

Door data anders naar schilderijen kijken

Operatie Nachtwacht levert 600 TB aan data op

Operatie Nachtwacht, hét onderzoeks- en restauratieproject van Rembrandts meesterwerk, levert zoveel data op, dat het schilderij straks vele malen sneller te onderzoeken is. *Bas Nederveen*

“We monitoren De Nachtwacht al heel lang”, zegt Katrien Keune, hoofd Science van het Rijksmuseum. “Bij De Nachtwacht vallen bepaalde gebieden op, zoals het hondje, daar zit een wittige waas over. Dat is niet zoals Rembrandt het bedoeld heeft.” Reden voor een grondig onderzoek, met als doel het schilderij optimaal te behouden voor de toekomst. “Het onderzoek levert ons beter begrip op van de oorspronkelijke gedaante en de huidige staat van het schilderij. Ook geeft het onderzoek inzicht in de vele veranderingen die het schilderij de afgelopen eeuwen onderging.”

Onderzoekstechnieken

Tijdens de onderzoeksfase van het project worden de nieuwste en meest geavanceerde onderzoekstechnieken ingezet, waaronder macro-röntgenfluorescentie (MA-XRF), reflectie imaging spectroscopie (RIS) en optische coherentiemografie (OCT). Enkele technieken zijn nooit eerder op deze schaal gebruikt binnen het museum, zoals macro-röntgendiffractie scanning (MA-XRD), ontwikkeld en uitgevoerd door de Universiteit van Antwerpen. Met MA-XRD wordt inzicht verkregen in kristallijne pigmenten in de verf.

Complex proces

Keune: “We hebben MA-XRD eerder op Rembrandts portretten van Marten en Oopjen toegepast, maar toen konden we alleen nog maar lijnscans maken. Tegenwoordig kunnen we een plaatje maken. Omdat De Nachtwacht zo groot is, is het meten met de verschillende onderzoeks-



Werk in uitvoering: onderzoekstechniek macro-röntgenfluorescentie (MA-XRF) maakt met behulp van röntgenstraling één foto per chemisch element. Doordat de straling tot in onderliggende lagen doordringt, kunnen o.a. verborgen schilderijen zichtbaar worden gemaakt met fascinerende mogelijkheden voor kunsthistorisch onderzoek als gevolg. Credits: Rijksmuseum

technieken een complex proces. Om het hele oppervlak van het schilderij te kunnen scannen, is voor dit project een speciaal imaging frame ontwikkeld.”

Ongekende hoeveelheid data

De onderzoeksfase genereert een voor het museum ongekende hoeveelheid data. Ruim 12.000 foto's leveren al een samengesteld beeld op van het schilderij van een miljoen pixels breed. Naar schatting leveren alle onderzoekstechnieken tezamen 600 TB aan data op. Ter vergelijking: tot nu toe beheert het Rijksmuseum ongeveer 14,5 TB

aan onderzoeksdata. Om alle nieuwe data te verwerken en te begrijpen, zal gebruik gemaakt worden van technieken als *data fusion*, *machine learning* en *interactive visualization*. Hierdoor is het schilderij straks vele malen sneller te onderzoeken dan menselijk te doen is. Keune: “De kwantiteit aan data is hier de kracht. Dit soort technieken gaan ons nieuwe inzichten geven en nieuwe verbanden laten zien. De manier waarop we in dit project met data omgaan, gaat ons leren om op een andere manier naar schilderijen te kijken. Dat is heel spannend.”

rijksmuseum.nl/nachtwacht

Nieuw platform maakt data beter beschikbaar voor extern onderzoek

CLaaS brengt software naar data

CLARIAH ontwikkelt CLaaS, een platform waarop software naar de data gebracht kan worden, in plaats van andersom. *Erica Renckens*

Onderzoek doen met anderens data kan logistiek en administratief een hele klus zijn. Neem de collectie digitale kranten van de KB. “Bij het delen van deze data is het auteursrecht soms een struikelblok”, vertelt Steven Claeysens, conservator digitale collecties. “Het maken van kopieën is een auteursrechtelijke handeling met risico op ‘lekken’ en de administratie is niet altijd even overzichtelijk.” Het online platform CLaaS moet dit in één klap oplossen.

Jauco Noordzij, ontwikkelaar van CLARIAH-as-a-Service (CLaaS) namens het Huygens ING: “Via CLaaS werkt de software voor analyse en visualisatie daar waar de data zijn, bijvoorbeeld

bij de KB, in plaats van andersom.” Onderzoekers kunnen zo sneller bij het materiaal, dat bovendien altijd up-to-date is. Naast de KB werkt CLARIAH ook samen met partners als Beeld en Geluid en het KNAW Humanities Cluster, die ook over grote hoeveelheden data beschikken.

Softwarestandaarden

“Het platform sluit aan bij de softwarestandaarden van verschillende vakgebieden. Dat betekent dat software die is ontwikkeld volgens die standaarden probleemloos werkt buiten de onder-

zoeksinstelling waar deze ontworpen is”, aldus Noordzij. Ook onderzoekers die met eigen, andersoortige scripts werken, kunnen CLaaS gebruiken. “We bieden dan alleen geen garanties voor toekomstig gebruik. Dat kan betekenen dat opgeslagen resultaten na verloop van tijd worden gewist.”

Onderzoekers die zelf niet programmeren, zullen weinig merken van CLaaS, vermoedt Noordzij. “Maar steeds meer onderzoekers willen scripts kunnen aanpassen voor specifieke onderzoeksvragen. Voor hen willen we de software op het platform zo aanbieden dat deze zonder programmeerkennis gemakkelijk is aan te passen voor eigen onderzoek.”

Het eerste prototype van CLaaS zal eind 2019 klaar zijn. Deze zal gedurende de looptijd van CLARIAH continu updates krijgen. De eerste bruikbare versie wordt begin 2021 verwacht.

clariah.nl



E-DATA & RESEARCH

Jaargang 14 | nummer 1

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, KNAW Humanities Cluster, de Koninklijke Bibliotheek en het Rijksmuseum.

INHOUD

3 ATHENA toont biodiversiteit in online portal

4 Deze data zijn ook sinds kort beschikbaar

5



Netwerkmanager Helmus: “Samen kom je verder”

6 Huygens ING zet 140.000 scans van brieven online

6 Inspiratie: werken met de Secure Supercomputer

7 Oud nieuws beschikbaar voor nieuw onderzoek

8 Gastcolumnist Van Ham: Hollandse microdata paradijs voor wetenschappers

E-data wordt gratis toegezonden aan relaties van de stakeholders. Ook een uitgave ontvangen? Mail de redactie: edata@dans.knaw.nl.



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

Landelijke workshop Onderzoekinformatie

Elly Dijk

Op 3 september organiseerden DANS en de Koninklijke Bibliotheek een middag over het uitwisselen van gegevens op het gebied van onderzoekinformatie en Open Science. De middag was deels de opvolger voor het zogenaamde WISH-overleg (Werkgroep Implementatie SHARE), dat ooit in het kader van de ontwikkeling van de Nederlandse wetenschappelijke infrastructuur was ingesteld. Zo'n 60 repositorymanagers en -beheerders, open access medewerkers en andere geïnteresseerden van universiteiten, hbo en andere onderzoeksinstituten waren aanwezig. Het eerste onderwerp van gesprek was het harvesten van wetenschappelijke publicaties door DANS (ten behoeve van de wetenschappelijke portal NARCIS) en het E-depot van de KB. In het E-depot worden wetenschappelijke publicaties duurzaam en open toegankelijk bewaard. Daarna volgden twee presentaties over NARCIS. Eerst een presentatie over vernieuwingen in NARCIS, waaronder de in het Europese FREYA-project ontwikkelde Persistent Identifiers. De tweede presentatie betrof de ontwikkeling van NARCIS en de doorlevering van de publicaties aan OpenAIRE. Tot slot waren er twee gastsprekers, Just de Leeuwe (bibliotheek TU/Delft) en Arjan Schalken (bibliotheek VU), die elk een presentatie gaven over de ontwikkeling van open access publicaties in Nederland. Just de Leeuwe ging daarbij in op de open access aantallen die door de universiteiten worden geleverd aan de VSNU, ten behoeve van het ministerie OCW. Het voorlopig percen-

GEHOORD & BIJGEWOOND



Het EYE-museum was decor van de iPRES-conferentie

tage ligt voor 2018 ruim boven de 50 procent; meer dan in 2017. Arjan Schalken sprak over de diverse uitdagingen die spelen om het overheidsdoel van 100% open access publicaties in 2020 te halen, welke beleidsbeslissingen er moeten worden genomen en hoe werkprocessen moeten worden ingericht. Na afloop bleven de deelnemers nog lang napraten over de diverse onderwerpen. Ook gaven zij aan dat de middag zeer informatief en inspirerend was geweest. De presentaties staan op de website van DANS.

