

Afelonne Doek, directeur Collecties en Digitale Infrastructuur IISG:

‘Erfgoedinstellingen in een spagaat’

Je zou zeggen dat er anno 2017 goede afspraken te maken zijn over het tonen van gedigitaliseerde erfgoedcollecties.

De praktijk wijst anders uit.

Machteld Maris

Rechtszaken inzake auteursrechtinbreuk maken duidelijk dat de wetgeving niet goed aansluit bij de manier waarop erfgoedinstellingen hun collecties online beschikbaar maken. Zo werd het Internationaal Instituut voor Sociale Geschiedenis (IISG) in 2014 door fotograaf Kors van Bennekom voor de rechter gedaagd wegens het ongeraagd online tonen van 221 van zijn foto's die zich in de collecties van het IISG bevonden. Ondanks het feit dat ze op klein formaat, in lage kwaliteit, met volledige naams- en bronvermelding getoond werden. Van Bennekom eiste 50.000 euro schadevergoeding.

Visuele metadata

Afelonne Doek, directeur Collecties en Digitale Infrastructuur van het IISG: “Het IISG stelde zich destijds op het standpunt dat het de beelden toonde als erfgoedinstelling en archief, in de context van de collecties, als visuele metadata en als uitdrukking van een historische gebeurtenis. Het is niet zo dat we posters van het werk van Van Bennekom verkopen in een webshop en daaraan verdienen. Het IISG is een onderzoeksinstituut. Van 270.000 beelden uit de collectie is een digitale representatie beschikbaar. Van het overgrote deel van deze beelden is geen maker of auteursrechtelijke bekend. Het regelen van toestemming vooraf is zeer arbeidsintensief en vrijwel ondoenlijk.”

Schadeclaim afgewezen

De rechtbank stelde weliswaar auteursrechtinbreuk vast, maar wees de schadeclaim van de fotograaf af. Het gevraagde bedrag stond in geen verhouding tot de mogelijk geleden schade. Slechts zeven van de 221 foto's waren in totaal 49 keer bekeken, één keer was er een scan tegen on-



Illustratie Auke Herrema

kostenvergoeding opgevraagd. Om de collectie voor onderzoek beschikbaar te kunnen maken, heeft het IISG een overeenkomst gesloten met Pictoright, een collectieve beheersorganisatie (CBO), die namens beeldmakers optreedt en auteursrechtelijke belangen behartigt. Tegen betaling van een bedrag (hoe meer beelden, hoe hoger het bedrag) verleent Pictoright erfgoedinstellingen het recht om beelden online te tonen onder vooraf bepaalde strikte voorwaarden. Doek: “Maar met een collectieve beheersovereenkomst ben je er niet. Het tonen van beelden zonder toestemming is nog steeds in strijd met de auteurswet, er kunnen altijd rechthebbenden opduiken die bezwaar maken tegen het gebruik van hun werk. Erfgoedinstellingen zijn terughoudend bij het beschikbaarstellen en zetten (delen van) hun beeldbank op zwart. Dat kan toch niet de bedoeling zijn in een tijd waarin de beschikbaarheid van digitale collecties dé nationale ambitie is?”

Opt-outregeling

Het wachten is op de invoering van een *extended collective licensing* (ECL). Een ECL is een rechtsfiguur waarbij het mandaat van een CBO kan worden uitgebreid tot rechthebbenden die niet zijn aangesloten. Zij komen in aanmerking voor

een licentie. Rechthebbenden moeten altijd een opt-outregeling kunnen gebruiken. En er zou zoiets moeten zijn als een toetsing van de representativiteit van aangesloten leden bij een CBO door een College van Toezicht. Inkomsten die niet ten goede komen aan rechthebbenden zouden idealiter moeten terugvloeien naar de erfgoedinstellingen ten behoeve van digitaliseringsprojecten.

Zonder resultaat

Doek benadrukt dat een wettelijk ECL niet zaligmakend is. Maar het zou voor erfgoedinstellingen en CBO's wel makkelijker zijn om hun taak uit te kunnen voeren. Rechthebbende partijen of personen die beslissen over beschikbaarstelling, staan bovendien duidelijker in hun recht. De Federatie Auteursrechtbelangen heeft, samen met erfgoedinstellingen, al in 2013 bij de staatssecretaris van Veiligheid en Justitie en de minister van Onderwijs, Cultuur en Wetenschap aangedrongen op een wettelijke invoering. Tot op heden zonder resultaat.

socialhistory.org/en/node/5146

(over de rechtzaak en het vonnis)

kl.nl/themas/auteursrecht

(over dossier Auteursrecht bij Kennisland)

Ingrid Dillo, interim secretaris-generaal van de RDA:

‘Bouwstenen klaar voor gebruik’

Vijf jaar na de start telt de Research Data Alliance (RDA) maar liefst 6.000 leden uit 128 landen. Deze community kwam al tot veel bruikbare bouwstenen, online beschikbaar voor iedereen.

Harmen van der Meulen

De RDA wil het delen en hergebruiken van onderzoeksdata faciliteren. Hiertoe worden producten ontwikkeld, bijvoorbeeld tools op het gebied van researchdatamanagement. Maar ook softwarecodes en beleids-

standaarden zijn voorbeelden van bouwstenen ontwikkeld door de RDA-community, beschikbaar voor iedereen die deze elementen kan gebruiken.

De RDA brengt data-experts bij elkaar: twee keer per jaar is er een werkconferentie met plenaire en parallelsessies voor de inmiddels ruim 80 werk- en interessegroepen. Ingrid Dillo, adjunct-directeur / directeur Beleid bij DANS en interim secretaris-generaal van de RDA:

Vervolg op pagina 2

CoreTrustSeal: nieuwe certificerings-standaard

Data Seal of Approval (DSA) en World Data System (WDS) van ICSU introduceren: CoreTrustSeal (CTS), een non-profitorganisatie ter bevordering van duurzame en betrouwbare data-infrastructuren. CTS certificeert data-opslagplaatsen op basis van de catalogus en procedures van de DSA WDS Core Trustworthy Data Repositories Requirements. Hiermee vervangt CTS zowel DSA- als WDS Regular Members-certificering. Mustapha Mokrane, voorzitter van de Standards and Certification Board van CoreTrustSeal.

‘Wij leveren professionele certificeringstools en -diensten aan data-opslagplaatsen en bieden onze gekwalificeerde reviewers optimale omstandigheden om hun beoordeling uit te voeren.’
coretrustseal.org



E-DATA & RESEARCH

Jaargang 12 | 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, Huygens ING, de Koninklijke Bibliotheek en het Rijksmuseum.

INHOUD

2 Verslagen van events in Gehoord en bijgewoond



3 Man-met-een-missie Mons over GO FAIR

4 Deze datasets zijn ook sinds kort beschikbaar

5 Eén database met duizenden beroepstitels

6 Spraakherkenning in een stroomversnelling

6 Een sticker voor het delen van data

7 K-PLEX vraagt om hulp: small data onder de loep

8 Clichés en stereotypen leveren veel data op

E-data & Research 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.
www.edata.nl

Man-met-een-missie Barend Mons:

‘GO FAIR laat data, tools en rekenkracht elkaar treffen’

Findable, Accessible, Interoperable en Reusable: vier principes om op een verantwoorde manier met wetenschappelijke data om te gaan. Aan het woord is Barend Mons, een van de grondleggers van FAIR.

Steven Claeysens

Barend Mons is een man-met-een-missie. “De wijze waarop de wetenschap vandaag de dag omgaat met de enorme hoeveelheden data moet anders. Als je belastinggeld aanneemt voor onderzoek, zorg dan ook goed voor de data die dat oplevert en maak hergebruik door anderen mogelijk. De FAIR-principes zijn er om hierbij te helpen.”

