

Pagina 3 • eHumanities • De nieuwe eHumanities Group van de KNAW heeft een eigen plek in Amsterdam. Nu kan het beginnen.

Pagina 4 • Webweegschaal • Duizend internetweegschalen meten gewicht en vetpercentage van deelnemers aan het LISS panel. Door deze *state-of-the-art technology* komen meer en preciezer gegevens beschikbaar.



WILLIAM HOOGTEYLING

Pagina 5 • Pieter Hooimeijer • Sociaal-geograaf Pieter Hooimeijer was altijd bezig met grote dataverzamelingen. Nu pleit hij voor een systeemaanpak van maatschappelijke problemen. En voor het toegankelijker maken van registerdata.

Pagina 6 • CLARIN-NL • In de rubriek Focus is de aandacht gericht op een project om bestaande bestanden in de geesteswetenschappen, en gereedschappen om ze te gebruiken, goed zichtbaar en toegankelijk te maken.

Pagina 8 • Darnton • Harvardhoogleraar en directeur van de universiteitsbibliotheek aldaar Robert Darnton sprak in Antwerpen over boeken en e-books. *e-data* was erbij.

EN VERDER

Agenda	2
Nieuws	3
Achtergrond	4,6
Focus	6
Sinds kort beschikbaar	7
Column	8
Gelezen	8

Geld voor e-projecten uit NWO Middelgroot

Verschillende onderzoeksvorstellen met een sterke e-dimensie zijn in de prijzen gevallen bij de subsidieronde van NWO die in maart plaatsvond onder de titel Investerings NWO-middelgroot. Die ronde, speciaal bedoeld voor infrastructurele voorzieningen, ging in totaal over 12,2 miljoen euro.

Zo werd een aanvraag gehonoreerd van een groep onder leiding van Christine Arnold (zie ook p.7) voor een databank over politieke responsiviteit. De bedoeling is om een geïntegreerde en openbare databank te maken met gegevens over voorkeuren van Europese electoraten, politieke partijen en regeringen, verkiezingsuitslagen en wetgeving. Met die databank moet onderzoek worden gedaan naar de mate waarin het publiek reageert op overheidsbeleid en de mate waarin de overheid zijn politiek uitdraagt door de tijd en

KNAW: Open Access ook voor data

De Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) kiest voor Open Access en vraagt van haar instituten om dat ook te doen. Dat heeft het bestuur van de Akademie dit voorjaar besloten. In een brief aan de instituten wordt gevraagd 'het beleid rondom publicaties zodanig vorm te geven dat het past bij de doelstellingen' van de KNAW, en dit nog dit jaar vast te leggen in een notitie. Daarbij kunnen ze gebruik maken van een door de moederorganisatie aangereikt 'handvat'. Om de digitale duurzaamheid door de hele organisatie heen van de grond te krijgen is een budget gereserveerd van acht ton over een periode van vier jaar te beginnen met 2011.

Opvallend is dat de KNAW expliciet spreekt over open toegang tot publicaties en data. Vaak wordt dat onderscheid in discussies over open access niet gemaakt en blijken de gedachten van discussianten en beleidsmakers vooral uit te gaan naar publicaties zoals tijdschriftartikelen.

Voor publicaties 'volgt de KNAW het voorstel dat de (...) Nederlandse universiteiten vorig jaar hebben aangenomen', aldus de beleidsnotitie waarop het bestuursbesluit is gebaseerd. Voor onderzoeksdata is het beleid dat data (...) duurzaam opgeslagen en voor iedereen toegankelijk worden, tenzij er bijvoorbeeld wettelijke bepalingen (privacy, nationale veiligheid), contracten of andere dwingende argumenten zijn die dat

belemmeren.' Uitzonderingen dienen te worden verantwoord en elk onderzoekplan moet een korte dataparagraaf bevatten die aangeeft 'hoe het hergebruik van de onderzoekdata bevorderd kan worden'.

