

Online toegankelijk maken en delen van data beloond

BoschDoc, AHCODA-DB en OpenML winnaars Dataprijs

“Indrukwekkende initiatieven en lichtende voorbeelden van het delen van data.” Tijdens de uitreiking van de Nederlandse Dataprijs 2016 sprak de juryvoorzitter vol lof over de winnende datasets. *Heidi Berkhout*

Afgelopen november werd voor de vierde keer de Nederlandse Dataprijs uitgereikt door Research Data Netherlands (RDNL). De Dataprijs is een prijs die waardering geeft aan een onderzoeker of onderzoeksgroep die extra bijdraagt aan de wetenschap door onderzoeksdata beschikbaar te maken voor nieuw of aanvullend onderzoek. Uit 52 inzendingen en 9 nominaties werden uiteindelijk drie winnaars gekozen. Namens RDNL beloonde juryvoorzitter prof.dr. José van Dijck (president KNAW) de winnaars met een Dataprijs 2016-sculptuur en € 5.000 om de dataset toegankelijk(er) te maken. E-data zet de winnaars op een rij.

Vereerd, trots en blij

Hoe voelt het om winnaar van de Dataprijs 2016 te zijn?

Loes Scholten, projectleider BoschDoc: “Wij zijn zeer vereerd dat de jury van de Dataprijs het internationale en interdisciplinaire belang van de dataset onderstreept. BoschDoc is immers het product van mensen vanuit verschillende disciplines en achtergronden zoals: (kunst)historici, archivariissen, mediëvisten, vertalers, programmeurs. We hebben getracht door vereende kennis een wetenschappelijk verantwoorde en functionele database te creëren. De combinatie van de sensationele Jheronimus Bosch tentoonstelling in 2016 en deze dataset maakte het mogelijk het grote publiek met wetenschap kennis te laten maken.” Bioinformatics software ontwikkelaar bij Sylics Bastijn Koopmans: “We zijn ontzettend blij met en trots op deze prijs. Het laat zien dat de jaren van ontwikkeling het meer dan waard waren om een systeem te ontwikkelen waar gedragsdata centraal verzameld kan worden, en direct beschikbaar is voor onszelf en on-



Foto (van links naar rechts): winnaars A.M. Koldewij, L. Scholten en R.C. Hage (BoschDoc), M. Loos en A. Steenbergen (AHCODA-DB) en J. Vanschoren (OpenML) Foto Annemiek van der Kuil

derzoekers wereldwijd. Ook de jury heeft onze doelen lovend verwoord: “AHCODA-DB is een enorm rijke dataset, van groot belang als aanvulling op reeds verzamelde data, als replicatie, als vergelijking en als voorbereiding op nieuw onderzoek. Het vrij beschikbaar maken van zowel gepubliceerde als ongepubliceerde data scheelt veel tijd, geld en spaart in dit geval ook dierenlevens.” Joaquin Vanschoren, universitair docent bij de Technische Universiteit Eindhoven: “We zijn uiteraard erg blij met de dataprijs. OpenML is een open source-project dat vooral ontwikkeld wordt door onderzoekers in hun vrije tijd. De Dataprijs is een goede opsteker voor al het werk dat ze er al in hebben gestoken. We zijn blij dat de jury heeft benadrukt dat we een heel breed platform zijn, met data uit verschillende disciplines, want een deel van ons doel is om beter te kunnen samenwerken met specialisten uit verschillende wetenschappelijke disciplines. Inmiddels zijn we benaderd door Microsoft om ons te helpen met grootschalige berekeningen in ‘de cloud’. Ik weet niet of dat direct met het winnen van de Dataprijs te maken heeft, maar het verbetert zeker onze uitstraling. Ook heeft de uni-

versiteit Eindhoven beslist om ons meer te ondersteunen, en daar heeft de Dataprijs zeker bij geholpen.”

Prijzengeld goed besteed

Wat gaan de winnaars met het prijzengeld doen?