De essentie van FAIR

FAIR staat voor Findable, Accessible, Interoperable en Reusable. FAIR richt zich op machines, want als informatie *machine readable* en *machine actionable* is, dan kan je het ook altijd aanbieden in een voor mensen leesbare, gestructureerde vorm. Mons: “Wil je bijvoorbeeld verwijzen naar een wetenschappelijk standaardbegrip, gebruik dan geen afkorting die tientallen betekenissen kan hebben, maar gebruik een URI of een andere unieke, persistente resolvable identifier. Zo weet de computer waarover je het hebt. Dát is de essentie van FAIR. Andersom, als iets begraven ligt in een tekst dan kan je die tekst wel minen, maar je vindt nooit alles terug. Tekst is een nachtmerrie voor machines en per definitie niet FAIR, zelfs als de tekst als zodanig Findable en Accessible is. Mijn voorstel is dan ook om binnen de wetenschap af te stappen van het eerst publiceren voor de mens. Laten we zaken eerst FAIR publiceren, zodat computers het begrijpen, en pas in tweede instantie ook leesbaar maken voor de mens.”

Internet of FAIR Data and Services

“Data, tools en rekenkracht moeten elkaar op de juiste plek treffen. De tijd van grote, centrale datawarehouses, waar alle data verzameld worden om te analyseren, is echter voorbij. Dat is contra-intuïtief voor veel mensen, maar het is wel de toekomst van big data. In een ideale situatie zijn alle data en de metadata van de workflows FAIR, weet de computer welke workflows op welke data kunnen draaien, en vermelden de metadata van de workflows hoeveel rekenkracht nodig is. Met zo’n infrastructuur kunnen vrijwel alle wetenschappelijke vragen beantwoord worden op een gefedereerde wijze, bijvoorbeeld met virtual machines die de data opzoeken. Het systeem dat die data, workflows en rekenkracht bij elkaar brengt, is het toekomstige Internet of FAIR Data and Services. De European Open Science Cloud is een belangrijk onderdeel van deze toekomstige infrastructuur, maar die kan niet uitsluitend Europees zijn, kan niet altijd open zijn, is niet enkel voor de wetenschap en is geen cloud.”

GO FAIR

“Maar hoe zorgen we ervoor dat zo’n infrastructuur er komt, en blijft? Dat bestaande kennis en kwaliteiten worden hergebruikt? En dat er structurele financiering komt, geen vierjarige geldinjecties? Want iedereen, inclusief NWO, wil investeren in *rocket science*, en bijna niemand in de *rocket launcher*. We gebruiken hiervoor de benadering GO FAIR. Iedereen die een component van het Internet of FAIR Data and Services wil in-



‘De computer laten weten waar je het over hebt, dat is de essentie van FAIR.’ foto Mariët Mons

INTERVIEW

‘Iedereen wil investeren in rocket science, bijna niemand in de rocket launcher’

een bedrijf, een individu of een heel land is, heeft dus de vrijheid om bijvoorbeeld een eigen ontologie te ontwikkelen. Het enige wat we de netwerken vragen, is om zich te houden aan de FAIR-principes en te tekenen tegen *vendor lock-in*. Verder is GO FAIR volledig bottom-up opgezet en dat blijft het.”

Internationaal support office

Ook de politiek heeft GO FAIR intussen omarmd. Staatssecretaris van Onderwijs, Cultuur en Wetenschap Sander Dekker en zijn Duitse collega jagen het in Europa actief aan en meer landen melden zich al. “We hebben nu de mogelijkheid om het voortouw te nemen. Daarom komt er, naast nationale support

brengen, kan bijdragen door een implementatienetwerk op te zetten. We herhalen daarmee de truc waarmee het internet zelf is ontstaan: we definiëren minimale standaarden waar iedereen zich aan hoort te houden, daarbuiten garanderen we maximale vrijheid voor implementatie. Elk implementatienetwerk, of het een instituut,

teams, een internationaal support office in Leiden.

Door het huidige systeem verbieden we jonge onderzoekers de facto om bijvoorbeeld nanopublicaties of RDF te publiceren. Ook hebben met name ICT en de andere wetenschapsdomeinen te weinig kennis van elkaars domein. Uitgevers, onderzoekers en subsidiegevers houden elkaar in een dodelijke omarming en houden zo een pervers systeem in stand. Het is de cultuur die ons tegenhoudt, niet de techniek. Laten we dat samen doorbreken!”

www.dtls.nl/fair-data/go-fair

Over Barend Mons

Barend Mons is hoogleraar Biosemantiek aan het Leids Universitair Medisch Centrum en verbonden aan het Leiden Centre of Data Science. Naast initiatiefnemer van de FAIR-principes was Mons voorzitter van de High Level Expert Group van de European Open Science Cloud. Ook is hij Scientific Lead DTL Data binnen het Dutch Techcentre for Life Science en E-science integrator bij NLeSC. Vanuit al deze affiliaties trekt hij, samen met Erik Fledderus, directeur van SURF, aan het GO FAIR-initiatief.

SINDS KORT BESCHIKBAAR

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

CentERdata

• De meest aantrekkelijke loterij

In april 2016 is het onderzoek Lottery Decisions afgenomen onder een deel van het LISS panel. Dit onderzoek bestaat uit een experiment met loterij-achtige situaties. Het onderzoek bestudeert hoe mensen keuzes maken. Tijdens het onderzoek werd men gevraagd om 26 keuzes te maken tussen loterijen met verschillende kansen op verschillende bedragen. Een willekeurige selectie van de deelnemers kreeg het bedrag van een van hun keuzes daadwerkelijk uitbetaald. Het onderzoek werd uitgevoerd door de onderzoekers Charles Bellemare (University of Laval, Canada), Alexander Sebald (University of Copenhagen) en Joachim Winter (LMU München). De data zijn beschikbaar via het LISS Data Archive.

lissdata.nl/dataarchive

Ook sinds kort beschikbaar:

Studies LISS panel

- Brouwers, L., maart 2016, Norm questionnaire NAZB
- CentERdata, juni 2016 - juli 2016, Economic Situation: Housing - Wave 9
- CentERdata, juni 2016 - juli 2016, Economic Situation: Income - Wave 9
- CentERdata, juli 2016 - augustus 2016, Economic Situation: Assets - Wave 5
- CentERdata, oktober 2016 - november 2016, Social Integration and Leisure - Wave 9
- CentERdata, mei 2017 - juni 2017, Personality - Wave 9
- CentERdata, juni 2017 - juli 2017, Economic Situation: Income - Wave 10



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

DANS

• Rookvrije schoolterreinen

Veel mensen beginnen gedurende hun schooltijd met roken. In opdracht van het Mulier Instituut heeft DUO Onderwijsonderzoek in het najaar van 2016 onderzoek gedaan op het gebied van rookvrije schoolterreinen. In dit onderzoek zijn drie verschillende deelonderzoeken uitgevoerd naar het primaire onderwijs, voortgezet onderwijs en MBO door middel van online enquêtes.