dans.knaw.nl

Open Science FAIR Porto 2019

Eliane Fankhauser, Ellen Leenarts
Na Athene in 2017 vond de tweede editie van de Open Science FAIR in Porto plaats. Deze conferentie onderscheidt zich van andere events door de workshops waarin deelnemers uitgebreid over Open Science en FAIR discussiëren. Dit jaar werd

het evenement gesponsord door vier grote EOSC-projecten: FAIRsFAIR, RDA Europe, OpenAIRE en FREYA. Na de plenaire opening volgden twee keynotes met verschillende invalshoeken voor Open Science: vanuit het perspectief van de Europese Commissie en vanuit de waarneming van een onderzoeker. De onderzoeker, Paula Masuzzo, betoogde dat het tijd is voor andere criteria voor goed onderzoek dan artikelen en subsidies zodat ook het leveren van goed gedocumenteerde data en code mee gaat tellen. In de vele workshops werd flink gediscussieerd om nieuwe aanbevelingen en acties te bedenken. Belangrijke onderwerpen die aan bod kwamen, waren: data stewardship, de FAIRness van EOSC-trainingsmateriaal, services voor de ondersteuning van FAIR en Plan S. Het blijkt dat men, bij het realiseren van Open Science en FAIR, vaak dezelfde uitdagingen ervaart zoals een gebrek aan samenwerking op regionaal en internatio-

naal niveau, onduidelijke verantwoordelijkheden van verschillende belanghebbenden en een benodigde cultuurverandering voor Open Science. Natuurlijk werd er ook over oplossingen en verbeteringen van bijvoorbeeld infrastructuur en diensten nagedacht. Niet zelden werden ook concrete acties geformuleerd die de komende maanden door de organisatoren kunnen worden opgepakt. Op de laatste conferentiedag werden een aantal kleine projecten, tools en diensten gepresenteerd in korte demonstraties. Dit was een handige manier om iets te leren over dit soort initiatieven en tegelijkertijd in gesprek te komen met de makers ervan. We vertrokken uit de mooie stad Porto en verlieten de Open Science FAIR 2019 met het gevoel dat we aan de verdere ontwikkeling van Open Science hebben bijgedragen.

opencsciencefair.eu

EYE on the Horizon

Sophie Ham

Het prachtige gebouw van het EYE Filmmuseum aan de overkant van het IJ in Amsterdam was dit jaar het decor van de iPRES-conferentie. Met zo'n locatie is een congres al voor de helft geslaagd - en deze 16e editie van de grootste conferentie op het gebied van digitale duurzaamheid trok dan ook maar liefst 420 deelnemers uit 33 landen. Van 16 tot en met 21 september waren er workshops, demonstraties, lezingen, presentaties en - een van mijn favorieten - Great Digital Preservation Bake-off sessies.

Digitale duurzaamheid was vijftien jaar geleden nog een soort niche in erfgoedland, waar vooral pionierende IT-ers zich mee bezig hielden. Ook iPRES ging jarenlang vooral om die technische kant: migratie, emulatie, het toevoegen van duurzame metadata en systemen die dat soort dingen heel goed kunnen. Maar door de digitale revolutie is zowel het vakgebied als het congres veel breder geworden. Meer en meer gaat de aandacht naar de praktijk van digitale duurzaamheid en bredere ontwikkelingen binnen het veld. Nieuwe bestandsformaten, onderzoeksdata, webarchivering en social media stonden dan ook prominent op de agenda vanuit een meer beleidsmatige visie, en soms zelfs een existentiële. Dat laatste bleek heel goed uit de uitstekende keynotes. Het congres opende met een vrij sombere - maar absoluut tot nadenken uitnodigende - bespiegeling van Geert Lovink over de downside van sociale media. Hij riep op onszelf te organiseren en los te breken uit de greep van de grote platformen. Dat stond in een bepaald opzicht in contrast met het imponerende verhaal van de laatste spreker, Eliot Higgins van Bellingcat. Hij liet indirect de kracht van sociale media en burgerjournalistiek zien door de *crowd* bijvoorbeeld in te schakelen bij het ontkrachten van Russisch nepnieuws. En zo vertelde deze conferentie vooral ook over het belang van sociale media in het algemeen - want of je ze nu verafschuwt of nodig hebt - ze vormen onze maatschappij. En dus aan ons, archivariissen en bibliothecarissen, de taak daar een *preservation policy* voor te verzinnen.

ipres2019.org

Een datasteward zorgt voor data gedurende de hele onderzoekscyclus

Datastewardship in vogelvlucht

Stel je een wereld voor waarin data perfect voorzien zijn van metadata, zodat iedereen ze altijd kan vinden en kan hergebruiken. Dat zou het leven van onderzoekers een stuk eenvoudiger maken. Kan de datasteward hiervoor zorgen?

Frans Huigen

Datastewardship staat voor het zorgen voor data, gedurende de hele onderzoekscyclus. De onderzoeker is verantwoordelijk voor de data, maar kan hierbij professionele ondersteuning gebruiken. Hier is een rol weggelegd voor de data expert, ook wel datasteward genoemd.

500.000 data experts

Volgens de High Level Expert Group (HLEG)

European Open Science Cloud (EOSC), ingesteld door de Europese Commissie, hebben we op korte termijn 500.000 data experts nodig. Maar waaraan moeten die data experts voldoen? FAIRsFAIR, een in maart gestart Horizon2020-project, bekijkt hoe het beroep van datasteward geprofessionaliseerd kan worden en welke beroepsvaardigheden nodig zijn om onderzoeksdata vindbaar, bruikbaar en toe-

gankelijk te maken én voor de langetermijn te houden: FAIR Stewardship Skills. Een ander initiatief voor aanstormende datastewards, een Research Data Science Summer School, werd deze zomer verzorgd door het Committee on Data for Science and Technology (CODATA) en de Research Data Alliance (RDA).

Nationaal aan de slag

Ook nationaal zijn er veel initiatieven op het gebied van het professionaliseren van het beroep van datasteward. Enkele voorbeelden zijn: de RDNL-cursus 'Essentials 4 Data-support', het datastewardshipproject van Wageningen University & Research, het door ZonMw gefinancierd project 'Towards FAIR Data Steward as profession for the Life Sciences' en een project van het platform Nationaal Plan Open Science (NPOS) voor

het professionaliseren van de dataprofessional. In september is de online cursus 'Delivering Research Data Management Services' gestart, een initiatief van RDNL en het Britse Digital Curation Centre en in november organiseert de European University Foundation samen met de Universiteit van Amsterdam en FAIRsFAIR een focusgroepbijeenkomst 'teaching (FAIR) data management and data stewardship'.

Deze en andere interessante feiten over de groeiende behoefte aan datastewards en de professionalisering van deze beroepsgroep staan in het rapport van het Landelijk Coördinatiepunt Research Data Management (LCRDM), 'Datastewardship op de kaart: Een verkenning van taken en rollen in Nederlandse onderzoeksinstituten'.

doi.org/10.5281/zenodo.2642066



Credits: Spresso Design Studio & LCRDM, CC-BY

OVERNEMEN ARTIKELEN

Wilt u een artikel uit dit blad overnemen? Dat mag altijd, maar vermeld wel de bron (E-data & Research) en de naam van de auteur van het artikel. Neem ook contact op met de hoofdredacteur (zie colofon) om door te geven waar artikelen geplaatst worden.

Van vergeet-me-nietjes tot nertsen

Online zoeken naar Nederlandse biodiversiteit

Afgelopen zomer is ATHENA gepresenteerd, een portal met een enorme hoeveelheid online informatie over biodiversiteit in Nederland door de eeuwen heen. *Erica Renckens*

Het gaat niet goed met de Nederlandse biodiversiteit, horen we regelmatig in het nieuws. Hoe komt dat? En hoe is hier in het verleden mee omgegaan? Om dergelijke vragen te kunnen adresseren, is in ATHENA informatie bij elkaar gebracht en gestandaardiseerd. Deze portal met historische bronnen, archeologisch materiaal en ecologische databestanden (zoals tellingen) maakt interdisciplinair onderzoek naar biodiversiteit mogelijk.

Toegang tot acht databases

“Dit is een gezamenlijk project van historici, archeologen en biologen”, vertelt projectleider Thomas van Goethem (Universiteit Utrecht). Van Goethem werkt samen met onderzoekers van de Radboud Universiteit en Wageningen Universiteit aan de portal. Via ATHENA krijgen natuurliefhebbers en onderzoekers toegang tot acht verschillende databases, waaronder BoneInfo (een database met archeozoologisch en fysisch antropologische informatie), het NLBIF (Netherlands Biodiversity Information Facility) en het Nederlands Instituut voor Kunstgeschiedenis (RKD).

Doorontwikkeling

Sinds de lancering zijn de gebruikers positief. “Opvallend veel mensen willen nieuwe bronnen en gegevens aandragen. Daar staan we erg voor open.” Ook op andere vlakken zal ATHENA nog doorontwikkeld worden. “We willen ervoor zorgen dat de portal optimaal functioneert, bijvoorbeeld door collecties beter doorzoekbaar te maken. Daarbij willen we gebruikmaken van tools die zijn ontwikkeld binnen CLARIAH, maar ook van *crowdsourcing*, bijvoorbeeld voor het annoteren van schilderijen.”



