



E-DATA & RESEARCH

Jaargang 12 | 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: CentERdata, CLARIAH, DANS, Huygens ING, de Koninklijke Bibliotheek en het Rijksmuseum.

INHOUD

2 Hoe open access is Nederland anno 2018?

3 Martijn Kleppe, hoofd Onderzoek KB



4/5 spread 'Privacy'
 • Gevolgen AVG voor data-archieven
 • Gouden tips van lopend onderzoek
 • ODISSEI biedt strikt beveiligde omgeving
 • DataTag-tool bepaalt toegangsniveau data

7 Stand van zaken CLARIAH-CORE

8 Ewoud heeft een mooie taak voor bibliotheken

E-data kunt u gratis per post en/of digitaal ontvangen. Mail uw verzoek naar edata@dans.knaw.nl.



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

Nederlab geeft flinke impuls aan diachroon onderzoek

Miljoenen teksten digitaal doorzoekbaar

Op 1 juli wordt Nederlab opgeleverd: met miljoenen Nederlandse teksten biedt het onderzoekers een schat aan informatie over de Nederlandse taal en cultuur.

Mathilde Jansen

In 2012 kreeg Nederlab 3,2 miljoen euro toegekend door NWO. Nu, 6 jaar later, geeft Nederlab met zo'n 25 collecties aan teksten, lopend van de 13^e tot 21^e eeuw, een flinke impuls aan diachroon geesteswetenschappelijk onderzoek. Van kranten teksten tot dagboeken: in Nederlab vind je al-

'Na inlog kunnen onderzoekers en studenten een eigen corpus samenstellen'

lerlei soorten teksten. Het oudst zijn de dertiende-eeuwse teksten uit het Corpus Gijsseling, gevolgd door de veertiende-eeuwse teksten uit het Corpus Van Reenen-Mulder. Meer van deze tijd zijn teksten uit grote krantencorpora of hele romans afkomstig uit de Digitale Bibliotheek voor de Nederlandse Letteren (DBNL). Voor onderzoekers biedt Nederlab een schat aan informatie over de Nederlandse taal en cultuur. Het Nederlab-consortium bestond uit het Meer-

tens Instituut, het Instituut voor Nederlandse taal, het Huygens ING, de Universiteit Nijmegen, de Koninklijke Bibliotheek, de DBNL en diverse dataleveranciers. Niet alleen veel data werden bijeengebracht, ook werd gezorgd voor een onderliggende infrastructuur en een groot aantal tools. Doordat de data taalkundig verrijkt zijn, is het mogelijk om bijzonder complexe zoekopdrachten uit te voeren over een heel breed corpus.

40 miljard annotaties

Hennie Brugman is IT-ontwikkelaar aan het Meertens Instituut en coördinator van Nederlab. Hij vertelt hoe de zoekmachine is opgebouwd: "Bij de oplevering bevat Nederlab dertig miljard woorden. Die hebben we zo veel mogelijk voorzien van taalkundige annotaties via automatische processen. Ieder woord is bijvoorbeeld gekoppeld aan een lemma en een woordsoort. Zo ontstaat een structuur van aan elkaar gekoppelde lagen van tekstwaarden. Het zoeken kun je doen door alle lagen heen; dat is de basis van de hele machine. Bij tien miljard woorden praat je dus al gauw over 40 miljard annotaties."

Ontwikkeling door tijd

Iedereen kan Nederlab gebruiken, maar onderzoekers en studenten hebben meer onderzoeksmogelijkheden na inlog. Zo kunnen ze hun eigen corpus samenstellen. Nicoline van der Sijs, taalkundige aan het Meertens Instituut, schreef destijds het projectvoorstel en is ook daarna nauw betrokken geweest bij Nederlab. Voor haar onderzoek maakt ze al regelmatig gebruik van het nieuwe onderzoeksportal: "Wat ik interessant vind, is dat je een ontwikkeling kunt volgen over een periode van eeuwen. Dat kon voorheen niet omdat alle corpora met Nederlandse teksten versnipperd waren. Ik denk dat we hierdoor nieuwe inzichten krijgen in de ontwikkeling van het Nederlands en de oorzaken van taalverandering." www.nederlab.nl/onderzoeksportal

🛒 Het evangelie volgens Matteüs

gerne: non-fictie, bijbel/bijbelteksten

collectie: SoNaR

aantal hits: 2

woord:	God	zal	ook	met	hen	medelijden	hebben
lemma:	God	zullen	ook	met	hen	medelijden	hebben
pos:	SPEC	WW	BW	VZ	VNW	N	WW
kenmerken:	deeleigen	pv		init	pers	soort	inf
		tgw			pron	ev	vrij
		ev			obl	basis	zonder
					vol	onz	
					3p	stan	
					mv		

De zoekterm 'medelijden hebben' binnen de Nederlabportal 'geavanceerd zoeken' in 'zoeken in tekst' toont binnen de dataset 'Het evangelie volgens Matteüs' 2 hits met de volgende taalkundige tags: lemma (woord), pos (woordsoort, bijvoorbeeld spec = speciaal, ww = werkwoord, vz = voorzetsel), kenmerken verwijst naar de functie in de zin (pv = persoonsvorm, tgw = tegenwoordige tijd, ev = enkelvoud). Met dit resultaat kan een onderzoeker nagaan hoe vaak de combinatie 'medelijden hebben' in deze tekst voorkomt. credits Nederlab

OPROEP



Wint u de Nederlandse Dataprijs '18?

Op 28 november wordt de Nederlandse Dataprijs uitgereikt. Een prijs voor een onderzoeker of onderzoeksgroep die extra bijdraagt aan de wetenschap door onderzoeksdata beschikbaar te maken voor aanvullend of nieuw onderzoek.

Nomineer voor 1 juli

Bent of kent u zo'n onderzoeker met een goed voorbeeld van gedeelde data? U kunt tot 1 juli uzelf of een ander via researchdata.nl nomineren.

De organisatie van de Nederlandse Dataprijs is in handen van Research Data Netherlands. Kijk voor meer informatie over de prijs, zoals de categorieën, de prijzenpot en de criteria, op de website van RDNL. (HB)

researchdata.nl/diensten/dataprijs

Amsterdam Time Machine: reizen door tijd en ruimte

Het consortium Amsterdam Time Machine (ATM) wil op termijn een 'Google Earth' maken van het verleden. Gebruikers kunnen zich hiermee door Amsterdam verplaatsen in ruimte en tijd, op wijk-, straat- en huishoudniveau. ATM verbindt de Amsterdamse erfgoedinstellingen in AdamNet, de creatieve industrie, de gemeente Amsterdam en sociaal- en geesteswetenschappelijk onderzoekers aan zowel de universiteiten als bij de KNAW. Door een financiële bijdrage uit CLARIAH komt de geo-infrastructuur HisGIS van de Fryske Akademy beschikbaar voor ATM. Zo komen data en tools ter beschikking voor onderzoek naar Amsterdam, en voor geesteswetenschappelijk onderzoek in het algemeen. (MM) www.create.humanities.uva.nl/amsterdam-time-machine