Veel mensen beginnen in hun schooltijd met roken. foto Shutterstock



Prize Papers credits Erik van der Doe

Gevraagd werd onder andere naar de communicatie over het rookverbod, de redenen om het plein rookvrij te maken en de mate van stiekem roken op een rookvrij schoolterrein. Het onderzoek is toegankelijk via EASY.

doi.org/10.17026/dans-z7v-t4ds

Ook sinds kort beschikbaar:

- Chen, MSc Z.; Veling, dr. H.P.; Dijksterhuis, prof. dr. A.J. en Holland, prof. dr. R.W. (allen RU Radboud Universiteit) (2016): How Does Not Responding to Appetitive Stimuli Cause Devaluation: Evaluative Conditioning or Response Inhibition?. DANS. <https://doi.org/10.17026/dans-xfv-at9w>
- Kadaster (2017): Kadastrale kaart - versie februari 2017. DANS. <https://doi.org/10.17026/dans-zwc-8g2y>
- Kastelein, MA D. (Gemeente Zutphen, Afd. Stadsbedrijven, team Archeologie) (2015): De eerste Sjoel. DANS. <https://doi.org/10.17026/dans-zgk-3aru>
- Kimenai, drs P. (BAAC bv) (2016): De Sint-Clemenskerk. Multidisciplinair onderzoek naar de geschiedenis van de middeleeuwse parochiekerk in Oud-Merkelbeek. DANS. <https://doi.org/10.17026/dans-zwg-9qz2>
- Samur, D. (VU University Amsterdam) (2017): Four Replication Experiments of Kidd and Castano (2013). DANS. <https://doi.org/10.17026/dans-zx6-h97m>

- Vahdat, M. (Eindhoven University of Technology); Carvalho, M.B. (Tilburg University); Funk, M. (Eindhoven University of Technology); Rauterberg, M. (Eindhoven University of Technology); Hu, J. (Eindhoven University of Technology); Anguita, D. (University of Genoa) (2016): Learning analytics for Lix puzzle-game. DANS. <https://doi.org/10.17026/dans-xk7-8f5r>
- Wolters, Mr. Dr. P.T.J. (Radboud Universiteit. Onderzoekcentrum Onderneming & Recht) (2016): De matiging van boetedingen bij koopovereenkomsten ten aanzien van onroerende zaken. DANS. <https://doi.org/10.17026/dans-z7u-xrbj>



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

Huygens ING

• Prize Papers

Sinds kort bevinden zich 72.000 scans uit het Londense archief van de High Court of Admiralty bij het Huygens ING. Het gaat om geselecteerde Prize Papers, Nederlandse documenten die in de periode 1652-1815 door Britse kapers zijn buitgemaakt. Denk aan scheepsjournalen, ladingoverzichten, rekeningen, plantagelijsten en ondervragingen van bemanningsleden én ook zo'n 38.000 zakelijke en particuliere brieven. Een subsidie van Bureau Metamorfoze maakte het scannen mogelijk. Het Huygens ING gaat nu aan de slag met de beschikbaarstelling van de scans. Om het gedigitaliseerde, maar niet geordende materiaal, toegankelijk te maken zal door Huygens ING een Virtual Research Environment (VRE) ontwikkeld worden. De VRE wordt gefinancierd door Samenwerkende Maritieme Fondsen. Voor meer informatie: jelle.van.lottum@huygens.knaw.nl



Deze publicatie is beschikbaar via huygens.knaw.nl/tools-en-data/. Bezoek deze site of scan de QR-code.

AGENDA

19 - 21 oktober • Tokio

DSAA2017

De vierde internationale conferentie met als thema Data Science and Advanced Analytics.

dslab.it.aoyama.ac.jp/dsaa2017

23 - 29 oktober • overall

Open Access Week

Internationale week om meer bekendheid te genereren voor open access in onderzoek. Dit jaar is het thema 'Open in order to...'

openaccessweek.org

26 oktober • Utrecht

Workshop Koppelen van Data

Tijdens deze workshop, georganiseerd door het CBS en ODISSEI, worden ervaringen gedeeld over het koppelen van data.

odissei-data.nl

27 oktober - 1 november •

Washington DC

ASIS&T

Informatiewetenschappers en onderzoekers van over de hele wereld komen samen tijdens dit jaarlijkse event.

asis.org/am17

6 november • Manchester

Working with data on political behaviour

Deze gratis workshop verkent gegevens die beschikbaar zijn in Europa voor het onderzoeken van politiek gedrag door middel van gesprekken, discussies en een hands-on sessie. De workshop wordt georganiseerd door CESSDA ERIC.

cessda.eu

8 - 10 november • Wenen

Conference on Cultural Heritage and New Technologies

Tijdens dit event staat 'urban archaeology and Integration' centraal.

chnt.at

9 november • Amsterdam

EHRI Workshop

Engaging New Generations - The Holocaust and Knowledge Dissemination in the Digital Age.

ehri-project.eu

9 november • Nieuwegein

KNVI Jaarcongres

'Informatie = m8' is het thema van het congres voor informatieprofessionals georganiseerd door de vakvereniging KNVI.

congres.knvi.info

30 november • overall

IDPD

Tijdens de International Digital Preservation Day (IDPD) wordt aandacht besteed aan de digitale collecties die wereldwijd beheerd worden. Kijk op NCDD.nl voor een overzicht van de Nederlandse activiteiten en volg #IDPD17 op Twitter.

dpconline.org

5 - 6 december • Lausanne

EDDI17

Negende jaarlijkse conferentie voor DDI-gebruikers.

eddi-conferences.eu

26 januari • Nijmegen

CLIN28

De jaarlijkse conferentie over computationele taalkunde in het Nederlandse taalgebied.

ru.nl

App voor analyse van middeleeuwse handschriften in de maak

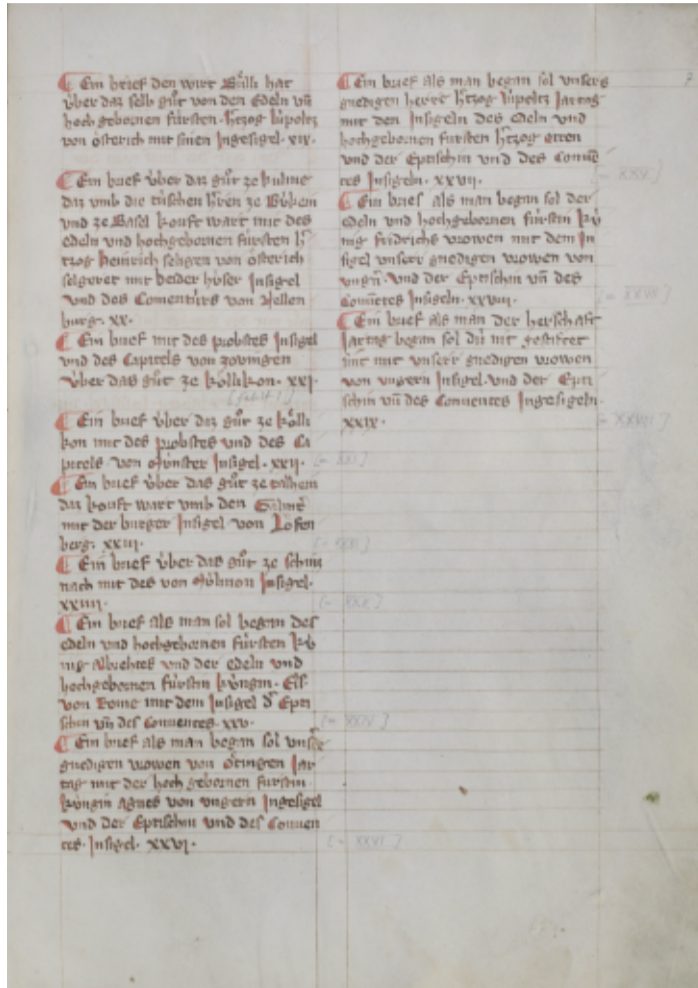
Graven in digitale data

Welke digitale tools zijn nodig voor handschriftenonderzoek? En welke vragen zou je met zo'n tool willen beantwoorden? Een onderzoeksgroep bestaande uit mediëvisten, softwareontwikkelaars, studenten, PhD's en postdocs buigt zich over deze vraag. *Mariken Teeuwen*