Scholten: “De inhoud van BoschDoc is nog niet volledig, waardoor het prijzengeld op diverse manieren nuttig besteed kan worden. Er moet bijvoorbeeld nog een slag gemaakt worden bij de vertaling van de contextuele toelichtingen naar het Engels. Wanneer dat voltooid is, zullen de archiefschatten in BoschDoc voor nog meer mensen, wereldwijd, gaan spreken.” Koopmans: “Het prijzengeld zal worden gebruikt om de database verder uit te breiden, nieuwe functionaliteiten toe te voegen en de data nog toegankelijker te maken voor de wetenschap.” Vanschoren: “We zijn er nog niet helemaal uit wat we met het prijzengeld gaan doen, maar bovenaan het verlanglijstje staan het nog toegankelijker maken van de data (bv. betere websites en andere interfaces) alsook gemakkelijke archivering van data om dit te linken aan publicaties.” researchdata.nl

Humaniora en sociale wetenschappen:

BoschDoc

BoschDoc bevat 1.000 schriftelijke documenten over de laatmiddeleeuwse schilder Jheronimus Bosch. De documenten zijn getranscribeerd en geannoteerd in de tool eLaborate, ontwikkeld door Huygens ING. Faceted search biedt zoekmogelijkheden zoals: publicatiejaar, auteur en documenttype. Van alle documenten zijn een facsimile en transcriptie opgenomen. Bij elke brontekst staat bovendien een contextuele toelichting. Dit alles maakt deze dataset bruikbaar voor een breed publiek.

boschdoc.huygens.knaw.nl

Medische en levenswetenschappen:

AHCODA-DB

De publieke website van AHCODA-DB bevat data van geautomatiseerde en conventionele gedragstesten van genetische muislijnen. Via meer dan 400 automatisch gegenereerde grafieken en meta-analyses kunnen onderzoekers hun eigen data vergelijken met de referentie data van eerdere onderzoeken in AHCODA-DB, of data opzoeken voordat ze beginnen met hun onderzoek. De gebruiksvriendelijke interface maakt de database makkelijk toegankelijk.

public.sylics.com

Exacte en technische wetenschappen:

OpenML

OpenML is een online open source-platform om samen machine learning modellen te ontwikkelen. OpenML is geïntegreerd in populaire machine learning tools (zoals R, scikit-learn en WEKA) om gemakkelijk data te delen voor reproductie en hergebruik in andere onderzoekstudies. Momenteel hebben zo'n 2.000 geregistreerde gebruikers via OpenML toegang tot zo'n 20.000 datasets, 3.000 algoritmes en 1,7 miljoen machine learning modellen en evaluaties daarvan.

openml.org

KORT

ODISSEI: het nieuwe dataplatform voor sociale wetenschap

De mens- en maatschappijwetenschappen bundelen hun krachten in een nieuw dataplatform: ODISSEI (Open Data Infrastructure for Social Sciences and Economics Innovation). Het initiatief was eerder bekend als NDSW, Nationale Data-infrastructuur voor de Sociale Wetenschappen. De wetenschappelijk trekker van het dataplatform is Pearl Dykstra van Erasmus Universiteit. Binnen ODISSEI werken Nederlandse universiteiten, projectleiders van dataverzamelingen, NWO, CBS, DANS, rijkskennisinstellingen, hogescholen, overheden en bedrijven samen op het gebied van koppeling van grote dataverzamelingen en verschillende databronnen. Het platform is opgenomen in de nationale Roadmap Grootschalige Wetenschappelijke Infrastructuur. (MdB) nwo.nl/onderzoek-en-resultaten/programmas/odissei

Onderzoeksfinitiering door NWO in 2017

Wetenschapsfinancier NWO verandert momenteel ingrijpend. Dit heeft gevolgen voor de onderzoeksfinitiering en de manier waarop commissies en besturen de beoordeling, selectie en besluitvorming uitvoeren. Op de website van NWO staat een handig overzicht van de NWO-onderzoeksfinitiering voor de sociale en geesteswetenschappen in 2016/2017. (NWO)

nwo.nl/over-nwo/organisatie/nwo-onderdelen/gw/financieringsinstrumenten+sociale+en+geesteswetenschappen+2016+2017