

Prof. dr. Lex Heerma van Voss, directeur Huygens ING en hoofdaanvrager CLARIAH:

‘CLARIAH biedt structuur aan de digitale onderzoeksrevolutie’

Vanaf 2015 werkt CLARIAH aan de ontwikkeling van een digitale infrastructuur voor geesteswetenschappers. De voorbereidingen daarvoor zijn momenteel in volle gang.

Erica Renckens

“Vanaf 1 januari mogen we geld uitgeven, maar voor het zover is, moet er nog heel wat overlegd worden over hoe we dat precies gaan doen,” aldus prof. dr. Lex Heerma van Voss, die als directeur van het Huygens ING hoofdaanvrager is van CLARIAH. Begin deze zomer kreeg CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities) een NWO-subsidie van twaalf miljoen euro voor het ontwikkelen van een digitale infrastructuur waarin geesteswetenschappers grote databestanden kunnen interpreteren en ontsluiten.

Voor alles van nut

Het is de tweede keer dat het consortium van geesteswetenschappelijke onderzoeksinstellingen de CLARIAH-aanvraag indiende. “De eerste keer kregen we als commentaar dat onze inzet op de gehele geesteswetenschappen te breed was,” vertelt Heerma van Voss. “Daarom hebben we nu de focus gelegd op drie deelgebieden: taalkunde met overwegend tekstuele data, mediastudies met audiovisuele bronnen en sociaal-economische geschiedenis met gestructureerde databestanden, databases.” Alle drie de vakgebieden zijn al voorloper in de Digital Humanities – een weloverwogen keuze. “Deze typen bronnen worden ook door andere geesteswetenschappelijke disciplines gebruikt. We hebben de verantwoordelijkheid voor de hele breedte van de geesteswetenschappen, dus wat we gaan maken, moet in principe voor alle disciplines van nut zijn.”

“Taalkundigen kunnen bijvoorbeeld al heel goed teksten automatisch parsen,” geeft Heerma van Voss als voorbeeld van een toepassing die breed inzetbaar is. “Nu willen historici of letterkundigen meestal niet speciaal weten of iets een lidwoord of een werkwoord is, maar zo’n techniek is wel een middel om grip te krijgen op een tekst. Je ziet steeds meer onderzoeken waar een parser wordt gebruikt om historische of letterkundige vragen te kunnen beantwoorden.”

Lab voor geesteswetenschap

“Net als in CLARIN zullen we bij CLARIAH calls uitschrijven waarin onderzoekers aanvragen kunnen indienen. We willen weten waar de geesteswetenschappers behoefte aan

hebben, dus we vragen mensen om zelf met voorstellen te komen voor nieuwe datasets of nieuwe tools om databestanden mee te ontsluiten.” Toch zullen er ook verschillen zijn met CLARIN. “In CLARIN zijn voornamelijk losse demonstrators gemaakt,” constateert Heerma van Voss. “Ik denk dat wij meer opschuiven naar een labsituatie, een volledige productieomgeving voor onderzoekers.”

In de naam CLARIAH is niet alleen ‘CLARIN’, maar ook ‘DARIAH’ te herkennen, beide grote Europese onderzoeksinfrastructuren – het eerste voor taalkundigen, het andere voor historici. “Nederland heeft een voortrekkersrol in CLARIN. In DARIAH minder, al is Nederland ook daar actief in. CLARIAH zal zowel in CLARIN als DARIAH de Nederlandse bijdrage leveren.” Daarnaast verbindt CLARIAH ook Nederlandse deelprojecten, als CLIO-Infra, Nederlab en EU Screen.

INTERVIEW

‘Oké, dit is zo interessant, dit ga ik ook gebruiken’

Eisen aan metadata

Goede afspraken omtrent metadata zijn voor zo’n brede samenwerking onontbeerlijk. “Daar zullen we het dit najaar – en ongetwijfeld gedurende het hele project – veel over moeten hebben,” beaamt Heerma van Voss. “Het is best ingewikkeld om tussen al die geesteswetenschappelijke disciplines te communiceren. CLARIN heeft eisen gesteld aan hoe de metadata eruit moeten zien. We zijn er nog niet over uit of het zin heeft die regels geesteswetenschappelijk breed voor te schrijven. In ieder geval streven we de interoperabiliteit ook internationaal na.”

“De demonstrators van CLARIN toonden aan waar onderzoekers in het veld behoefte aan hadden; daar werden standaarden voor ontwikkeld. Die standaarden zijn nog altijd volop in ontwikkeling. Deze revolutie zal ook doorzetten als CLARIAH niet gefinancierd zou zijn, maar nu hebben we in Nederland de mogelijkheid hier goed mee om te gaan en daar een structuur voor aan te bieden.”

Een grote vlucht

“Ik denk dat over vier jaar elke student een cursus Digital Humanities volgt,” voorspelt Heerma van Voss. “Niet omdat wij dat willen

Moral panic over prostitutie

Ook voor zijn eigen onderzoek ziet Prof. dr. Lex Heerma van Voss de mogelijkheden van de CLARIAH-infrastructuur. Hij onderzoekt arbeidsomstandigheden wereldwijd. “Ik heb bijvoorbeeld een artikel geschreven over prostitutie. Rond 1900 en rond 2000 zie je een vergelijkbare omslag in het denken over prostitutie: van gereguleerd en geaccepteerd naar verbieden vanwege dwang en slavernij. Inmiddels weten we uit literatuurstudies dat de moral panic van een eeuw geleden overdreven was, maar daaruit kun je natuurlijk niet concluderen dat dat nu weer zo is. Er blijken veel variabelen een rol te spelen in dit verhaal, meer dan ik als mens kan overzien. Met de data die we nu in CLARIAH verzamelen, heb ik een veel betere set in handen waarmee ik variabelen als ‘mate van migratie’ en ‘vrijheid van de vrouw op plek van herkomst’ beter in de vingers kan krijgen.”



Prof. dr. Lex Heerma van Voss foto Bart van Vliet

of geven, maar omdat steeds meer mensen er de mogelijkheden van inzien. We hebben verschillende onderzoeksscholen onze aanvraag laten lezen en gevraagd hoeveel van hun promovendi gebruik zouden maken van een dergelijke infrastructuur. Momenteel zou dat zo’n veertig procent zijn, maar ze gaven allemaal aan dat dit op korte termijn een nog veel grotere vlucht zal nemen.”

Bij het onderzoeksinstituut van Heerma van Voss zelf, het Huygens ING, is de geesteswetenschappelijke revolutie al voelbaar. “Wij

houden ons veel bezig met het maken van tekstedities. We wisten precies hoe dat moest; daar waren gewoon regels voor. Maar de digitale revolutie maakt dat alle basale handelingen voor geesteswetenschappers in beweging zijn: welke informatie is beschikbaar, hoe vind ik die, hoe bevestig ik de data, hoe laat ik de resultaten zien, hoe zorg ik dat alles teruggevonden kan worden. Ook al heb je een briljant idee, je kunt niet zeggen ‘Voortaan gaan we het zó doen.’ Alle relevante spelers werken mee aan CLARIAH, zodat we precies weten waar onderzoekers behoefte aan hebben.”

CLARIAH heeft voor vier jaar subsidie gekregen. Wanneer kan gesproken worden over een geslaagd project? “We hebben natuurlijk beloofd een infrastructuur te ontwikkelen die in de drie gebieden relevant is,” aldus Heerma van Voss. “Maar ik zou het echt een succes vinden als we iets maken waardoor ook andere geesteswetenschappelijke disciplines zeggen: ‘Oké, dit is zo interessant, dit ga ik ook gebruiken.’”