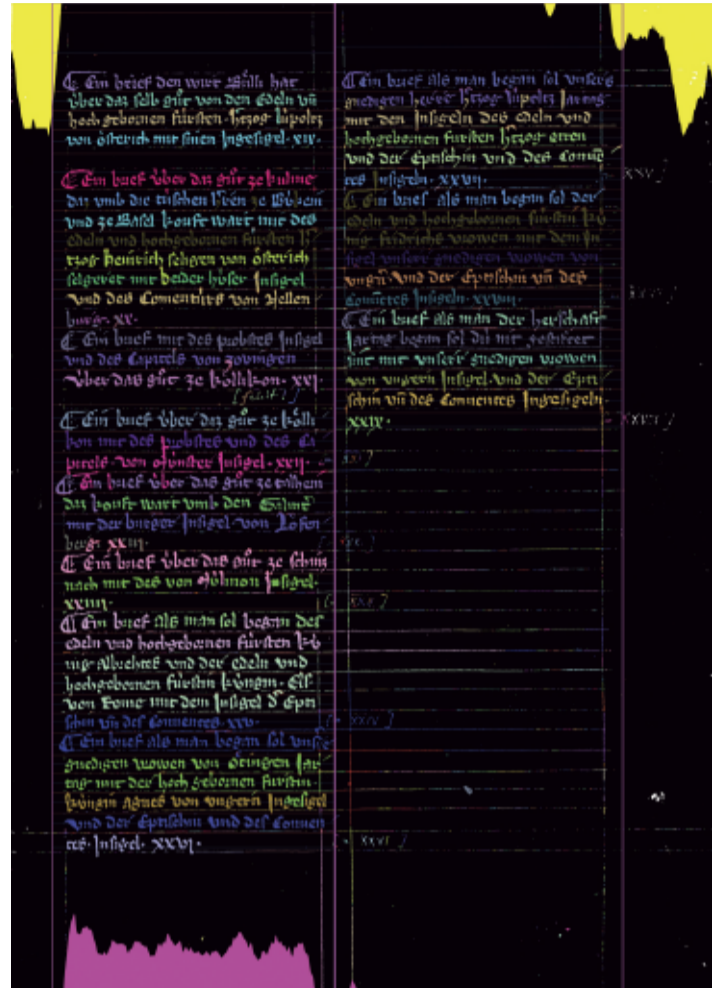
Bij Huygens ING is een denktank gestart voor ideeën en experimenten voor het ontwikkelen van digitale tools voor handschriftenonderzoek. Deze groep van onderzoekers en software engineers onderzoekt de vragen die we zouden willen (en kunnen) benaderen vanuit een digitaal perspectief, en wat we daarvoor zouden kunnen bouwen. Er is in ons vakgebied een aardverschuiving gaande. De digitalisering van handschriftencollecties was al wat langer aan de gang (sinds ca. 2000), maar steeds meer grote bibliotheken die verantwoordelijk zijn voor de opslag en ontsluiting van collecties middeleeuwse handschriften, omarmen het International Image Interoperability Framework (IIIF). Dit framework is een internationaal protocol voor het opslaan van digitale images.

Verschillende analyses

Er gaat een wereld voor ons open. Nu is het mogelijk om één viewer te bouwen waarmee handschriften uit landen van over de hele wereld naast elkaar kunnen worden getoond op één scherm (Mirador). En omdat de images vergezeld gaan van gestructureerde metadata (kenmerken zoals datum, inhoud, afmetingen, geschiedenis, etcetera), is het in principe dus ook mogelijk om met één handeling zoekopdrachten gelijktijdig uit te voeren in ver-



Links staat de originele afbeelding. Rechts staat het resultaat van de digitale toolbox: één image van waaruit handschriftenonderzoekers verschillende analyses automatisch kunnen uitvoeren. credits Aarau, Staatsarchiv Aargau, AA/0428, f. 7r – Cartulary I of Königsfelden (<http://www.e-codices.unifr.ch/en/list/one/saa/0428>)



schillende collecties. En is het mogelijk om dezelfde analyses uit te voeren op images uit verschillende collecties. Dit heeft het idee gevoed om een app te ontwikkelen waarmee de handschriftenonderzoeker vanuit één image verschillende analyses automatisch kan laten uitvoeren. De app heeft nog geen naam, maar laten we het voor nu The Digital Toolbox for the Manuscript Scholar noemen. Met deze Digital Toolbox willen we moge-

lijkheden gaan operationaliseren in één app. De onderzoeker kan de lay-out analyseren, de tekst automatisch laten transcriberen met behulp van Transkribus (een handschrift-OCR systeem dat ontwikkeld wordt met Europees geld) én de metadata zien die beschikbaar zijn in IIIF-collecties. Vervolgens heeft hij of zij mogelijkheden te over om te zoeken naar vergelijkingsmateriaal, naar lokalisering (de bepaling waar een manuscript oorspronkelijk vervaardigd is), naar datering (de bepaling

wanneer het vervaardigd is), etcetera. Kortom: het wordt een tool om snel en effectief naar materiaal te zoeken, en dat materiaal van historische, culturele en inhoudelijke context en interpretatie te voorzien. Wordt vervolgd!

Prof. Dr. Mariken Teeuwen is senior wetenschapper afdeling Wetenschapsgeschiedenis, Huygens ING.
huygens.knaw.nl

Kea Tijdens, coördinator WP8-SERISS-project:

‘Zelf beroepstitels kiezen’

Dankzij het SERISS-project is er een online database met duizenden gecodeerde beroepstitels beschikbaar, inclusief software en handleiding. De coördinator van deze taak, Kea Tijdens, vertelt.

Maurice Martens

“Bijna alle sociaal-economische surveys vragen respondenten naar beroep en opleiding. Het zijn belangrijke indicatoren voor de identiteit van een werkende, voor zijn sociaal-economische status. Ieder land heeft tienduizenden beroepstitels. Om titels tussen landen te kunnen vergelijken, moeten ze geclassificeerd zijn. Het achteraf omzetten naar classificaties kost tijd en geld. Het is efficiënter om respondenten zelf te laten kiezen uit een lijst met beroepstitels. Deze keuzelijst moet

groot genoeg zijn om een beroepstitel te vinden of ten minste één die ‘dicht’ genoeg bij het eigen beroep ligt om tot dezelfde classificatie te worden gecodeerd”, steekt Kea Tijdens van wal.

Wageindicator

Het begon eind jaren negentig. “Paulien Osse, directeur van WageIndicator, maar destijds journaliste, kwam bij me langs. Weekblad *Intermediair* had een salarischek op z'n website geplaatst, maar deze was volledig gericht op academische mannenberoepen. Paulien vroeg of het mogelijk was ook voor vrouwen een salarischek te maken. We hebben toen voor een heel aantal beroepen uitgezocht wat vrouwen verdienen. We hebben verschillende enquêtes opgenomen in de grootste vrouwenbladen en ook een online

versie. 15.000 werkende vrouwen gaven ons hun loon en hun beroep, een groot succes. Later hebben we mannenberoepen toegevoegd en een loonwijzer voor iedereen gemaakt, deze draait inmiddels in 92 landen en heeft 40 miljoen bezoekers per jaar.”

Sterker onderzoek

SERISS staat voor Synergies for Europe's Research Infrastructures in the Social Sciences. Dit vierjarig project richt zich op het versterken van sociaalwetenschappelijk onderzoek in Europa. Eén van de werkpakketten binnen SERISS gaat over het online beschikbaar maken van codeerbare items en de gereedschappen om deze items te integreren in vragenlijsten. Ook de beroepenlijst die Tijdens heeft samengesteld, hoort hierbij. Op surveycodings.org kunnen de sets met

duizenden gecodeerde beroepstitels, industrienaamen, opleidingen en vraagteksten worden opgevraagd. Er wordt uitgelegd wat de beste manier is om deze vragen te stellen en hoe deze te integreren in webvragenlijsten.

surveycodings.org

Over Kea Tijdens:

Kea Tijdens is socioloog en promoveerde in 1989. Zij werkt sinds 1982 bij de Universiteit van Amsterdam en is sinds 2000 onderzoekscoördinator bij AIAS. Ze was gedurende 10 jaar (2004-2014) hoogleraar Vrouw en Arbeid aan de faculteit Sociologie, Erasmus Universiteit Rotterdam. Sinds 2001 is ze ook wetenschappelijk coördinator van de WageIndicator/Loonwijzer web survey.

