over de personen, die zijn er ook niet).

Wilbert Spooren vertelde ons het volgende. Alleen voor het materiaal in (4) (video-opnamen van 5 rechtszittingen) is toestemming verkregen om het nog niet te vernietigen. Dat betekent zeker niet dat het aan anderen beschikbaar gesteld mag worden. Voor (1)-(3) is zelfs geen toestemming gegeven om het nog niet te vernietigen. De transcripties van (3) en (4) zijn niet geanonimiseerd.

Wilbert had van Arjan van Hessen begrepen dat de DCS ook materiaal kan cureren om het voor de toekomst veilig te stellen zodat het toegankelijk kan worden gemaakt zodra die toestemming wel is verkregen. Ook bij DANS liggen zulke collecties, puur om te zorgen dat het bewaard blijft (archivering dus).

We begrijpen nu waarom we een voorbeeld van (4) hebben gekregen. Dat is namelijk de enige categorie waarvan is gezegd dat het nog niet vernietigd hoeft te worden. Dit materiaal is echter niet geanonimiseerd (i.t.t. (1) en (2)).

In conclusie:

- IPR van het materiaal is niet geregeld, het mag aan niemand ter beschikking worden gesteld. Wellicht dat dit later nog eens gebeurt, maar daar is nu echt niets van te zeggen
- Moet de DCS er desondanks toch (veel) werk insteken om het te cureren: oplijning, anonimiseren van tekst en audio, metadata, PIDs ?
- Of is het beter het materiaal (bijv. bij DANS) te archiveren "as is", eventueel met CMDI metadatafiles in afwachting van betere tijden?

Ons advies zou het laatste zijn.

## **IPNV**

DANS heeft besloten CLARIN-B-center te willen worden en daarmee voor een CLARIN-conforme PID-afhandeling zorg te dragen. DANS is tevens bereid om bij te dragen aan de curatie van de ongeveer 750 resterende interviews. De audio daarvan is al aanwezig en het meeste zit ook in Easy. Zaak is nu PIDs te genereren en CMDI metadata files en die harvestable te maken. In juli maken we nadere afspraken. In september begint het werk en in oktober is de curatie door de DCS naar verwachting klaar.

## **Gelders Dialectwoordenboek**

Charlotte Glesbers die werkt aan de Gelderse dialect woordenboeken heeft contact opgenomen met de DCS om de mogelijkheden voor curatie ervan te verkennen.

Voor Gelderland bestaan er 3 dialectwoordenboeken: 1 voor het Rivierengebied, 1 voor de Veluwe en 1 voor de Achterhoek en Liemers.

-De data voor het Rivierengebied is zo goed als klaar. Charlotte denkt hier nog een maand de tijd voor nodig te hebben. In totaal zijn het 190.000 records.

-De data voor de Veluwe is digitaal, maar moet nog eens goed nagelopen worden. Hiervoor heeft Charlotte nog een maand of 6 nodig.

-De data voor de Achterhoek en Liemers is nog niet digitaal en zou eerst nog ingevoerd moeten worden.

De data zijn de resultaten van de vragenlijsten, de data bestaan uit:

- lemma
- dialectwoord
- betekenis
- plaats
- kloekecode
- beschrijvingen

In overleg met Roeland van Hout heeft Charlotte besloten allereerst het Rivierengebied te digitaliseren om vervolgens ook de Veluwe eraan te hangen. De Achterhoek en Liemers kost meer tijd en geld. Daarvoor moeten we kijken wat de mogelijkheden zijn.

### **Leiden, Marjan Klamer**

Marjan Klamer heeft een voorstel voor een curatieproject ingediend in Call 3 (closed call). Naast het materiaal wat in dit voorstel al wordt gecureerd, ligt er in Leiden nog veel meer materiaal dat voor curatie in aanmerking komt. Er is contact geweest met Marjan Klamer. Zij wil graag in samenspraak met de DCS bezien welk materiaal door de DCS gecureerd zou kunnen worden. Aangezien Marjan momenteel in Indonesië verblijft, praten we hierover na de zomer.

### **LESSLA**

Met Ineke van de Craats is gesproken over de curatie van het LESSLA corpus. De laatste ideeën hierover zijn dat de data in ieder geval veilig gesteld worden (door overdracht aan de DCS) en dat er enkel een inspanning geleverd wordt om het corpus middels de metadata te ontsluiten

### **Verzoek Nicoline van der Sijs:**

Het omzetten van ca. 480.000 microfiches van de vragenlijsten van het Meertens Instituut naar jpg/pdf. Sinds 1931 heeft het Meertens Instituut jaarlijks vragenlijsten uit doen gaan naar informanten in Nederland, Friesland en België met dialectvragen, vanaf 1934 ook met volkskundevragen; daarnaast zijn er een tiental naamkundevragenlijsten. Vanaf 2005 worden de vragenlijsten op de Meertens website digitaal beantwoord, maar tot die tijd kwamen de vragenlijsten met handgeschreven antwoorden per post retour.

De eerste vraag is of het omzetten van de 480.000 microfiches naar jpg/pdf kan vallen onder de DCS: er zijn al veel kosten en moeite gedaan om de data van papier om te zetten naar een digitaal medium, maar inmiddels voldoet het medium microfiche niet meer aan de eisen die we momenteel aan data stellen. De tweede vraag is of het scannen van de papieren antwoorden wellicht ook onder de DCS kan vallen. (65 vragenlijsten = 290.000 pagina's)

Indien akkoord voor EB, dan kan de klus wellicht aan BREED worden uitbesteed (in januari 2013).

### **Aanbeveling voor Sponsoring door CLARIN-NL:**

Het International Digital Curation Conference (IDCC). Zie <http://www.dcc.ac.uk/events/idcc13/become-sponsor>.

DCS wil hier present zijn met poster en paper.