

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



Verschillen 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 computationale 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