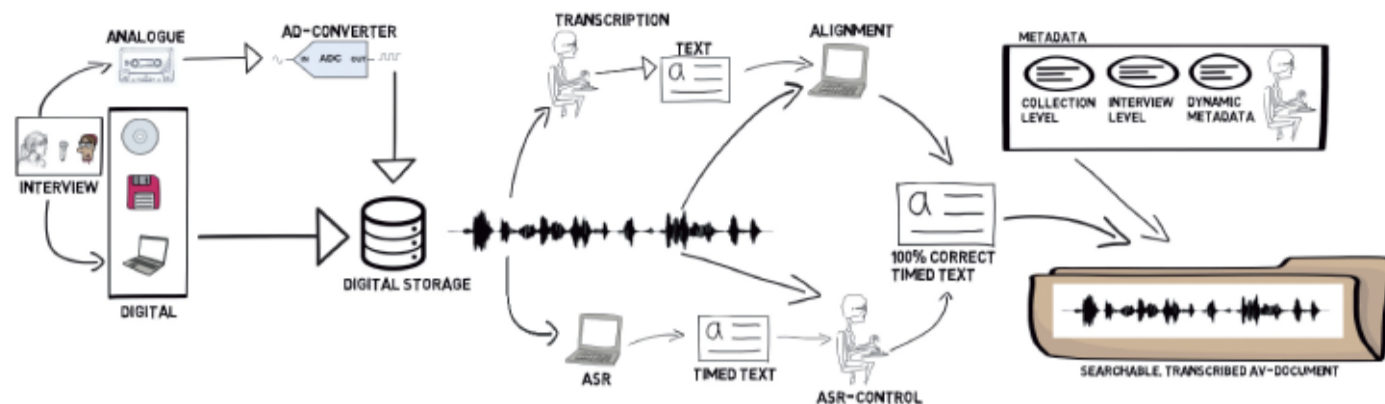
Linked Data en Syrische bronnen

Reeds in de oudste getuigen van het Syrische christendom zien we een complexe interactie tussen het bijbelse literaire erfgoed en de hellenistische cultuur. De bestudering hiervan vereist een optimaal gebruik van encyclopedische, geografische en lexicografische bronnen. Er komen steeds meer van dergelijke bronnen online beschikbaar, maar ze zijn vaak moeilijk toegankelijk en missen de verbinding met elkaar en de teksten waarop zij betrekking hebben. In de CLARIAH research-pilot LinkSyri: Linking Syriac Data, uitgevoerd door het Eep Talstra Centre for Bible and Computer (VU) en DANS worden deze bronnen met Linked Data toegankelijk gemaakt en met elkaar verbonden. (WvP) clariah.nl/projecten/research-pilots/linksyri

Transcriptie met 95 procent correcte teksten

Spraakherkenning in stroomversnelling

De komst van Deep Neural Networks maakt het analyseren en doorzoeken van spraakopnames een stuk makkelijker. Steeds meer oral history-onderzoekers maken er gebruik van. *Erica Renckens*



Automatische spraakherkenning (ASR) bestaat inmiddels al een tijdje, maar de resultaten lieten tot voor kort nogal wat te wensen over. Getraind op één gebruiker ging het nog vrij aardig, maar algemene software kon vaak weinig chocola maken van een audiosignaal. Dat maakte het analyseren en doorzoeken van spraakopnames erg lastig. De komst van Deep Neural Networks (DNN), die een beter akoestisch model genereren dan de oude methodes, heeft hier verandering in gebracht.

Tot op de milliseconde

Een DNN-gebaseerde herkenner voor het Nederlands is ontwikkeld door de Universiteit Twente en het

De transcription chain leidt de gebruiker langs de verschillende stappen om zijn analoge opname te verwerken tot een digitaal formaat dat analyse mogelijk maakt. credits Arjan van Hessen

Nederlands Instituut voor Beeld en Geluid. Zij gebruikten hiervoor Kaldi, een gratis open-source toolkit. In samenwerking met de Radboud Universiteit en CLARIN-EU wordt de herkenner opgenomen in een transcription chain: een workflow waarin onderzoekers hun opnamen kunnen digitaliseren, transcriberen, corrigeren en voorzien van metadata. Het resultaat is een reeks woorden die tot op de milliseconde is opgelijnd met de opname.

Taalmodel

“Als de omstandigheden optimaal zijn, is 90 tot 95 procent van de her-

kende tekst nu correct”, aldus Arjan van Hessen, spraaktechnoloog aan de Universiteit Twente. “Veel gebruikers hebben echter geen idee hoe een goede opname te maken. Dikwijls legt men gewoon de smartphone op tafel waardoor ieder contactgeluid keihard in de opname komt. Ook realiseren veel interviewers zich niet dat hun terugkoppeling – ‘ja, ja’, ‘ga door’ – ook in de transcriptie belandt, wat de leesbaarheid niet verhoogt. Liever neem je het gesprek op met aparte headsets.”

Het resultaat van de spraakherkenning hangt niet alleen af van de

kwaliteit van de opname. “Het taalmodel voorspelt de kans op woord X gegeven de woorden Y en Z. Het woordgebruik verschilt per gespreksonderwerp, dus idealiter zijn er ook verschillende taalmodellen. Maar dat is nu nog te arbeidsintensief. We werken aan de mogelijkheid om onderzoekers zelf woordenlijsten te laten toevoegen.”

Toekomstig gebruik

Ook hoopt het team de transcription chain verder uit te breiden met een editor waarmee de output van de spraakherkenner eenvoudig gecorrigeerd kan worden. “Daar zijn

wel al toepassingen voor, maar die zijn niet altijd gebruiksvriendelijk. Of de tijdinformatie gaat verloren waardoor zoeken in het geluidssignaal niet meer mogelijk is.” Van Hessen ziet het gebruik van ASR bij de overheid de laatste tijd sterk toenemen. Zo wordt de herkenner momenteel getest door de politie en de FIOD. “En ook onderzoekers die gebruikmaken van gesproken getuigenissen ontdekken steeds meer het gemak ervan. Met de transcription chain zal dit hopelijk nog verder toenemen.”

oralhistory.eu/workshops/transcription-chain

Hoe meet je Open Science?

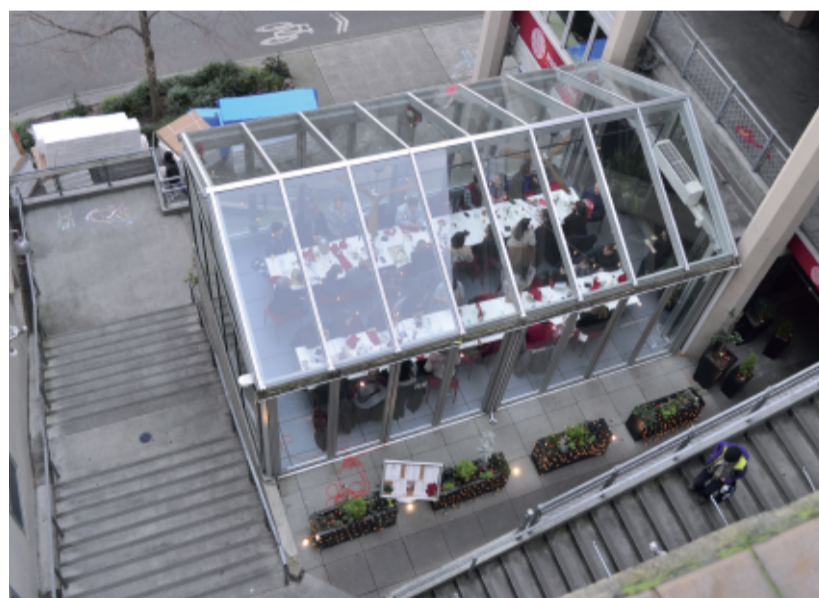
Een sticker voor goed gedrag

Rens van de Schoot schreef in het boek *Hoe zwaar is licht. Meer dan 100 dringende vragen aan de wetenschap dat wetenschappers in een glazen huis op een marktplein moeten gaan werken.* *Heidi Berkhout*

Naar aanleiding van diverse fraudegevallen, meldingen in de media van onverantwoord toepassen van wetenschap, en de replicatiecrisis die diverse vakgebieden boeit, zijn tal van initiatieven ontstaan om de goede naam van de wetenschap te herstellen. Zo verklaarden diverse organisaties de ambities uit het Nationaal Plan Open Science na te streven.

