

13 onderzoeksfaciliteiten in agenda KNAW

Boukje Cuelenaere

Op 31 oktober is de KNAW-agenda grootschalige onderzoeksfaciliteiten gepresenteerd. Eerder dit jaar waren alle disciplines uitgenodigd om een droom te formuleren: welke onderzoeksfaciliteit is ideaal over een periode van 20-30 jaar? Dit leverde een grote variatie aan dromen en wensen op die door de KNAW zijn beoordeeld.

Uiteindelijk zijn dertien faciliteiten in de agenda opgenomen. Enkele voorbeelden: DESIDERIA, een flexibele semantische infrastructuur voor digitale exploratieve tekstanalyse, ADVANT, een geavanceerd programma voor video-analyse en M3. M3 staat voor Molecuul - Mens - Maatschappij, een interdisciplinaire data-infrastructuur voor populatieonderzoek, waarbij biometrische data, survey-data en administratieve data gekoppeld kunnen worden. Elke faciliteit werd kort gepitched in minder en meer begrijpelijke taal, met een geanimeerde borrel achteraf. De agenda is aangeboden aan NWO en aan het ministerie van OCW.

knaaw.nl/nl/actueel/publicaties

GEHOORD & BIJGEWOOND



Hay Kranen (links) ontvangt van juryvoorzitter Sander van Kempfen (rechts) de beker voor het winnen van de juryprijs tijdens de Hack-a-LOD 2016 foto Jacqueline van der Kort (KB)

Een web van erfgoeddata tijdens Hack-a-LOD 2016

Steven Claeysens

Tijdens de nacht van 11 op 12 november 2016 weerklonken voor één keer weer stemmen in de steriele cellen van de voormalige Utrechtse gevangenis De Vrije Wolf. Het was Hack-a-LOD-tijd. Dertig programmeurs hadden de uitdaging aan-

vaard. In 24 uur tijd moesten ze een slimme toepassing van Linked Open Erfgoeddata bouwen. De volgende ochtend presenteerden ze hun creaties, maar niet dan nadat computerwetenschappers Herbert van de Sompel en Wouter Beek, Koninklijke Bibliotheek-directeur Lily Knibbeler en acteur-fotograaf Thom Hoffman elk vanuit hun ei-

gen perspectief op het belang van openheid van digitaal erfgoed hadden gehamerd. Lily Knibbeler trapte af met een pleidooi voor een overgang van een gecentraliseerde aanpak van erfgoeddata op het web, naar een gedistribueerde aanpak. De datasilo's moeten worden opengemaakt. Van de Sompel maakte een snelle rond-

gang langs zijn indrukwekkend palmares en verklapte zelf zijn rode draad: denk altijd vanuit het web. Hoffman verhaalde in een interview over zijn inspanning om een beeldbank over de geschiedenis van Nederlands-Indië op te zetten. Wouter Beek, ten slotte, gaf praktische tips om de evolutie naar een decentraal web van erfgoeddata te bespoedigen: verzamel data, maak ze schoon, doorzoekbaar en bevrage ze met complexe vragen. Na vastgesteld te hebben dat de eensgezindheid groot is maar de klus nog allesbehalve geklaard, was het aan de celbewoners van die nacht om hun werk aan publiek en jury te pitchen. De publieksprijs ging naar het CLARIAH-team bestaande uit Marieke van Erp (VU) en Richard Zijdeman (IISG) voor hun *If a building could talk*-app. De juryprijs was voor Hay Kranen met zijn *Facebook Messenger-erfgoedbot*. Hij mag zijn idee samen met KB-programmeurs ontwikkelen. Hack-a-LOD werd georganiseerd door het Netwerk Digitaal Erfgoed en de KB. De ochtend werd mede mogelijk gemaakt door CLICKNL en CLARIAH. Slechts een enkeling ontging de ironie van de locatie. hackalod.com

Betrouwbare data-archieven én datasets

Do you want to play FAIR?

Als één woord de gemoederen in Dataland bezighoudt, dan is het het woord FAIR.

Maar hoe vertaal je deze principes voor betrouwbare datasets naar de praktijk?