In ATHENA staan veel afbeeldingen, zoals deze pelikaan en ander gevogelte bij een waterbassin, bekend als ‘Het drijvend veertje’, Melchior d’Hondecoeter, ca. 1680
Credits: Rijksmuseum

CLARIAH, de digitale infrastructuur voor de geesteswetenschappen, heeft de ontwikkeling van ATHENA sinds 2015 gefinancierd. “We proberen de drie pilaren van CLARIAH - taal- kunde, mediastudies en gestructureerde data - te overbruggen. Daarvoor overlegden we met onderzoekers uit die vakgebieden om te kijken of we onderdelen van hun werk konden gebruiken. Ook maken we gebruik van de database-infrastructuur die binnen CLARIAH is ontwikkeld.” Verschillende wetenschappers

maken bij hun onderzoek al gebruik van de nieuwe portal. Zo is de thesaurus al gebruikt in het CLARIAH-project SERPENS, waarin deze werd gekoppeld aan krantenarchief Delpher om de geschiedenis van ongedierte-overlast te bestuderen. Van Goethem: “Samen met Jan Luiten van Zanden gebruik ik ATHENA voor een onderzoek naar de ontwikkeling van de Nederlandse biodiversiteit in de laatste honderd jaar.”

athena-research.org

Welkom voor KNAW Humanities Cluster

Het KNAW Humanities Cluster (HuC) is een nieuwe stakeholder van dit blad. Het HuC bestaat sinds 1 oktober 2016 uit drie instellingen: het Huygens ING, het Meertens Instituut en het Internationaal Instituut voor Sociale Geschiedenis (IISG). Alle drie zijn onderzoeksinstituten, waarbij het IISG sociale en economische geschiedenis onderzoekt op wereld-

schaal, het Huygens ING zich concentreert op Nederlandse geschiedenis, wetenschapsgeschiedenis en letterkunde en het Meertens Instituut zich bezighoudt met de Nederlandse taal en cultuur. Alle drie werken ze met data en ontwikkelen ze tools en websites. Samen hebben ze een infrastructurele taak, willen ze interdisciplinair onderzoek mogelijk maken, datasets aan elkaar koppelen en discipline-onafhankelijke standaarden ontwikkelen. Het HuC ziet in E-data een professioneel kanaal om over samenwerkingen, resultaten en projecten te berichten. (HB)

huc.knaw.nl

Nieuwe open standaard voor kwalitatieve data

Er is een nieuwe open standaard voor kwalitatieve data: de REFI-QDA-standaard. Deze standaard maakt interoperabiliteit mogelijk tussen programma's voor kwalitatieve data analyse. Het stelt gebruikers in staat om gegevens tussen programma's uit te wisselen en duurzaam te archiveren. Al bestaande programma's voor de analyse van kwalitatieve data, zoals ATLAS.ti, MAXQDA en NVivo, gebruiken eigen bestandsformaten, wat het uitwisselen van bestanden tussen verschillende

programma's moeilijk maakt. Het uitvoer-bestandsformaat van REFI-QDA is .qdp.

DANS beschouwt dit als een voorkeursbestandsformaat voor langetermijn toegang tot in QDAS gecreëerde projecten en codeboeken. Het is een open standaard en elk programma kan het implementeren, waardoor het aantal softwareprogramma's dat met elkaar kan 'praten' toeneemt. (RvH)

qdasoftware.org

KORT

OpenSoNaR: luisteren naar spraak

Afgelopen voorjaar heeft het Instituut voor de Nederlandse Taal een nieuwe webversie van OpenSoNaR gelanceerd. Deze applicatie geeft niet alleen toegang tot de 500 miljoen geschreven woorden Nederlands uit het SoNaR-corpus, maar nu ook tot de 900 uur spraak uit het Corpus Gesproken Nederlands (CGN). De geluidsfragmenten zijn online te beluisteren. Gebruikers kunnen de data doorzoeken op woordniveau of met een complexere zoekopdracht, zoals annotaties. OpenSoNaR is gratis toegankelijk met een gebruikersaccount van een universiteit of met een CLARIN-account. (ER)
openonar.clarin.inl.nl

CLARIAH gaat Europees

De CLARIAH-beweging breidt zich steeds verder uit over Europa. Afgelopen juni vond in het Belgische Gent de eerste Benelux-bijeenkomst plaats, waarbij naast Nederlandse ook Belgische en Luxemburgse onderzoekers en ontwikkelaars hun kennis en ideeën samenbrachten voor de ontwikkeling van een digitale onderzoeksinfrastructuur voor de geesteswetenschappen. Daarnaast zullen in Duitsland de komende twee jaar de nationale projecten CLARIN-D en DARIAH-DE met steun van de Duitse overheid samengaan in het project CLARIAH-DE. (ER)
clariah.nl/over/internationaal

Nieuwe online cursus Delivering RDM Services

In september ging de cursus *Delivering Research Data Management Services* van start, georganiseerd door het Digital Curation Centre (DCC) en Research Data Netherlands (RDNL). De cursus, in de vorm van een Massive Open Online Course (MOOC), richt zich op personen die onderzoekers ondersteunen bij het beheren en delen van hun onderzoeksgegevens. De cursus behandelt verscheidene diensten ter ondersteuning van wetenschappelijk onderzoek. De cursus wordt aangeboden via het platform FutureLearn en is tot 21 oktober vrij toegankelijk. In 2020 volgt waarschijnlijk een tweede ronde. (FvG)
futurelearn.com/courses/delivering-research-data-management-services

Nieuwe call voor digitale infrastructuren

In de nieuwe call Digitale Infrastructuur kunnen onderzoekers uit de sociale en geesteswetenschappen financiering aanvragen voor zowel de versterking en opschaling van bestaande digitale infrastructuur als voor het opzetten van nieuwe initiatieven op dit gebied. De call is uitgezet door het Platform Digitale Infrastructuur Social Sciences & Humanities (PDI-SSH). Dit platform is onlangs opgericht door het SSH-Beraad, CLARIAH en ODISSEI om een deel van het sectorplan van het Ministerie van Onderwijs, Cultuur en Wetenschap voor de sociale en geesteswetenschappen strategisch vorm te geven. Voorstellen kunnen tot uiterlijk 31 oktober 2019 worden ingediend. (ER)
pdi-ssh.nl

SINDS KORT BESCHIKBAAR

Dit overzicht toont databestanden die recent beschikbaar zijn gekomen bij CentERdata en Data Archiving and Networked Services.

CentERdata

Het Wiv-referendum 2018

In opdracht van dr. Kristof Jacobs (Radboud Universiteit) en Stichting KiezersOnderzoek Nederland (SKON) zijn in het LISS panel van februari t/m april 2018 drie vragenlijsten afgenomen over de effecten van het referendum van 21 maart 2018 over de Wet op de inlichtingen- en veiligheidsdiensten (Wiv). In december 2018 is het rapport 'Het Wiv-referendum; Nationaal Referendum Onderzoek 2018' gepubliceerd. Dit rapport is te vinden via <https://kennisopenbaarbestuur.nl/rapporten-publicaties/wiv-referendumonderzoek-2018/>. De data van deze drie vragenlijsten in het LISS panel - Wiv Referendum 2018 - zijn beschikbaar via LISS Data Archive.

lissdata.nl

Ook sinds kort beschikbaar:

Studies LISS panel

- Suetens, S.; Cettolin, E.; Perez Padilla, M., december 2016, Social Status
- Meerkerk, G.J., februari 2017, Gambling Policy
- Mari, G., december 2016 - maart 2017 - juni 2017, Gender, parenthood and labor market discrimination: a survey experiment
- Denissen, J.; Geenen, R.; van Aken, M., augustus 2017, Big Five Inventory (BFI-NL)
- Van der Voet, J., februari 2018 - april 2018, ABW Waste disposal - part 2 & 3
- CentERdata, januari 2017 - december 2017, Initial Questionnaire
- CentERdata, augustus - september 2018, Religion and Ethnicity - Wave 11
- CentERdata, december 2018 - maart 2019, Politics and Values - Wave 11
- CentERdata, mei - juni 2019, Personality - Wave 11



Deze bestanden zijn kosteloos beschikbaar via lissdata.nl. Bezoek deze site of scan de QR-code.

DANS

Franciscaners op de Voorstraat

Een bijzonder rijke dataset is toegevoegd aan EASY. Het betreft de onderzoeksdata vanaf 1982 over het dertiende eeuwse Minorbroederklooster in Dordrecht. De data geven een goed inzicht in de fasering en aard van het klooster gedurende de gehele bestaansperiode van de bouw tot 1572. Naast gegevens over het klooster is ook informatie verzameld over de bewoningsperiode in de periode 1200-1246 en de bewoning gedurende de post-kloosterperiode. De diverse dataset bevat onder meer gedetailleerde veldtekeningen en objecttekeningen alsmede uitgewerkt beeldmateriaal van vondsten en fraaie graffiti's.