Nationaal Plan

Sander Dekker, demissionair staatssecretaris van Onderwijs, Cultuur en Wetenschap, staat aan de basis van dit plan. Hij roept op tot meer Open Science, waarmee hij bedoelt: (1) het open access publiceren van wetenschappelijke publicaties – dat wil zeggen gratis te lezen; (2) optimaal (her)gebruik van onderzoeksdata – door bijvoorbeeld data FAIR te deponeren en te hergebruiken via archieven, zoals



EASY, het online archiveringstelsel van DANS; (3) het aanpassen van evaluatiesystemen – in het landelijke protocol voor evaluatie van wetenschappelijk onderzoek (het Standard Evaluation Protocol, opgesteld door de KNAW, VSNU en NWO) is bijvoorbeeld het punt ‘productiviteit’ komen te vervallen.

Sticker voor open data

Een belangrijk thema is het beschikbaar maken van onderzoeksdata. “Tegenwoordig krijg je bij sommige tijdschriften een sticker

als je je data openbaar hebt gepubliceerd. Dat is een mooie bijdrage aan Open Science. We moeten echter oppassen dat zo’n sticker zich niet tegen ons keert”, stelt Rens van de Schoot, universitair hoofddocent bij de Universiteit Utrecht, lid van De Jonge Akademie en betrokken bij de ontwikkeling van het Nationaal Plan Open Science. “Het aantal stickers gaan tellen bij beoordelingen lijkt een slim plan. Immers, meer stickers betekent meer Open Science. Maar wat te doen met al die onderzoekers die hun data niet openbaar mogen maken en dus

Van de Schoot adviseert in zijn bijdrage ‘Wat zijn de regels van de wetenschap en liggen die voor altijd vast?’ in het boek ‘Hoe zwaar is licht. Meer dan 100 dringende vragen aan de wetenschap’ dat wetenschappers uit hun ivoren toren moeten komen om de wetenschap opener te maken. “Beter is om de wetenschap in een glazen huis uit te voeren. Hiermee wordt wetenschapsbeoefening zichtbaar voor iedereen, maar we moeten niet doorslaan.” foto Joe Mabel

nooit zo’n sticker kunnen verdienen? Denk aan medische gegevens. Moet je dan alleen maar werken aan projecten waarvan de data openbaar gemaakt kunnen worden?”

Zelf stappen zetten

Gelukkig zijn ook voor dit soort typen data genoeg oplossingen. “Tijdens een bijeenkomst over De Huiskamer van de Wetenschap, een initiatief van De Jonge Akademie (zie kader), noemde Peter Doorn, directeur van DANS, bijvoorbeeld dat data nooit ‘zomaar’ gedeponerd worden. Onderzoekers kunnen zelf

bepalen of data openbaar toegankelijk zijn.” Van de Schoot vervolgt: “Het gaat erom dat we met z’n allen bedenken wat per wetenschappelijk veld nodig is om de volgende stappen te zetten. We hoeven echt niet allemaal in een glazen huis op de markt te gaan werken, maatwerk kan en mag. Maar als we allemaal vandaag nog één stap zetten om een beetje opener te zijn, dan zou dat al een enorme impact hebben.”

Kijk voor alle initiatieven op het gebied van Open Science op openscience.nl

Huiskamer van de wetenschap

De Jonge Akademie wil weg van een onderzoekscultuur van ‘zo doen we het nu eenmaal hier’ naar een opener discussie over wetenschapsbeoefening. Het daartoe in het leven geroepen project Huiskamer van de Wetenschap bestaat uit rondreizende huiskamerbijeenkomsten en een website (in wording). Zo kunnen jonge wetenschappers kennismaken van discussies over de onderzoeksmores in hun vakgebied en daar zelf over meepreten. Denk aan thema’s als auteurschap, diversiteit en werkdruk.

De Jonge Akademie is onderdeel van de KNAW. dejongeakademie.nl

KORT

International Digital Preservation Day

Op 30 november vindt de eerste International Digital Preservation Day (IDPD) plaats, een initiatief van de Digital Preservation Coalition (DPC) en internationale partners. Tijdens IDPD wordt aandacht besteed aan de digitale collecties die wereldwijd beheerd worden door allerlei organisaties. In Nederland coördineert het Netwerk Digitaal Erfgoed/NCDD de activiteiten die door collectiebeherende instellingen in de domeinen archief, musea, bibliotheken, wetenschap en AV/media worden opgezet. Kijk op de website van de NCDD voor een overzicht van de activiteiten en volg #IDPD17 op Twitter. (HB)

ncdd.nl

ERIC-status voor CESSDA

CESSDA, het consortium van Europese sociaalwetenschappelijke data-archieven, is op 14 juni een Europese onderzoeksinfrastructuur (ERIC) geworden. CESSDA bestaat al sinds 1976 als een informele netwerkorganisatie, maar als ERIC heeft het nu een officiële juridische status gekregen waardoor de data-archieven van de lidstaten makkelijker kunnen samenwerken. CESSDA richt zich op het samenbrengen en integreren van sociaalwetenschappelijke datacollecties uit de verschillende Europese landen om zo nationaal en internationaal onderzoek te bevorderen en samenwerking te ondersteunen. Er zijn 14 landen lid van CESSDA. (MW)

cessda.eu

Partnerschap Dryad en DANS

Dryad en DANS zijn een samenwerking aangegaan waarmee de toegankelijkheid en het behoud van wetenschappelijke gegevens in de Dryad Digital Repository verzekerd is voor de lange termijn. De meer dan 50.000 wetenschappers die hun onderzoeksdata al bij Dryad hebben gedeponereerd, kunnen rekenen op blijvende toegang tot hun data, plus een extra laag beveiliging als direct gevolg van deze samenwerking. De openbaar toegankelijke inhoud van de Dryad-servers, met meer dan 15.000 datasets en 50.000 bestanden, zal regelmatig als back-up worden gekopieerd naar DANS. Zo wordt het risico van dataverlies of -beschadiging ook op de lange termijn tot een minimum beperkt. Daarnaast garandeert DANS de bruikbaarheid van de digital object identifiers (DOI's) van Dryad nu en in de toekomst. (HB)

datadryad.org

Big-data-aanpak mist cultureel erfgoeddata

K-PLEX zoekt verborgen data

Het koppelen van grote hoeveelheden data lijkt het antwoord op allerlei onderzoeksvraagstukken. Maar veel data blijven bij deze aanpak verborgen. Het K-PLEX-project onderzoekt de gevolgen daarvan. *Marion Wittenberg*

Binnen de geesteswetenschappen en de cultureel erfgoedsector zijn data vaak ongestructureerd of zelfs helemaal niet gedigitaliseerd, en hierdoor onbruikbaar voor een big-data-aanpak. Wat voor invloed heeft dit op de kennisproductie? Die vraag probeert het Horizon 2020-project K-PLEX (Knowledge Complexity) te beantwoorden. Want: "Schuilt er geen gevaar in het feit dat wanneer de focus van wetenschapsbeoefening op het analyseren van grote databestanden komt te liggen, een groot deel van de werkelijkheid buiten beschouwing wordt gelaten, omdat dit niet te vertalen is in gestructureerde data?"