Het budget voor Open Access wordt over vier gelijke porties per jaar verdeeld, waarbij steeds 25 duizend euro voor scholing en voorlichting wordt begroot en een zelfde bedrag voor het uitbreiden van

de reeds bestaande repository voor publicaties van de KNAW. Stimuleringsfondsen voor open access publicaties en voor permanente toegankelijkheid van data krijgen ieder steeds 75 duizend euro. (MdG)

Sonttolregisters te raadplegen



WIKIPEDIA

Willem van de Velde (1611-1693): Slag op het Sont, 8 nov 1658; v.l.n.r. de Carolus, de Eendraght, de Mercurius, de Rotterdam, de Pelikan, de Morgen-stjarna, de Halve Maen, de Krona en de Brederode (detail)

Dit voorjaar is het eerste deel van de database Sonttolregisters-online online beschikbaar gekomen, met de gegevens over meer dan 125 duizend doorvaarten uit de laatste dertien jaar van de achttiende eeuw. De gegevens zijn te raadplegen op www.soundtoll.nl.

De koning van Denemarken hief

van ca. 1400 tot 1857 tol op schepen die door de Sont voeren, de zeestraat tussen de Oostzee en de Noordzee. De registers zijn bewaard vanaf 1497. De helft van de schepen in het register kwam uit Nederland. In de zeventiende eeuw ging het vooral om Hollandse schepen, in de loop van de achttiende eeuw namen schepen uit

Friesland die rol over. De tolboeken bevatten gegevens over 1,8 miljoen doorvaarten en vormen daardoor een belangrijke bron voor onderzoek.

De data worden in een database ingevoerd op initiatief van de Universiteit van Groningen en het Fries historisch en letterkundig centrum Tresoar. Het project duurt tot 2013.

Ook tweets in SoNaR-corpus

Ruim vierhonderd Nederlandse twitteraars hebben tot nu toe gegeven aan een oproep van taalkundigen om hun tweets beschikbaar te stellen aan de wetenschap. Die oproep werd afgelopen februari gedaan voor SoNaR, een groot corpus geschreven Nederlands.

'We hebben ook gezocht naar de accounts van bekende Nederlanders zoals politici, waarbij we er vanuit gaan dat we vrij gebruik mogen maken van de informatie die zij via Twitter verspreiden', vertelt Henk van den Heuvel, coördinator van de verzameling nieuwe media van SoNaR. 'Verder verzamelen onze Vlaamse collega's ook nog twitteraccounts in Vlaanderen.'

Naast tweets worden ook chats uit chatboxen, posts van internetfora, blogs en e-mails in het corpus opgenomen. 'Uiteindelijk zullen bijna alle tekstcategorieën voorkomen

in SoNaR', legt Van den Heuvel uit. 'Zo hebben we boeken en brochures, bijvoorbeeld patiëntenfolders en voorlichtingsmateriaal. Maar ook rapporten en jaarverslagen en zelfs werkstukken en scripties van scholieren en studenten.'

SoNaR zal in december dit jaar afgerond zijn en bevat dan ruim 500 miljoen woorden geschreven Nederlands. 'De Centrale voor Taalen Spraaktechnologie (TST-Centrale) van de Nederlandse Taalunie gaat het corpus vervolgens beheren', aldus Van den Heuvel. 'Onderzoekers en ontwikkelaars kunnen daar aankloppen om de data te gebruiken voor hun project.'

Wie deze gebruikers precies zullen zijn weten de samenstellers niet; daarom proberen ze het corpus zo

compleet mogelijk te maken. De verzamelde teksten moeten een realistisch beeld geven van de manier waarop mensen in het Nederlands schrijven. Dit gaat lang niet altijd volgens de officiële regels voor grammatica en spelling. Verder worden de teksten voorzien van allereerste metadata zodat onderscheid naar bijvoorbeeld leeftijd, geslacht of woonplaats mogelijk is.

Het SoNaR-corpus wordt gefinancierd door het Nederlands-Vlaamse STEVIN-programma voor spraak- en taaltechnologie. Aan het project werken onderzoekers mee van de universiteiten van Nijmegen, Tilburg, Enschede, Utrecht, Gent en Leuven. Projectleider is Nelleke Oostdijk van de Radboud Universiteit Nijmegen. (ER)