

In het artikel *Het Digitale Drama* (NRC 10 sep. 2011) van Karel Berkhout wijzen een aantal onderzoekers erop dat de massale digitalisatie van geschreven bronnen in de geesteswetenschappen grote mogelijkheden biedt om nieuwe onderzoeksvragen voor het eerst aan te pakken, bestaande vragen veel efficiënter te beantwoorden, en de kwaliteit van het onderzoek op te schroeven (door verifieerbaarheid, repliceerbaarheid van de onderzoeksresultaten, etc.). Zij wijzen er echter terecht ook op dat de “humaniora vanouds versplinterd zijn”, dat er geen duidelijke standaard is voor digitalisatie, dat er nog te veel fouten in de ge-OCR-de data zitten, en dat daardoor het potentieel dat de digitalisatie biedt onderbenut wordt. Data zijn verspreid over vele websites, en vaak moeilijk of helemaal niet vindbaar. Ik onderschrijf deze beweringen maar zou er ook op willen wijzen dat er wel degelijk acties gaande zijn om verandering in deze situatie te brengen. CLARIN-NL (www.clarin.nl) bouwt sinds 2009 aan een digitale infrastructuur voor de humaniora waar via een webportaal alle data (en software om met deze data om te gaan) zichtbaar, vindbaar en voor alle onderzoekers bruikbaar zullen zijn; een eerste prototype van een browser kan men vinden op www.clarin.eu/vlo. Geavanceerde software om in de data en metadata te zoeken wordt ontwikkeld en zal gedemonstreerd worden op de *European Innovation Conference* (www.eic2011.com/).

Hoewel er altijd fouten zullen worden gemaakt bij het OCR-en van (vooral) oudere teksten, worden er steeds vaker technieken gebruikt die het aantal fouten sterk reduceren. Daarnaast worden de gedigitaliseerde teksten voorzien van hoge kwaliteit metadata (data die op eenduidige wijze de gedigitaliseerde data beschrijven) in overeenstemming met door CLARIN ontwikkelde standaarden. Dit alles om huidige en toekomstige gedigitaliseerde teksten beter te kunnen ontsluiten.

Alle in het artikel genoemde instellingen doen actief mee in CLARIN-NL, en CLARIN-NL zelf kadert in een Europese inspanning om een dergelijke digitale infrastructuur voor de humaniora op te zetten (het heeft immers geen zin om zoiets op nationaal niveau te organiseren; daar is de wetenschap veel te internationaal voor). Bibliotheken (KB, universiteitsbibliotheken, DBNL) en datacentra (bijv. Beeld & Geluid) ontwikkelen nu plannen om ook hun data via de CLARIN infrastructuur zichtbaar, vindbaar en doorzoekbaar te maken. CLARIN-NL wordt gefinancierd door NWO, en wordt gecoördineerd door de Universiteit van Utrecht, dat ook een leidende rol speelt in CLARIN op Europees niveau (samen met MPI Nijmegen).

Ook ten aanzien van de traditionele versplintering van de humaniora zijn er belangwekkende ontwikkelingen: zo is onlangs een infrastructuurproject aangevraagd door een consortium van *alle* geesteswetenschappelijke onderzoekers in Nederland om een gemeenschappelijk virtueel digitaal laboratorium te ontwikkelen voor geesteswetenschappelijk onderzoek. Dit toont duidelijk aan dat de geesteswetenschappers de traditionele fragmentatie van hun veld achter zich willen laten, nauw samen willen werken en optimaal gebruik willen maken van de nieuwe mogelijkheden die de digitale data en software bieden.

Prof.dr. Jan Odijk, Directeur CLARIN-NL, Universiteit Utrecht