Bewustwording

Het project – uitgevoerd door het Trinity College Dublin, de Freie Universität Berlin, Tilde in Riga en DANS – onderzoekt welke strategieën geesteswetenschappers hebben in het omgaan met ongestructureerde data en wat voor kennis theoretische consequenties dit heeft. Het doel is om bij te



Versillen in data, verschillen in conventies en gebruiken. Iedereen heeft zijn eigen blik. credits Mike Priddy

dragen aan de ontwikkeling van ICT-onderzoek binnen en buiten de geesteswetenschappen en bewustwording te vergroten over de specifieke eigenschappen van geesteswetenschappelijke data. Mike Priddy en Nicola Horsley, informatiekundige en research fellow bij DANS, onderzoeken voor een deelproject van K-PLEX hoe binnen de geesteswetenschappen omgegaan wordt met informatie. Wat leidt ertoe dat data verborgen of ongebruikt blijven? Priddy en Horsley zijn op zoek naar verhalen en anekdotes van onderzoekers, archivariissen en andere professionals werk-

zaam in archieven, bibliotheken, galeries, musea binnen de cultureel erfgoedsector over werkprocessen, conventies en gebruiken. Veranderen deze praktijken nu er meer nadruk komt te liggen op de analyse van grote bestanden? En wat gebeurt er met gegevens die niet aan bod kunnen komen binnen de digitale geesteswetenschappen? Het onderzoek richt zich op ervaringen met zowel digitale collecties als met meer traditionele gegevens en bestaat uit een enquête (van ongeveer een half uur) eventueel gevolgd door een interview. kplex-project.eu/take-part

Bij de KB in de weer met oude kranten en nieuwe algoritmes

Sinds wanneer illustreren foto's het krantennieuws?

Historicus Thomas Smits speurt met automatische beeldherkenningssoftware naar foto's in gedigitaliseerde historische kranten en ontdekt steeds meer. *Steven Claeysens*

De Koninklijke Bibliotheek (KB) nodigt jaarlijks twee jonge, beloftevolle onderzoekers uit om met behulp van digitale technieken gezamenlijk een vraagstuk naar keuze te lijf te gaan. Dit jaar is onder meer Thomas Smits te gast. Hij promoveert aan de Radboud Universiteit op een onderzoek naar de transnationale handel in illustraties van het nieuws in de negentiende eeuw. "Tegenwoordig wordt er vaak gezegd dat bepaalde beelden 'de wereld overgaan'. In mijn proefschrift wil ik aantonen dat de wortels van dit proces in de negentiende eeuw liggen."

Visuele nieuwscultuur

In de KB richt hij zijn pijlen op de visuele nieuwscultuur aan het eind van de negentiende en het begin van de twintigste eeuw: "Hoewel de fotografie rond 1830 werd uitgevon-

JONG TALENT



Thomas Smits foto Laura Smits

den, werden pas vanaf 1880 foto's in kranten en tijdschriften afgedrukt. Het op grote schaal reproduceren van foto's werd namelijk pas mogelijk met de uitvinding van de halftone-druktechniek. In mijn KB-project probeer ik te onderzoeken wanneer Nederlandse kranten foto's gingen gebruiken. Daarnaast wil ik een belangrijke these uit de mediageschiedenis onderbouwen die stelt dat kranten lange tijd foto's en illustraties naast elkaar gebruikten om het nieuws af te beelden."

'We gebruiken convolutional neural networks om de afbeeldingen te delen'

Computer vision

"We doen dit onderzoek met behulp van zogenaamde *computer vision-technieken*. We hebben eerst alle afbeeldingen uit gedigitaliseerde kranten op Delpher gehaald en deze in een database gezet. Vervolgens keken we met een combinatie van algoritmes of een afbeelding een foto of een illustratie is. We gebruiken *convolutional neural networks* om de afbeeldingen in onze database in te delen. Zo kunnen we alle kaarten en weersvoorspellingen herkennen.

Ook lukt het steeds beter om afbeeldingen met een bepaald onderwerp te herkennen, zoals illustraties en foto's van grote menigtes, of een bepaalde stijl, zoals politieke cartoons. Als het ons lukt om de afbeeldingen beter in te delen, kan de KB dit ook gebruiken om het zoeken naar informatie makkelijker te maken."

Traditionele vragen

"Ik zie mezelf als een traditioneel historicus die geïnteresseerd is in vrij traditionele historische vragen. Ik raak er echter steeds meer van overtuigd dat nieuwe computationele technieken ons kunnen helpen om oude vragen van de kunstgeschiedenis, de mediageschiedenis of de zogenoemde *visual culture studies* te beantwoorden. Hoe zit het bijvoorbeeld met stijl? Wat is het verschil en de overlap tussen visuele en conceptuele gelijkenis? Hoe hangt de betekenis van een afbeelding af van de omliggende tekst? Na dit project hoop ik dan ook een nieuwe aanvraag te schrijven waar ik verder kan met dit soort vragen."

illustratednewspictures.tumblr.com

Ook in 2018 nieuwe voorzieningen en diensten voor erfgoedsector

Netwerk Digitaal Erfgoed ook na eerste fase goed op weg

Netwerk Digitaal Erfgoed ging afgelopen zomer de tweede fase in. Wat is in de eerste fase bereikt en welke stappen zijn er nog te zetten?

Erica Renckens

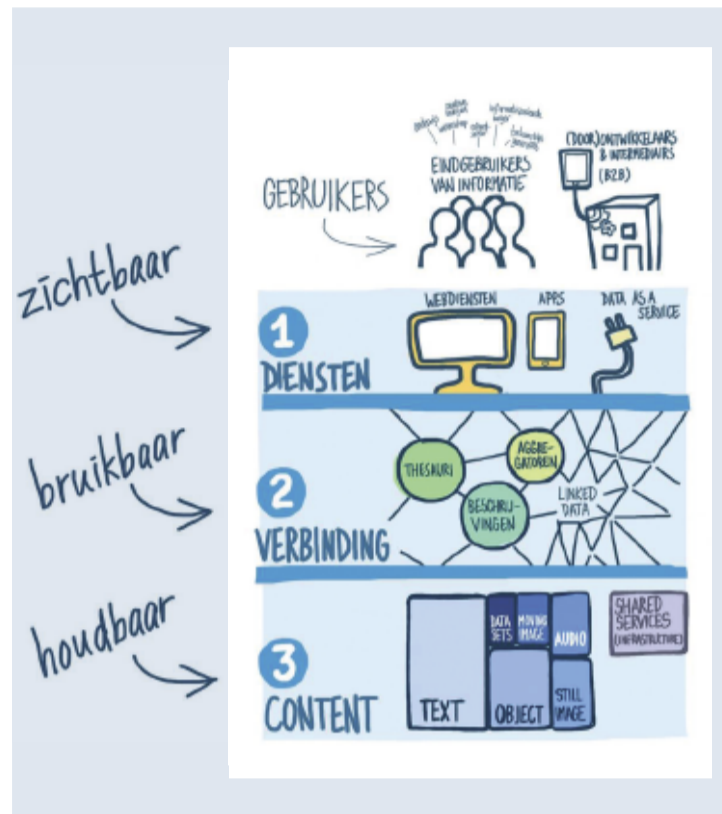
De afgelopen twee jaar werkte het Netwerk Digitaal Erfgoed (NDE) aan de uitvoering van de Nationale Strategie Digitaal Erfgoed, die de erfgoedsector in 2015 samen met het ministerie van Onderwijs, Cultuur en Wetenschap (OCW) opstelde. Doel was om de zichtbaarheid, bruikbaarheid en houdbaarheid van digitaal erfgoed te verbeteren door een stelsel van landelijke voorzieningen en diensten te ontwikkelen.