DOI: [10.17026/dans-25r-2aam](https://doi.org/10.17026/dans-25r-2aam)



Ook sinds kort beschikbaar:

De volgende datasets zijn open access beschikbaar via het online archiverings-systeem EASY van DANS:

- Bazelmans, dr J. (Rijksdienst voor het Cultureel Erfgoed); Purmer, dr M. (Natuurmonumenten); Kort, J.-W. de (Rijksdienst voor het Cultureel Erfgoed); Verspaandonk, A. (2017): Klei- en porseleinen pijpen Kamp bij Oirschot 1832-1834. DANS. DOI: [10.17026/dans-xgx-kj27](https://doi.org/10.17026/dans-xgx-kj27)
- Berkhout, Dr D.J. (University of Amsterdam); Statsch, MSc P.D. (University of Amsterdam) (2019): Dataset leden Provinciale en Gedeputeerde Staten 2018. DANS. DOI: [10.17026/dans-x2j-zbcd](https://doi.org/10.17026/dans-x2j-zbcd)
- Daniël, A.A.W.J. (Gemeente Nijmegen) (2017): Sporen vanaf het midden-neolithicum tot en met de Tweede Wereldoorlog in Nijmegen-Noord. Archeologisch onderzoek op 't Klumke; project Ngk4, Archeologische

Berichten Nijmegen - Rapport 67.

- DANS. DOI: [10.17026/dans-zzh-ah3f](https://doi.org/10.17026/dans-zzh-ah3f)
- Essen, MSc M. van (ResearchNed) (2019): Monitor Cultuureducatie primair onderwijs 2018-2019. DANS. DOI: [10.17026/dans-xmc-ftmw](https://doi.org/10.17026/dans-xmc-ftmw)
- Hanzon, MA C.A. (JSO) (2019): Leren van herhaald beroep in de JeugdzorgPlus. DANS. DOI: [10.17026/dans-xkd-48aj](https://doi.org/10.17026/dans-xkd-48aj)
- Kamp, drs. J.S. van der (Erfgoed gemeente Utrecht) (2019): Schachten vol schatten. LR86: Archeologisch onderzoek in en rond het castellum in De Meern (Utrecht). Basisrapportage Archeologie 95. DANS. DOI: [10.17026/dans-xev-z43s](https://doi.org/10.17026/dans-xev-z43s)
- Luijten, PhD M (Radboud University) (2019): Goal-Directed and Habitual Control in Smokers. DANS. DOI: [10.17026/dans-zsf-94p8](https://doi.org/10.17026/dans-zsf-94p8)
- Maassen, Prof. Dr. B.A.M. (University of Groningen) (2001): Early Precursors of Familial Dyslexia: A Prospective Longitudinal Study. DANS. DOI: [10.17026/dans-x6c-wxyw](https://doi.org/10.17026/dans-x6c-wxyw)
- Schrickx, C.P. (Archeologie Hoorn) (2019): Valsmunterij op een boeren erf. DANS. DOI: [10.17026/dans-z6u-d7hk](https://doi.org/10.17026/dans-z6u-d7hk)
- Sheill, Dr G (Trinity College Dublin) (2020): The Feasibility of Implementing an Exercise Programme for Deconditioned Cancer Survivors in a National Cancer Centre: FIXCAS Study. DANS. DOI: [10.17026/dans-26u-qejk](https://doi.org/10.17026/dans-26u-qejk)
- Shelach-Lavi, Prof. G.S. (Hebrew University) (2018): The Origins of Agriculture and Sedentary Communities in Northeast China. DANS. DOI: [10.17026/dans-xs7-rmqh](https://doi.org/10.17026/dans-xs7-rmqh)
- Sociaal en Cultureel Planbureau (SCP) (2014): Lesbische, homoseksuele, biseksuele ouderen (55+) - LHBO2014. DANS. DOI: [10.17026/dans-zej-55ex](https://doi.org/10.17026/dans-zej-55ex)
- Sociaal en Cultureel Planbureau (SCP) (2013): Lesbische, homoseksuele, biseksuele jongeren - LHBJ2013. DANS. DOI: [10.17026/dans-257-nqz4](https://doi.org/10.17026/dans-257-nqz4)
- Zoolingen, MA R.J. van (Gemeente Den Haag) (2019): Den Haag - Madepolderweg 45 en 55-57 - IVO-p. DANS. DOI: [10.17026/dans-xtw-r95w](https://doi.org/10.17026/dans-xtw-r95w)



Via easy.dans.knaw.nl zijn deze bestanden beschikbaar. Bezoek deze site of scan de QR-code.

AGENDA

19 - 23 oktober • Melbourne

ASIS&T
Met de nieuwste ontwikkelingen op het gebied van informatica.
Asist.org/am19

21 - 27 oktober • wereldwijd

Open Access Week
Dit jaar is het thema 'Open for Whom? Equity in Open Knowledge'.
Openaccessweek.org

22 oktober • Utrecht

ODISSEI Community Conference
Met voorbeelden van de ODISSEI Secure Supercomputer, LISS-call 2018 en Microdata Access Grant.
odissei-data.nl

22 - 24 oktober • Dublin

i-Society 2019
Internationale conferentie over 'the information society'.
i-society.eu

4 - 6 november • Wenen

Conference on Cultural Heritage and New Technologies
Jaarlijkse conferentie over cultureel werelderfgoed en de laatste technologische ontwikkelingen.
Chnt.at

7 november • wereldwijd

World Digital Preservation Day 2019
Volg #WDPD2019 op Twitter.
dpconline.org

14 november • Amsterdam

Smart Humanity 2019
Opvolger van het KNVI-jaarcongres, met als thema 'Shape Your Own Future'.
smarthumanity2019.knvi.nl

20 november • Den Haag

PID NL
Een workshop over de waarde en het gebruik van Persistent Identifiers.
rd-alliance.org

21 november • Amsterdam

National eScience Symposium 2019
Zesde jaarlijkse symposium over digitale uitdagingen in Open Science.
sciencecenter.nl

22 november • Amsterdam

Sharing is Caring
Dit event heeft als thema 'Expand the Source', met de focus op Research Data Management in musea en andere erfgoedinstellingen.
sharecare.nu/amsterdam-x-2019

3 - 4 december • Tampere

EDDI19
Elfde jaarlijkse Europese conferentie voor DDI-gebruikers.
Eddi-conferences.eu

9 - 12 december • Los Angeles

IEEE Big Data 2019
Het thema van de conferentie is 'Volume, Velocity, Variety, Value and Veracity'.
bigdataieee.org/BigData2019

30 januari • Utrecht

CLIN30
De jaarlijkse conferentie over computationele taalkunde in het Nederlandse taalgebied.
clin30.sites.uu.nl

17 - 20 februari • Dublin

15e IDCC
Het thema van de internationale Digital Curation Conference is 'Collective Curation: the many hands that make data work'.
dcc.ac.uk/events/idcc20

Meer dan 300 cursisten in 5 jaar tijd

Essentials 4 Data Support: een update

Research Data Netherlands biedt al 5 jaar de cursus 'Essentials 4 Data Support' aan, een inleidende cursus voor degenen die onderzoekers (willen) ondersteunen bij het opslaan, beheren, archiveren en delen van hun onderzoeksgegevens.

Research Data Netherlands (RDNL) is een samenwerkingsverband van 4TU.Centre for Research Data, DANS en SURFsara. Met deze coalitie, die ook voor andere partijen open staat, bundelen de drie dataarchieven hun krachten op het gebied van

langdurige data-archivering. 'Essentials 4 Data Support', beschikbaar in het Nederlands en het Engels, heeft al meer dan 300 datasupporters opgeleid. Studenten zijn vaak bibliotheekmedewerkers die werken als informatiespecialisten, maar ook is de cursus nuttig gebleken voor beleidsadviseurs en ICT-specialisten. Afgelopen zomer is de cursus uitgebreid met, onder andere, de nieuwste technische ontwikkelingen, juridische kwesties en overkoepelende thema's zoals FAIR en EOSC. (EV)

datasupport.researchdata.nl



Wilbert Helmus, netwerkmanager voor het Netwerk Digitaal Erfgoed:

‘Alleen ga je sneller, samen kom je verder’

Het Netwerk Digitaal Erfgoed laat met de extra middelen van OCW voor het intensiveringsprogramma *Erfgoed digitaal voor iedereen* ‘de trein die in 2015 vertrok vooral harder rijden’. Steven Claeysens

INTERVIEW

‘We inventariseren de behoeftes bij het gebruik van digitaal erfgoed’

In 2015 presenteerde het kersverse netwerk van Nederlandse erfgoedpartijen een nationale strategie die steunt op drie programma's: Houdbaar digitaal erfgoed, Bruikbaar digitaal erfgoed en Zichtbaar digitaal erfgoed. Het Netwerk Digitaal Erfgoed (NDE) wordt bij de start getrokken door vijf knooppunten: het Nationaal Archief, de Koninklijke Bibliotheek, de Rijksdienst voor het Cultureel Erfgoed, het Nederlands Instituut voor Beeld en Geluid en het KNAW Humanities Cluster. Vier jaar later spreken we met netwerkmanager Wilbert Helmus over het werk van het netwerk, het intensiveringsprogramma en de relatie tot de wetenschap.

