

ePistolarium maakt zeventiende-eeuwse correspondentie doorzoekbaar

Online grasduinen in brieven van geleerden

Binnenkort vindt de lancering plaats van het ePistolarium, een online onderzoeksomgeving rond ruim twintigduizend brieven van zeventiende-eeuwse geleerden uit negen verschillende verzamelingen. *Erica Renckens*

“Het begint eigenlijk pas net leuk te worden!” Charles van den Heuvel is vanuit het Huygens ING betrokken bij het onlangs afgeronde Geleerdenbrievenproject, waar de afgelopen vijf jaar ook het Descartes Centre, de Koninklijke Bibliotheek, DANS en de universiteiten van Amsterdam en Utrecht aan werkten. Gedurende die tijd is het karakter van het project iets bijgesteld. “Oorspronkelijk was het de bedoeling om een *collaboratory* te bouwen”, aldus Van den Heuvel. In een virtuele omgeving zouden wetenschaps-historici kunnen samenwerken en het materiaal verrijken. “Maar het bleek geen goed idee om het systeem direct open te stellen; de initiële interface en resultaten van het automatisch herkennen van *topics* in meertalige brieven waren nog niet goed genoeg.”

350 spellingsvarianten

De mogelijkheden van het ePistolarium zijn daarom uitgebreid met verschillende taaltechnologische methoden. “De meest geavanceerde is waarschijnlijk de taalafhankelijke *topic modeling*”, vertelt Walter Ravenek, die als ontwikkelaar vele uren in het project heeft gestoken. Op basis van het gezamenlijk voorkomen van woorden bepaalt deze methode de onderwerpen die in de brief behandeld worden. “Daarmee



De correspondentie van Constantijn Huygens, ruim 7.000 brieven, is één van de negen verzamelingen die momenteel zijn opgenomen in het ePistolarium credits The National Gallery, London

Antwoorden op veel zorgvragen

Gegevens over volksgezondheid en zorg zijn eenvoudig online te vinden. Maar door het enorme aanbod is het lastig om de betrouwbaarheid van deze gegevens in te schatten. Zorggegevens.nl geeft antwoorden. *Rutger Nugteren*

“Zorggegevens.nl is een online catalogus met informatie over gegevensbronnen die betrekking hebben op volksgezondheid en zorg.” Aan het woord is Laurens Zwakhals, projectleider Zorggegevens bij het RIVM. “Het gaat om beschrijvende informatie, de metagegevens, van gegevensbronnen zoals zorgregistraties, enquêtes, monitors en langlopende (cohort) onderzoeken. De catalogus biedt zoveel mogelijk in-

formatie zodat het vinden van de gegevens zo eenvoudig mogelijk wordt. De gegevens zelf kunnen onderzoekers vervolgens bij de bronhouders opvragen.”

Handige dossiers

Zwakhals vervolgt: “Onderzoekers vinden op de site informatie over bijvoorbeeld organisaties op het terrein van zorggegevens. Met korte teksten en links naar websites wordt een beeld geschetst wie welke gegevens verzameld en waarom. Handig als je wilt weten voor welke informatie je bij wie kan aankloppen. De site beschikt ook over handige dossiers zoals ‘Ouderen en gezondheid’ of ‘Infectieziektesurveillance’.

Gevraagd naar het belangrijkste onderdeel van de website antwoordt

Zwakhals: “Dat is de catalogus met beschrijvende informatie, de meta-database. Hierin staat algemene informatie zoals de naam van de gegevensbron. Je vindt er ook een beschrijving van de data, de gebruiksdoelen en de beschikbaarheid van de gegevens. Zo kun je bijvoorbeeld opzoeken of de bron al of niet online beschikbaar is en zo niet, met wie je dan contact kunt opnemen. Daarnaast is informatie over de inhoud van de bron beschikbaar zoals de geografische dekking, de steekproefgrootte en het responspercentage.”

Routekaart

“Ten slotte is er ook operationele informatie beschikbaar. Daarbij moet je denken aan het eerste en meest recente meetjaar, de op-

drachtgever en de belangrijkste financier van de dataverzameling.” Sinds begin dit jaar werkt Zorggegevens.nl samen met anderen (CentERdata, DANS, BBMRI en NIVEL) aan een datamanagement routekaart voor ZonMw. Deze routekaart is bedoeld om onderzoekers te helpen met een subsidieaanvraag bij ZonMw. “Voor de routekaart onderzoeken we bijvoorbeeld hoe bestaande databronnen beter vindbaar gemaakt kunnen worden en hoe de (duurzame) opslag van data gefaciliteerd kan worden. Ook beoordelen we het financieringsmodel voor (duurzame) opslag van data, bijvoorbeeld hoeveel ruimte (tijd en geld) projecten moeten reserveren voor adequaat datamanagement.”

zorggegevens.nl



E-DATA & RESEARCH

Jaargang 7 | nummer 3

Nieuwsbrief over data en onderzoek in de alfa- en gamma-wetenschappen.

E-data & Research verschijnt drie keer per jaar en wordt mogelijk gemaakt door: CBS, CentERdata, CLARIN-NL, DANS, Huygens ING, de Koninklijke Bibliotheek en het RIVM.

INHOUD

2 **Gehoord en bijgewoond: events besproken**

3 **80.000 KB-boeken online in Google Books**

3 **Volkstellingen aan Linked Data onderworpen**

4 **Deze datasets zijn sinds kort beschikbaar**

4 **Agenda: interessante bijeenkomsten op een rij**

5



Interview directeur Center for Digital Humanities

6 **Utrecht Data School visualiseert data**

7 **3D-illustraties in Dynamic Drawings-project**

8 **Schatgraven in de statistieken van het CBS**

8 **Ewoud Sanders zoekt een digitale betaalkrant**



Scan deze QR-code met een smartphone om de website van E-data te bezoeken.
www.edata.nl

VERVOLG OP PAGINA 2