Unieke links

Marcel Ras, coördinator van het werkpakket Digitaal Erfgoed Houdbaar en programmamanager van de Nationale Coalitie Digitale Duurzaamheid (NCDD), noemt als voorbeeld van een succesvolle dienst vanuit het NDE het project rondom Persistent Identifiers (PID's), unieke en blijvende links naar digitale objecten. "Het belang van PID's was duidelijk, maar ze werden nog niet in groten getale gebruikt – het bleef vooral bij praten. Inmiddels hebben softwareleveranciers op onze aanwijzingen een module ontwikkeld waarmee erfgoedinstellingen zelf PID's kunnen toewijzen aan hun collectie."

Niet zonder anderen

"Geen erfgoedinstelling kan zonder de hulp van anderen voorzieningen en services aanbieden",



Online beschikbaar

Alle tools, websites, rapporten en overige kennis uit het werkpakket Digitaal Erfgoed Houdbaar is beschikbaar via www.ncdd.nl/news/nieuwe-tools-en-tricks-duurzame-toegang-tot-collecties/

Afbeelding links:

De drie pijlers, en het nodige werk aan de winkel, van de nationale strategie op een rij: zichtbaar, bruikbaar en houdbaar *credits Elco van Stavereen, denkschets.nl*

stelt Ras. "De komst van het NDE twee jaar geleden bracht financiële en bestuurlijke armslag om ideeën uit te voeren. Hiermee gaan we in de tweede fase verder, waarbij de vijf betrokken knooppunten (Koninklijke Bibliotheek, Nederlands Instituut voor Beeld en Geluid, Rijksdienst voor Cultureel Erfgoed, Nationaal Archief en de KNAW) nog intensiever gaan samenwerken." Zo wordt begin 2018 een coöperatie in het leven geroepen; de directeurs van de knooppunten vormen het bestuur. Ras: "Deze coöperatie NDE krijgt een eigen bureau van waaruit samenwerkingsprojecten uitgevoerd zullen worden."

Voor de tweede fase heeft Ras duidelijke doelen voor ogen. "Collecties moeten onderling beter aan elkaar gekoppeld worden, zodat eindgebruikers makkelijker de informatie kunnen vinden die ze zoeken, onafhankelijk van welke instelling die bezit. Ook moeten voorzieningen nog meer gedeeld worden, zodat niet elke instelling die zelf hoeft te ontwikkelen. Daarnaast moet de samenwerking echt gestalte krijgen; tussen de knooppunten, maar ook daarbuiten. Ik heb er het volste vertrouwen in dat dat gaat lukken."

den.nl/nde

GELEZEN

Historici.nl is vernieuwd!

Historici.nl, de startpagina voor historisch Nederland, is vernieuwd.

Kayleigh Goudsmit

De website is een initiatief van het Koninklijk Nederlands Historisch Genootschap (KNHG) en Huygens ING. Het nieuwe design is responsive, overzichtelijk en laat de dynamiek in het vakgebied zien. Het design is nieuw, maar de vertrouwde elementen zijn gebleven: historici.nl geeft een overzicht van gedigitaliseerde bronnen van verschillende instituten. De bronnen en data zijn op een zorgvuldige en wetenschappelijk verantwoorde wijze online gepubliceerd. De bronnen zijn hier doorzoekbaar op onderwerp, type en periode. Daarnaast blijft historici.nl toegang bieden tot de resources van verschillende instituten, en staan op de site nieuws en debatten over het vakgebied. Correspondenten hebben een centrale rol op de vernieuwde website: vanuit verschillende werkvelden houden zij vakgenoten op de hoogte van de beroepspraktijk en de uitdagingen waarvoor historici staan.

historici.nl



COLUMN

Hik bin gezmond

En half jaar geleden heb ik mijn proefschrift verdedigd.

Dat gaat over jeugdverhalen over jodenbekering. Om die verhalen te vinden heb ik onder meer intensief gebruikgemaakt van Google Books en van Delpher, een gigantische databank met ruim zestig miljoen pagina's uit Nederlandse kranten, boeken en tijdschriften.

Het duurde even voordat ik in de smiezen had hoe je die verhalen het best kunt vinden. Uiteindelijk bleek het heel vruchtbaar om te zoeken op een joodse naam in de nabijheid van Jezus of Messias, want dat zijn vaste verhaal-elementen.

In de tachtig bekeringsverhalen die ik uiteindelijk heb gevonden en geanalyseerd heten de meeste joodse meisjes en vrouwen Debora, Esther, Lea, Myriam, Rachel(tje), Rebecca, Saar, Sara(h) of Saartje. En de meeste joodse jongens en mannen Iza(a)k, Isa(ac), Levi(e), Moos, Mozes, Nathan, Sam(met)je, Samuel, Simon of Zadok. Dat geldt niet alleen voor verhalen

over jodenbekering, maar voor allerlei verhalen over joden.

Inmiddels heb ik besloten om mijn onderzoek naar joden in de (jeugd)literatuur te verbreden en daarom moest ik nieuwe zoektermen verzinnen voor Google Books en Delpher. Ik ben daarbij tot een inzicht gekomen dat ook voor andere onderzoekers nuttig zou kunnen zijn, namelijk: zoek (ook altijd) op stereotyperingen.

Om literaire personages herkenbaar te maken voor een breder publiek nemen veel schrijvers hun toevlucht tot stereotypeeringen. Die namen zijn daar een voorbeeld van: wie Levi(e) heet is joods, dus dat hoeft je er als schrijver niet per se bij te vermelden. Bij joden vind je zulke stereotypeeringen niet alleen terug in de namen, maar ook in de clichématige beschrijvingen van hun uiterlijk

(zwart haar, zwarte ogen, grote neus) en zelfs van hun taalgebruik.

Met name laagopgeleide joden spraken het Nederlands vroeger anders uit dan niet-joden. Daarnaast gebruikten zij een aantal kenmerkende woorden en uitdrukkingen. Voor schrijvers was dit natuurlijk *gefundenes Fressen*: om een joods personage neer te zetten hoefde je hem of haar bijvoorbeeld alleen maar 'blijf gezond' of 'zal je gezond blijven' te laten zeggen, een zegewens die zeker



foto Leo van Velzen

tweehonderd jaar als typisch joods is beschouwd. Schrijvers die de joods-Nederlandse uitspraak wilden benaderen, om zo een volks joods type tot leven te wekken, deden dat door tamelijk willekeurig de letter H aan een woord toe te voegen. Of door de E in een woord te vervangen door een I. 'Ik ben' werd daardoor 'hik bin' en de zegewens blijf gezond werd bijvoorbeeld blijf gezmond.

Neem de proef op de som en zoek in Delpher op het woord gezmond. U krijgt 21 zoekresultaten die vrijwel allemaal relevant zijn. In de afdeling Boeken Basiscollectie bijvoorbeeld het bekeringsboekje Jodenkerstfeest. En in de afdeling kranten onder meer het obscure feuilleton 'Het paard van Moos'. Bij Google Books vindt u met dezelfde zoekopdracht een Levi

(hoe kan het anders) die in een vergeten bron uit 1869 zegt: 'Ghod dhoe je gezhond blijven'. Plus een Mozes die in een bron uit 1863 zegt: 'Blijf gezmond! Nah, jonkheer.' Nah is trouwens ook een goede zoekterm: schrijvers beschouwden dit als een typisch joodse uitroep. Probeer eens in Delpher nah of nha PROX (in de nabijheid van) hik.

Zoek in digitale bronnen naar dikke lippen en dikke, platte neuzen en u zult zwarten vinden. Zoek naar waarzegster in de nabijheid van woonwagen en u zult Sinti en Roma vinden. Kortom: zoek naar onweerstaanbare clichés en stereotyperingen in namen, uiterlijk, karakter en gedragingen en uw materiaalcollectie zal met prachtige, doorgaans niet eerder gebruikte bronnen worden verrijkt.

Ewoud Sanders

Taalhistoricus en journalist. Sanders is vaste medewerker van onder meer NRC Handelsblad en Onze Taal.