Een wereld te winnen

“Het ministerie van OCW heeft extra middelen beschikbaar gesteld in 2019 en 2020. Deze middelen besteden we aan deskundigheidsbevordering en het sneller beschikbaar maken van generieke voorzieningen. Wat het eerste betreft is er nog een wereld te winnen. De 10 erfgoedcoaches die sinds kort aan de slag zijn, constateren een grote behoefte aan kennis bij erfgoedprofessionals. Zo zetten we vol in op *linked data*, maar de kennis bij instellingen over hoe je *linked data* maakt, is nog beperkt. Dat is het geval bij de instellingen zelf, maar ook bij leveranciers van erfgoedsoftware. Je kan het vergelijken met digitaliseren. Tien



Wilbert Helmus, netwerkmanager Netwerk Digitaal Erfgoed: “Het netwerk is succesvol op het moment dat het volledig zelforganiserend is geworden.” Foto: Marco De Swart

jaar geleden dachten we een enorme sprong te maken met het digitaliseren van ons erfgoed. Intussen weten we dat het een druppel op een gloeiende plaat is, er is nog zoveel materiaal niet gedigitaliseerd. In plaats van alle erfgoed informatie in het netwerk zo snel mogelijk als *linked data* beschikbaar te krijgen, volgen we een stapsgewijze aanpak, met software waarmee het eenvoudig wordt om *linked data* te publiceren. Vervolgens enten we elk nieuw project op het *linked data*-gedachtengoed en ontwikkelen we samen met OCW de Digitaal Erfgoed Referentie Architectuur.”

Praktische vragen

“Ondertussen stellen wij de praktische vragen uit het erfgoedveld aan de wetenschap. Waar het netwerk eerder vooral zélf nadacht over bijvoorbeeld de technische inrichting van een gedistribueerd netwerk van erfgoed informatie met toepassing van *linked data*-principes,

maakt onze recente koppeling met CLARIAH het mogelijk om die vraag te stellen aan de CLARIAH-gemeenschap, want daar zit de onderzoekskracht. We willen die onderzoekskracht heel graag gebruiken in de realisering van het tweede belangrijke onderdeel van de intensivering: het sneller beschikbaar maken van de generieke voorzieningen. Alleen ga je sneller, maar samen kom je verder. Samen met de Universiteit Gent zoeken we nu zelfs een promovendus om toegepast onderzoek te doen naar zoeken in gedistribueerde bronnen. De vacature gaat binnenkort uit.”

Rondetafelgesprekken

“Van het algemeen publiek hebben we een behoorlijk beeld hoe zij digitaal erfgoed gebruikt. Van een aantal belangrijke gebruikersgroepen hebben we dat veel minder scherp. Daarom organiseren we dit najaar een aantal rondetafelgesprekken met onderzoe-

kers, de creatieve industrie en het onderwijs. We willen nagaan wat de behoeftes zijn bij het gebruik van digitaal erfgoed. Ook willen we toetsen of we digitaal materiaal wel op de juiste manier beschikbaar stellen. De uitkomsten van deze gesprekken moeten resulteren in activiteiten voor volgend jaar.”

“NDE geeft binnen het intensiveringsprogramma bijvoorbeeld ook gehoor aan de oproep van de UNESCO om software als erfgoed te beschouwen. Wat doen we met software die gebruikt wordt in digitale kunst? Op dat vlak is het internationaal platform voor mediakunst, LIMA, een mooi voorbeeld. LIMA heeft een behoud- en beheerfunctie en verzorgt de distributie van digitale kunst voor kunstenaars. Ook de door kunstenaars gebruikte software veroudert, en is dus op termijn niet meer bruikbaar. Samen met enkele internationale onderzoeksinstituten en met grote musea zoals Tate Modern doet LIMA onderzoek naar de toekomstbestendigheid van software. Het is een voorbeeld hoe de erfgoedwereld onderzoekers weer voedt met kennis.”

Netwerkdelen

“Momenteel werken zo'n twintig projectleiders voor NDE. Als je dat vermenigvuldigt met het aantal mensen in een werkgroep, gemiddeld 5 tot 10, en het aantal actieve instellingen, dan hebben we al een behoorlijk groot bereik. Maar de kern is het denken vanuit het netwerk. De kracht van het gezamenlijk aanpakken van een issue, vanuit verschillende perspectieven en deskundigheden, blijkt in de praktijk telkens weer. Het resultaat daarvan is veelal evengoed werkbaar voor andere instellingen in het netwerk. Als iedereen steeds even denkt ‘laten we krachten bundelen, ik ben niet de enige met deze vraag’, dan is het netwerk geslaagd.”

netwerkdigitaalnerfgoed.nl

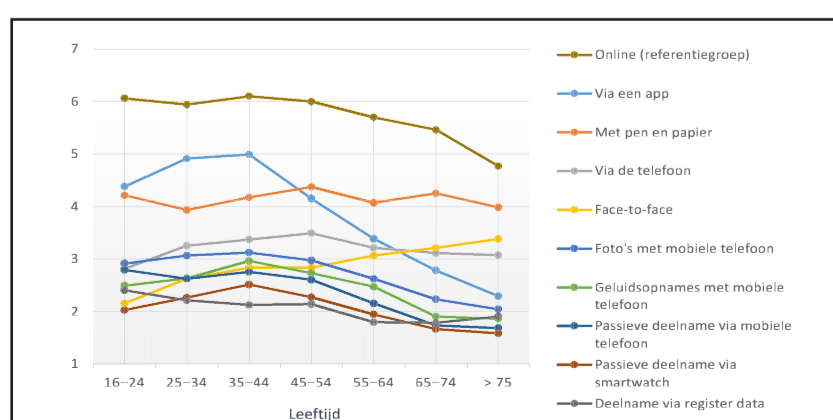
Netwerk Digitaal Erfgoed

Gevraagd naar het belang van het Netwerk Digitaal Erfgoed is de voorzitter van het NDE-bestuur Lily Knibbeler (KB) resoluut: “Het netwerk brengt alles samen: het digitale erfgoed en het erfgoedveld. Digitaal erfgoed bijeenbrengen is goed nieuws voor iedereen, zowel het brede publiek als de wetenschap. Met NDE is er nu bovendien één aanspreekpunt om de Nederlandse digitale erfgoedwereld te bereiken.”

Week van Digitaal Erfgoed

Van 25 tot en met 30 november organiseert het Netwerk Digitaal Erfgoed de Week van het Digitaal Erfgoed. Het is een week voor en door erfgoedinstellingen, in samenwerking met het ministerie van Onderwijs, Cultuur en Wetenschap. Verspreid over Nederland vinden allerlei bijeenkomsten plaats.

weekvanhetdigitaalnerfgoed.nl



Wie werkt mee aan welk soort onderzoek?

Deze grafiek laat de bereidheid van online panelleden per leeftijdscategorie en techniek zien, waarbij 1 staat voor een heel onwaarschijnlijke deelname en 7 voor een heel waarschijnlijke deelname. Uit dit onderzoek van Joris Mulder en Marika de

Bruijne (CentERdata) blijkt dat onderzoek met een app tot een bepaalde leeftijd hoog scoort, maar daarna afneemt. De bereidheid om mee te werken aan passieve vormen van dataverzameling en deelname via registerdata scoren binnen alle leeftijdsgroepen het laagst.

Benieuwd naar meer resultaten? Het volledige onderzoek *Willingness of Online Respondents to Participate in Alternative Modes of Data Collection* staat online.

(EdC)