Peter Doorn

FAIR staat voor Findable, Accessible, Interoperable en Reusable. Het fraaie acroniem FAIR is gedoopt tijdens de workshop *Jointly designing a data FAIRPORT* in Leiden in 2014. Sindsdien is de populariteit alleen maar toegenomen. FAIR-data zijn data die vindbaar, toegankelijk, uitwisselbaar en herbruikbaar zijn. En wie wil dat nu niet? Zelfs de laatste richtlijn voor datamanagement, vereist voor aanvragen in het Horizon 2020-programma van de Europese Unie, verlangt dat onderzoekers FAIR-data produceren.

DSA als voorbeeld

Voor het toepassen van de FAIR-principes in de praktijk kan het kader van het Data Seal of Approval (DSA) worden gebruikt. In 2006 bedacht DANS dit keurmerk voor de betrouwbaarheid van digitale data-archieven, gebaseerd op vijf princi-

DSA principes for data repositories	FAIR principes for datasets
data can be found (internet)	F indable
data are accessible	A ccessible
data are in a usable format	I nteroperable
data are reliable	R eusable
data can be referred to	(citable, included in findable)

FAIR en DSA lijken elkaar perfect aan te vullen: het DSA geeft een kwaliteitsstempel voor databewaarplaatsen, FAIR geeft principes waaraan individuele datasets moeten voldoen credits Peter Doorn, DANS

pes. Ook DSA heeft de afgelopen tien jaar een grote rol gespeeld in Dataland: inmiddels zijn ruim 60 archieven DSA-gecertificeerd. FAIR en DSA lijken elkaar perfect aan te vullen: het DSA geeft een kwaliteitsstempel voor databewaarplaatsen, FAIR geeft principes waaraan individuele datasets moeten voldoen.

Wat betekent dit voor de praktijk? DANS ziet graag dat iedere dataset in een repository met een DSA-stempel een duidelijke FAIR-score krijgt. Die score moet in één oogopslag duid-

lijk maken of een dataset aan de 4 FAIR-principes voldoet. Op die manier zal de FAIR-score tevens een aanduiding zijn voor datakwaliteit.

Verder uitwerken

Een eerste stap is het bereiken van overeenstemming over de uitwerking van de FAIR-principes. Ook hier dient DSA als voorbeeld: de principes zijn uitgewerkt in een 16-tal nauwkeurig toegelichte criteria. Over de FAIR-principes bestaat op hoofdlijnen al overeenstemming, maar in de details zijn er verschillen

in interpretatie. Want wat betekent het precies dat een dataset vindbaar, toegankelijk of interoperabel moet zijn? En is herbruikbaar niet de optelsom van de eerste drie? Kunnen dezelfde criteria voor alle vakgebieden gelden?

Rol datamanager

DANS stelt ook voor om een dataset die bij een archief binnenkomt, direct door een datamanager te laten beoordelen op de mate van FAIRness: zijn de metadata en documentatie op orde? Zijn de data toegankelijk, en zijn eventuele beperkende voorwaarden helder? Zijn de gegevens eenvoudig te koppelen aan andere datasets en wordt gebruik gemaakt van standaard coderingschema's? Uiteraard kan ook aan gebruikers van datasets worden gevraagd hun oordeel te geven over de FAIRness van de datasets door dezelfde vragen te beantwoorden. DANS buigt zich dit jaar verder over de vraag, hoe dit er in de praktijk uitziet.

datafairport.org

force11.org

datasealofapproval.org

dans.knaw.nl

Prioriteiten (33) voor nationale wetenschap

In december nam staatssecretaris van Onderwijs, Cultuur en Wetenschap Sander Dekker de Roadmap Grootschalige Wetenschappelijke Infrastructuur in ontvangst.

Op deze Roadmap staan 33 faciliteiten en clusters die de komende vier jaar de hoogste prioriteit hebben voor de wetenschap in Nederland. In de loop van 2020 vindt de volgende update van de Nationale Roadmap plaats. Een zogenoemde Permanente Commissie zal na de aanstaande financieringsronde de opzet van de huidige Roadmap evalueren.

Tussentijds kunnen in urgente situaties beperkte aanvullingen gedaan worden op deze Nationale Roadmap. Kijk voor een inventarisatie van het aanbod aan bestaande grootschalige infrastructuur in Nederland op de website. (NWO)

onderzoeksfaciliteiten.nl

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.