doi.org/10.29115/SP-2019-0001

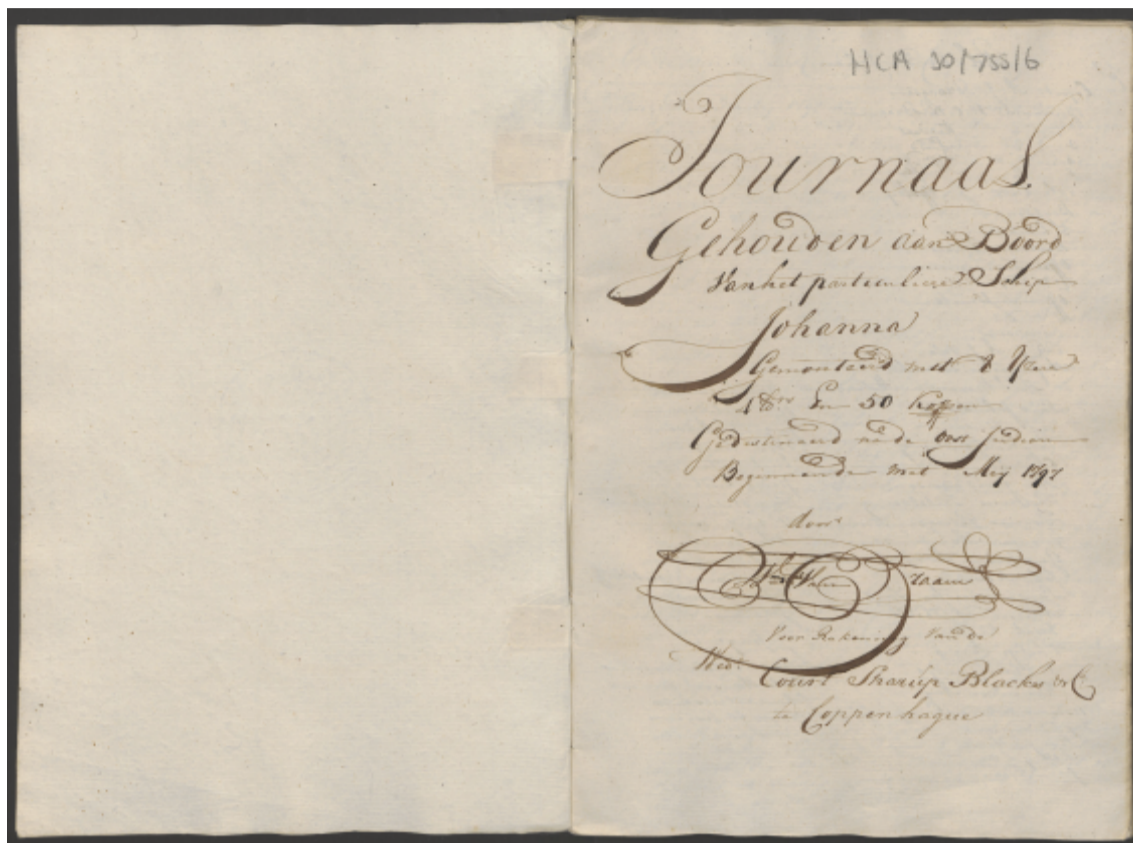
Huygens ING zet ruim 140.000 scans online

Dutch Prize Papers geven uniek inkijkje in leven in de 17e en 18e eeuw

Duizenden Nederlandse brieven werden door Britse kapers buitgemaakt tussen 1652 en 1815. De buit is onlangs door het Huygens ING gedigitaliseerd en als Dutch Prize Papers online beschikbaar gesteld.

Mathilde Jansen

De originele documenten lagen in het archief van het Londense High Court of Admiralty. De collectie bestaat onder andere uit scheepsjournalen, ladingoverzichten, rekeningen, plantagelijsten en onderzochtingen van bemanningsleden. Ongeveer een kwart van de totale collectie is afkomstig van Nederlandse schepen. De Dutch Prize Papers bevat zo'n 38.000 zakelijke en particuliere brieven. Deze zijn afkomstig uit alle lagen van de bevolking en geven daardoor een uniek inkijkje in het alledaagse leven in de zeventiende en achttiende eeuw. De meeste brieven hebben hun bestemming overigens nooit bereikt. De collectie bevat naast papieren erfgoed ook diverse textielstalen en



Titelblad van het scheepsjournaal van het Schip de Johanna, (HCA 30/755/6), kan worden geraadpleegd in de nieuwe VRE. Credits: Huygens Instituut voor Nederlandse Geschiedenis

zelfs enkele sieraden. Door hun grote variëteit zijn de Prize Papers zeer geschikt voor verschillende typen onderzoek naar bijvoorbeeld ontwikkelingen in taal en dialect, handel, materiële cultuur en sociale relaties.

De buitgemaakte documenten en objecten geven niet alleen een beeld van de Nederlandse scheepvaart en handel in de zeventiende en achttiende eeuw, maar ook van het leven van opvarenden en hun thuisfront.

In de collectie kan bijvoorbeeld gezocht worden op naam van de kapitein, op de scheepsnaam en op de datum van de kaping.

Jelle van Lottum (Huygens ING) gebruikte de Prize Papers bijvoorbeeld

voor zijn onderzoek naar de migratie van zeelieden. "Buitenlandse zeelieden speelden een cruciale rol in de Nederlandse economie. Door de Prize Papers als onderzoeksbron te raadplegen, weten we dat de meerderheid van de bemanning op Nederlandse schepen uit het buitenland kwam."

Virtual Research

Om al het materiaal nog toegankelijker te maken voor onderzoekers en geïnteresseerden, heeft het Huygens ING een subsidie ontvangen voor de ontwikkeling van een *Virtual Research Environment* (VRE) van de Samenwerkende Maritieme Fondsen. Samen met de afdeling Digitale Infrastructuur van het KNAW Humanities Cluster wordt deze VRE ontwikkeld. Op de VRE kunnen bezoekers elk document tot in detail bekijken. Ook biedt de VRE een vrije *text search* op basis van automatische handschriftherkenning (HTR - *Handwritten Text Recognition*) en automatische herkenning van gedrukte tekst (OCR - *Optical Character Recognition*). Zowel de HTR als de OCR zijn nog in ontwikkeling en worden de komende jaren verbeterd.

prizepapers.huygens.knaw.nl

Voorbeeld van gebruik van de ODISSEI Secure Supercomputer

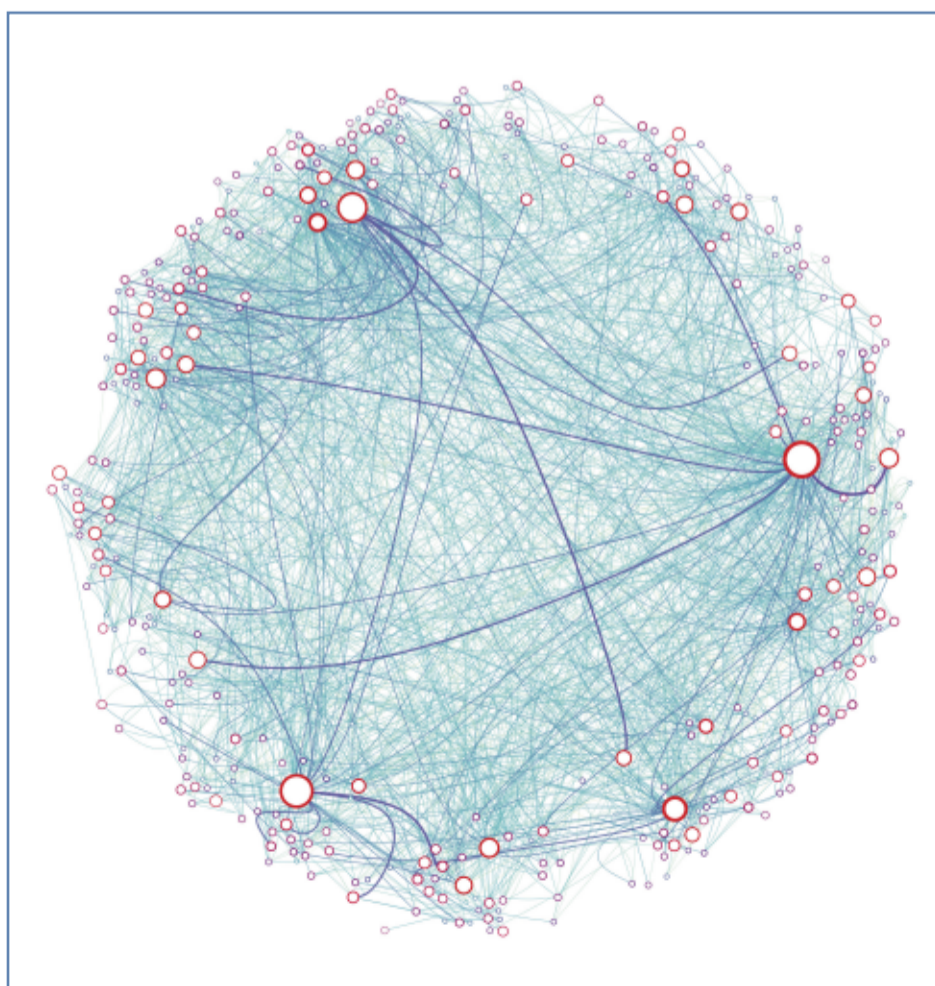
Sociale netwerkanalyse met 1 miljard links

Marjolijn Das, Edwin de Jonge en Jan van der Laan doen met behulp van de ODISSEI Secure Supercomputer onderzoek naar sociale netwerken van inwoners in Nederland. Evi de Cock

Van iedereen die in Nederland staat ingeschreven, heeft het CBS data beschikbaar. Uit deze data kunnen familiebanden, collegiale banden en relaties via school en buurt afgeleid worden die tezamen een netwerk vormen. Het resultaat is een omvangrijke, complexe dataset met een schat aan (geanonimiseerde) informatie, klaar voor vervolgonderzoek. De analyses vergen state-of-the-art technische voorzieningen. De ODISSEI Secure Super Computer (OSSC) biedt deze mogelijkheid met behulp van de high-performance computer van SURFsara.

Segregatie-index

Eén onderzoek richt zich op segregatie, hiervoor worden segregatie-indicatoren per persoon afgeleid. "Hierbij zijn 17 miljoen (geanonimiseerde) inwoners van Nederland met elkaar verbonden met 1 miljard links. De afbeelding toont een voorbeeld van dergelijke links: de intensiteit van eerste-orde familieconnecties tussen verschillende gemeenten in Nederland. Voor het uitrekenen van de segregatie-index op persoonsniveau was een computer met 250 GB aan werkgeheugen nodig.



Het netwerk verbindt (de lijnen) ouders, kinderen, broers en zussen die in verschillende gemeenten woonden in 2014 (de knopen). Het is slechts een voorbeeld van de schat aan informatie in een netwerk. Bron: CBS, bewerking door ODISSEI

De OSSC biedt deze rekenkracht en voldoende geheugen, en ook een veilige omgeving voor het analyseren van potentieel privacygevoelige data," aldus Edwin de Jonge (CBS).

Marjolijn Das (CBS, Erasmus Universiteit en LDE Centre for BOLD Cities): "Eén van onze vragen is hoe segregatie zich door de tijd ontwikkelt: neemt die toe of af in Nederland? Dit longitudinale aspect van sociale netwerkanalyses is niet eerder onderzocht op een dergelijk grote schaal. Een volgende vraag kan zijn hoe segregatie op persoonlijk niveau samenhangt met individuele kenmerken van personen, zoals maatschappelijk succes."

Vervolgonderzoek

De Jonge: "We hebben verschillende plannen voor vervolgonderzoek. We willen de segregatie-index verfijnen. Ook willen we één of meerdere van die indexen gebruiken voor het detecteren van netwerkcommunities en kijken naar de dynamiek van het netwerk: hoe verandert de *social fabric* van Nederland?"

Op dit moment maken meerdere onderzoeken gebruik van de OSSC. Na succesvolle afronding zal de OSSC verder worden opgeschaald en opengesteld voor de bredere onderzoeksgemeenschap.

odissei-data.nl/nl/odissei-secure-supercomputer

Alle zeventiende-eeuwse kranten in Delpher getranscribeerd

Oud nieuws voor nieuw onderzoek

De transcriptie van een grote hoeveelheid kranten maakt het mogelijk om taal- en cultuurhistorische veranderingen in de Gouden Eeuw grootschalig te onderzoeken.

Nicoline van der Sijs

Naar de ontwikkelingen van de Nederlandse Gouden Eeuw wordt veel onderzoek gedaan. Drukwerk uit de zeventiende eeuw vormt daarbij een belangrijke bron. En alhoewel er veel drukwerk is gepubliceerd en bewaard, bestond er tot nu toe geen aaneengesloten digitaal tekstcorpus waarmee taal- en cultuurhistorische veranderingen systematisch konden worden onderzocht.

20 miljoen woorden

Maar er is goed nieuws. Vrijwilligers hebben alle 17e-eeuwse kranten op Delpher - lopend van 1618 tot 1700 - getranscribeerd. Hiermee is het beschikbare digitale onderzoekscorpus van de zeventiende eeuw in één klap bijna verdubbeld. Het krantencorpus bestaat uit 6.184 verschil-

lende kranten die samen een kleine 20 miljoen woorden bevatten. Ter vergelijking: de DBNL-teksten voor deze eeuw bestaan uit circa 24 miljoen woorden. Door het transcriberen kunnen krantenteksten voor het eerst systematisch met de computer worden onderzocht. Tot nu toe was dat niet mogelijk omdat de optische tekenherkenning waarmee de teksten op Delpher waren gelezen, niet overweg kon met het gotische schrift en Oudnederlands. Medio 2020 komt het getranscribeerde krantencorpus beschikbaar via Delpher.

Verrijking van data

Het Meertens Instituut werkt aan het verder verrijken van de digitale tekstbestanden. Zo worden de metadata opgeschoond en uitgebreid en worden afzonderlijke artikelen semi-automatisch afgesplitst en voorzien van informatie over de tekstsoort (zoals advertentie, binnenlands nieuws, officiële mededeling). Ook de geografische namen die in de krantenkoppen voorkomen, worden verrijkt en benut: aan iedere naam wordt de moderne spelling toegevoegd. Die moderne schrijfwijzen kunnen vervolgens worden ingevoerd in een kaartprogramma, dat week voor week visualiseert waar

het nieuws binnen en buiten Europa vandaan kwam, en hoe de geografische focus in de loop van de eeuw veranderde.

Lacunes in kennis

De opgeschoonde en verrijkte krantenteksten komen in 2020 ook beschikbaar via een aparte interface. Dan kan iedereen zijn eigen onderzoeksvragen stellen, bijvoorbeeld naar maatschappelijke veranderingen of veranderingen in het taalgebruik. De teksten kunnen allerlei lacunes in kennis en gegevensbronnen aanvullen: zo ontdekte het Meertens Instituut al dat kranten een groot aantal woorden en spellingen bevatten die ontbreken in de bestaande historische lexica van het Nederlands. Het krantencorpus kan een proeftuin worden voor het testen van tools en modellen, zoals semantische vectoren en *topic modelling*. En de liefhebber kan natuurlijk ook gewoon het laatste nieuws van een bepaalde datum lezen.

meertens.knaw.nl

Nicoline van der Sijs is projectleider bij het Meertens Instituut. Heeft u vragen of suggesties, of wilt u meewerken aan dit project? Neem dan contact op: post@nicolinevdsijs.nl.



In plaats van woord voor woord lezen, kunnen onderzoekers de computer de getranscribeerde krantenteksten laten doorzoeken.

Credits: Detail uit Amsterdamsche Courant, 1684, via delpher.nl

Artificiële Intelligentie achter de Liederbank

Melodiegelijkenissen opsporen met algoritme

In de Nederlandse Liederbank, waarin ruim 175.000 Nederlandse liederen zijn ontsloten, kun je zoeken naar melodiegelijkenissen. Peter van Kranenburg ontwikkelde het algoritme achter deze functie.

Mathilde Jansen

Wie in de Nederlandse Liederbank 'Elf november is de dag' intypt, komt via 'vergelijkbare melodieën' terecht bij 'Daar was laatst een meisje loos'. Die mogelijkheid om naar melodiegelijkenissen te zoeken, is de verdienste van Peter van Kranenburg. Hij is computationeel musicoloog aan het Meertens Instituut en onderzoekt muziek aan de hand van computermodellen.

In 2010 promoveerde hij op een uitlijningsalgoritme. "Dat schrijft de melodieën zo onder elkaar, dat de overeenkomende noten precies onder elkaar komen te staan", legt Van Kranenburg uit. "Het algoritme zoekt uit op welke plekken ruimte toegevoegd moet worden, zodat de corresponderende delen onder elkaar staan. Hoe meer ruimte, hoe slechter de gelijkheid. Grof gezegd."

Tune families

Als je een uitlijning maakt van een query-melodie met alle melodieën uit de Liederbank, en die sorteert, dan komen de meest gelijkende melodieën bovenaan. "Net als Google-resultaten", verduidelijkt Van Kra-

In de Nederlandse Liederbank is het mogelijk om naar melodiegelijkenissen te zoeken. Een van de zoekresultaten bij 'Elf november is de dag' is 'Daar was laatst een meisje loos'. Credits: Meertens Instituut

nenburg. Zo zie je welke melodieën varianten zijn van elkaar, en kun je ze onderverdelen in families, ook wel 'tune families' genoemd. Tegenwoordig is veel kunstmatige intelligen-

tie gebaseerd op neurale netwerken. Daarom onderzocht Van Kranenburg samen met collega's van het Meertens Instituut en de Universiteit Antwerpen of dit ook werkte bij melo-

JONG TALENT

'Een uitlijningsalgoritme schrijft de melodiën en overeenkomende noten precies onder elkaar'



diegelijkenissen. "Om het neurale netwerk te trainen, werden steeds twee melodieën aangeboden die wel op elkaar lijken en twee die niet op elkaar lijken. Als je dat lang genoeg doet, met heel veel verschillende melodieën, in ons geval zo'n zesduizend, dan hoop je dat zo'n netwerk op een gegeven moment leert wat het betekent dat twee melodieën op elkaar lijken." En dat lukte. Het model vond melodiegelijkenissen met een betrouwbaarheid van 70 tot 80 procent. Iets beter dan het uitlijningsalgoritme. "Nog geen grote verbetering, maar het laat wel zien dat het model werkt. En dat biedt perspectief voor de toekomst en vormt nieuwe uitdagingen. Want het neurale netwerkmodel is misschien wel intelligent, voor mensen is het soms moeilijk te interpreteren wat het allemaal doet. Het begrijpelijk maken van die netwerken is een belangrijk onderzoeksgebied. Daar willen we in een volgende stap aan bijdragen door te onderzoeken wat ons netwerk geleerd heeft over melodische gelijkheid."

liederbank.nl

Peter van Kranenburg

Van Kranenburg studeerde Musicologie aan de Universiteit Utrecht en Electrical Engineering aan de TU Delft. Hij promoveerde in 2010 aan de Universiteit Utrecht. Hij werkt als computationeel musicoloog bij het Meertens Instituut en de Universiteit Utrecht.

CLARIAH-project Digital Film Listings

Reconstructie van oude filmpladders

Vroeger stond in de krant waar en wanneer bioscoopfilms draaiden. Onlangs is de informatie uit deze zogenoemde Filmpladders uit kranten van 1948 tot 1994 digitaal gereconstrueerd.

Kaspar Beelen

Het CLARIAH-project Digital Film Listings (DIGIFIL) spitst zich toe op het automatisch extraheren, digitaliseren en publiceren van de informatie uit Filmpladders. Het project demonstreert hoe machinegetranscribeerde tekst, geproduceerd met behulp van *Optical Character Recognition* (OCR), op een automatische manier kan worden geconverteerd naar verrijkte en gestructureerde data, voorzien van semantische annotaties.

Verrijkt en gelinkt

De eerste fase bestond uit een klassiek 'naald in een hooiberg'-probleem: de Filmpladders vissen uit de gigantische stapel artikelen in de Delpher-krantendatabank. Vervolgens maakte *machine learning* de impliciete structuur van de pladders expliciet: elk woord in de pladders werd op basis van de context voorzien van een label ('titel', 'tijdstop', 'bioscoop'). De

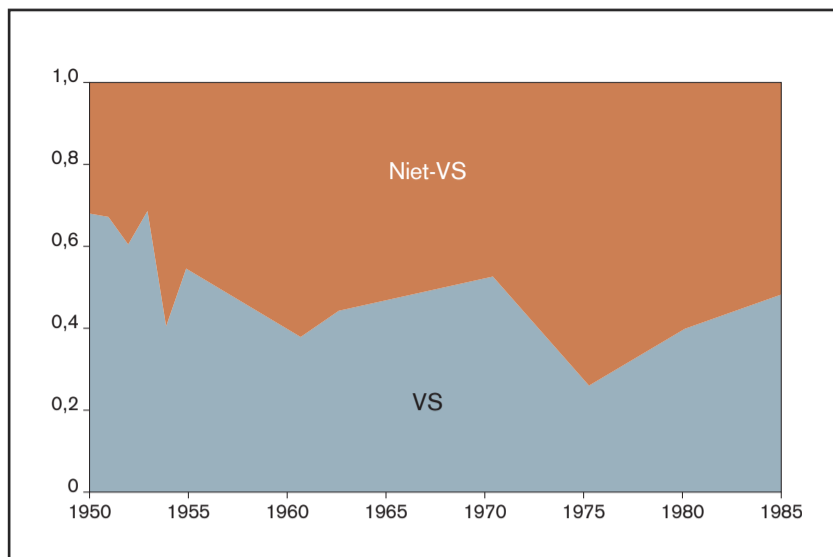
geïdentificeerde filmtitels werden daarna gekoppeld aan externe databanken, zoals de Internet Movie Database (IMDb). Met de verrijkte en gelinkte DIGIFIL-data is het mogelijk om ontwikkelingen in het naoorlogse filmpladderschap kwantitatief te onderzoeken, en de patronen te vergelijken met kwalitatief onderzoek (zie afbeelding).

Grotere ambities

DIGIFIL richt zich op de extractie van één type micro-evenementen, namelijk filmvertoningen, maar de onderzoekers koesteren grotere ambities. De tools kunnen misschien ook worden toegepast voor de extractie van andere soorten informatie, zoals theateragenda's of scheepsberichten. In die zin poogt DIGIFIL om alledaagse geschiedenis en digitale geesteswetenschappen met elkaar te verbinden.

De onderzoekers binnen het DIGIFIL-project zijn: Kaspar Beelen, Ivan Kisjes, Thunnis van Oort, Kathleen Lotze en Julia Noordegraaf.

Op Gitlab staan diverse scripts: gitlab.com/uvacreate/digifil/



De afbeelding toont het aandeel Hollywoodfilms (blauw) versus niet-in-Amerika-geproduceerde films (bruin) in de periode 1948 tot 1994 in Nederlandse bioscopen. De data laten een geleidelijke afname zien van het marktaandeel van Amerikaanse films, gedefinieerd in termen van het aantal vertoningen. Dit is in strijd met de bevindingen van Bart Hofstede, die in zijn proefschrift over de mondiale positie van de Nederlandse film uit 2000 een andere dynamiek waarnam: een relatieve daling van het Amerikaanse marktaandeel in de jaren 1960 en 1970, gevolgd door een sterke stijging in de jaren 1980 tot een aandeel van meer dan 80%. Of er werkelijk sprake is van een nieuw inzicht zal vervolgonderzoek moeten uitwijzen.



Voorbeeld van een filmpladder, *Algemeen Handelsblad*, 14-03-1952

Credits: Delpher

GELEZEN

LCRDM positioning Paper 2019 en verder
LCRDM

De adviesgroep van het Landelijke Coördinatiepunt Research Data Management (LCRDM) heeft een Positioning Paper 2019 en verder opgesteld. In dit paper wordt - aan de hand van een thematische prioritering en drie bredere (beleidsmatige) werkgebieden - beschreven wat wel en niet binnen de scope van het LCRDM valt.

Belangrijk is dat sinds 2018 het LCRDM werkt met een pool van experts. Deze pool is inmiddels uitgegroeid tot ruim 190 deelnemers uit 60 Nederlandse onderzoeksinstituten.

De experts werken in taakgroepen aan diverse aspecten van Research Data Management uit de praktijk van Nederlandse onderzoeksinstituten. De samenwerking in de taakgroepen plaatst de activiteiten binnen de instellingen in een breder landelijk perspectief. Dit zorgt voor samenhang, herkenbaarheid en onderbouwing. De paper en resultaten van werkgroepen staan op de LCRDM-website.

[LCRDM.nl](https://www.lcrdm.nl)

COLUMN

Micro-data voor macro-onderzoek

In 2011 verhuisde ik na een aantal jaar werken in het Verenigd Koninkrijk (VK) terug naar Nederland. Voor mijn onderzoek in het VK maakte ik veel gebruik van de fantastische Britse longitudinale survey data en de longitudinale data van de volkstellingen. In het VK is er nog iedere tien jaar een echte volkstelling en voor een kleine steekproef worden de individuele bestanden aan elkaar gekoppeld. Daardoor is het mogelijk om mensen door de tijd te volgen met iedere tien jaar een meetopname van het leven van mensen; niet ideaal, maar als je vier volkstellingen aan elkaar koppelt, kun je mensen toch al dertig jaar volgen.

Enmaal terug in Nederland stuitte ik op de microdata van het Centraal Bureau voor de Statistiek. De Britse data was al mooi, maar de Nederlandse microdata is het ware paradijs voor sociaalwetenschappers: individuele longitudinale data van de hele Nederlandse bevolking vanaf eind jaren '90; geen steekproef,

maar gewoon iedereen. De Nederlandse microdata is zeer vergelijkbaar met wat er beschikbaar is voor onderzoek in Zweden en Finland waardoor (in theorie) vergelijkbaar onderzoek mogelijk is.

Microdata is registerdata op basis van de Basisregistratie



Personen. Een uniek kenmerk van deze data is de mogelijkheid om surveydata te koppelen aan de CBS registerdata. Bijvoorbeeld: informatie uit een survey over zoekgedrag naar werk kan worden gekoppeld aan de microdata met informatie of mensen werk hebben gevonden, ook jaren na het afnemen van de survey. Het slim linken van survey data aan registerdata heeft enorme potentie voor onderzoek.

Met de enorme schat aan gegevens op individueel niveau kunnen onderzoekers onder zeer strikte voorwaarden zelf onderzoek doen. Hiermee ontstaan ook nieuwe uitdagingen, zoals het waarborgen van privacy versus de enorme maatschappelijke baten

van onderzoek.

De schat aan data stelt ook nieuwe eisen aan de hardware en software die sociaalwetenschappers nodig hebben om hun onderzoek te kunnen doen. Waar het gebruik van een supercomputer het domein was van klimaatonderzoekers en de kwantummechanica, ook sociaalwetenschappers hebben in toenemende mate grote rekenkracht nodig.

Erder dit jaar had ik de eer om samen met een van mijn promovendi binnen een pilot van de ODISSEI Data Facility te werken met de high-performance computer van SURFsara. Deze supercomputer heeft enorme potentie voor de sociale wetenschap-

pen. Ik hoop dan ook zeer dat er financiering beschikbaar komt voor het verder ontwikkelen en opschalen van ODISSEI (Open Data Infrastructure for Social Science and Economic Innovations) waardoor Nederland zich kan ontwikkelen tot de absolute wereldtop qua data infrastructuur voor de sociale wetenschappen.

Maarten van Ham

Maarten geeft de volgende column graag aan Sander Steijn van het Sociaal en Cultureel Planbureau.

Maarten van Ham is hoogleraar stadsgeografie aan de Faculteit Bouwkunde van de Technische Universiteit Delft. Hij promoveerde als geograaf in Utrecht en werd in 2011 hoogleraar aan de universiteit van St Andrews in het Verenigd Koninkrijk. In 2014 ontving hij een European Research Council (ERC) subsidie voor onderzoek naar de oorzaken en gevolgen van ruimtelijke ongelijkheid in Nederland, Zweden, Estland en het Verenigd Koninkrijk.