

## e-Humanities Group van start bij Meertens

Sinds maart van dit jaar is de nieuwe e-Humanities Group van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) gevestigd in het gebouw aan de Joan Muyskenweg in Amsterdam dat ook onderdak biedt aan het Meertens Instituut.

Volgens programmaleider Sally Wyatt is dat een belangrijke stap omdat het samenbrengen en samen laten werken van onderzoekers een van de belangrijke functies is van de groep. Juist omdat de onderzoekers die zich bij de groep aansluiten, merendeels in dienst blijven bij hun eigen instituut of universiteit, spelen ontmoeting en nabijheid een belangrijke rol, licht Wyatt toe. 'Er is op allerlei plaatsen al heel veel gaande op het gebied van eHumanities. Wat wij daaraan toe kunnen voegen is een plek die synergie bevordert door goede omstandigheden en allerlei activiteiten zoals bijeenkomsten, symposia en workshops. Aan de Muyskenweg hebben we prettige werkruimtes met grote kamers die zich lenen voor samenwerking'. Het is de bedoeling om voor dat doel in elk geval wekelijkse bijeenkomsten te organiseren.

Er werken nu vijf mensen bij de eHumanities Group, die de KNAW opdracht nadat vorig jaar werd besloten

de Virtual Knowledge Studio (VKS) op te heffen. Dat kunnen er snel meer worden als de lopende competitie om financiële steun van de Akademie voor projecten in de sfeer van de computationele geesteswetenschappen resultaten gaat opleveren. Het ziet ernaar uit dat dat in de komende weken gaat gebeuren, denkt Wyatt, die zelf ook afkomstig is van de VKS en daarnaast in Maastricht een leerstoel bezet in de *Digital Cultures in Development*. Er is enige vertraging opgetreden want de *Call for Proposals* voor deze competitie ging al in het voorjaar van 2010 uit (zie *e-data* 5-1). Wyatt: 'Er is een aantal voorstellen ingediend, allemaal met een flinke omvang van zo rond de zeven ton. Elk voorstel wordt minstens door twee instituten en een universiteit gedragen. De KNAW kiest er vijf uit. Als die projecten worden toegewezen komen de mensen die ze gaan uitvoeren voor een groot deel van hun tijd bij ons werken. Dan groeit vanzelf het niveau van de activiteiten maar ook het synergetisch effect.'

Het is een 'ingewikkelde constructie', geeft Wyatt toe, en ook een die nieuw is voor de KNAW: een groep die zelf weinig mensen in dienst heeft maar als centrum voor samen-



MICHEL VAN DUSSELDORP

Programmaleider Sally Wyatt

werking en inspiratie dient voor onderzoekers die elders op de loonlijst staan. 'Maar als je innoveert neem je altijd risico'.

De eHumanities Group moet ook een soort erfopvolger worden van het project Alfabab dat binnenkort afloopt, aldus Wyatt. 'Er zijn al een paar projecten in het kader van Alfabab afgerond. Andere komen in de komende maanden zo ver. Op 29 september komt er een grote bijeenkomst om te laten zien wat er is bereikt. Eén van de problemen met digitalisering in de geesteswetenschappen is dat er een aantal jaren wordt geïnvesteerd in mooie diensten of gereedschappen, en dat die daarna in verval raken als de projecten zijn afgesloten. De e-Humanities Group moet dat voorkomen, vooral door in te spelen op de behoeften van de geesteswetenschappers zelf.' (MdG)

<http://ehumanities.nl/>

## Oude teksten bruikbaar voor nieuw onderzoek

Begin dit jaar zijn twee CLARIN-NL projecten afgerond die het mogelijk maken om historische Nederlandse teksten te gebruiken als bron voor onderzoek. Voorheen leidde dit vaak tot problemen door de vele spellingsvarianties die er in de teksten voorkomen.

TICCLops (*Text-Induced Corpus Clean-Up Online Processing System*), een project geleid door Martin Reynaert (Tilburg University), biedt een online dienst die deze spellingsvariantie automatisch corrigeert nadat de teksten zijn ingescand en herkend met behulp van OCR. Het project Adelheid, onder leiding van Hans van Halsteren (Radboud Universiteit), biedt de mogelijkheid om historische teksten automatisch te voorzien van onder meer lemma's en zinsontleding. Het richt zich op Nederlandse teksten uit de veer-

tiende eeuw. Daarbij wordt gebruik gemaakt van hetzelfde systeem als TICCLops, dat zelf niet is gebonden aan een specifieke taal of periode. De diensten zijn beschikbaar via de CLARIN-centra.

Adelheid en TICCLops zijn twee van de elf projecten die in 2010 zijn gehonoreerd binnen de eerste oproep van CLARIN-NL. Al deze projecten zijn nu afgerond of in de afrondende fase. Na de tweede oproep, in de zomer van vorig jaar, werden negen projecten gehonoreerd die op dit moment in volle gang zijn. In de loop van dit jaar zal nog een derde (open) oproep volgen. De gehonoreerde projecten dragen bij aan de ontwikkeling van de CLARIN-infrastructuur, een Europese digitale onderzoeksomgeving voor geesteswetenschappers. (ER)

## Sociale wetenschap vraagt om andere manier van webarchiveren

Onderzoekers in de sociale wetenschappen vragen de webarchieven om zich niet alleen te concentreren op de inhoud van websites, maar ook informatie te gaan verzamelen over interacties op internet.

Dat bleek tijdens een conferentie van het International Internet Preservation Consortium (IIPC) op 9 mei bij de Koninklijke Bibliotheek in Den Haag. De IIPC is een internationaal samenwerkingsverband van circa veertig instellingen die websites archiveren; in de praktijk zijn dat vooral nationale bibliotheken.

Voor de jaarlijkse conferentie had de IIPC een aantal onderzoekers uitgenodigd om te discussiëren over het gebruik van de gearcheverde websites. Paul Girard van SciencesPo, Eric Meyer van het Oxford Internet Institute, en Anne Helmond en Esther Weltevrede van de Universiteit van Amsterdam lieten zien dat voor hen juist de dynamiek van het interactieve internet onderwerp van onderzoek is. Maar daarover verzamelen de webarchieven nog maar nauwelijks gegevens. De techniek is de grootste barrière. Gewone html/http-bestanden zijn goed te archiveren, maar het interactieve web zit vol met technische functionaliteiten die webarchieven nog niet aan kunnen.

Een andere vraag van onderzoekers is om letterlijk alles wat er met een bepaald webarchief gebeurt nauwkeurig te documenteren, zodat de wetenschappers precies weten op welke data ze hun onderzoek baseren. Daarbij gaat het niet alleen om de keuzes die bij selectie zijn gemaakt, maar bijvoorbeeld ook om preserveringsacties, om gege-

vens over mislukte *crawls* en grote systeemstoringen. (IA)

<http://tinyurl.com/3brxrac>

### Digitale editie Anne Frank

Het Huygens ING en de Anne Frank Stichting gaan, m.m.v. het NIOD, nieuw onderzoek doen naar het dagboek en de andere geschriften van Anne Frank. Het onderzoek richt zich op de ontwikkeling van Anne Frank als schrijfster en op de gebeurtenissen die zij beschrijft. Het zal resulteren in een Nederlands-Engelstalige webeditie, waarin het mogelijk wordt om het schrijfproces van het dagboek op de voet te volgen. Annotaties en illustraties zullen de teksten ondersteunen. (PB)

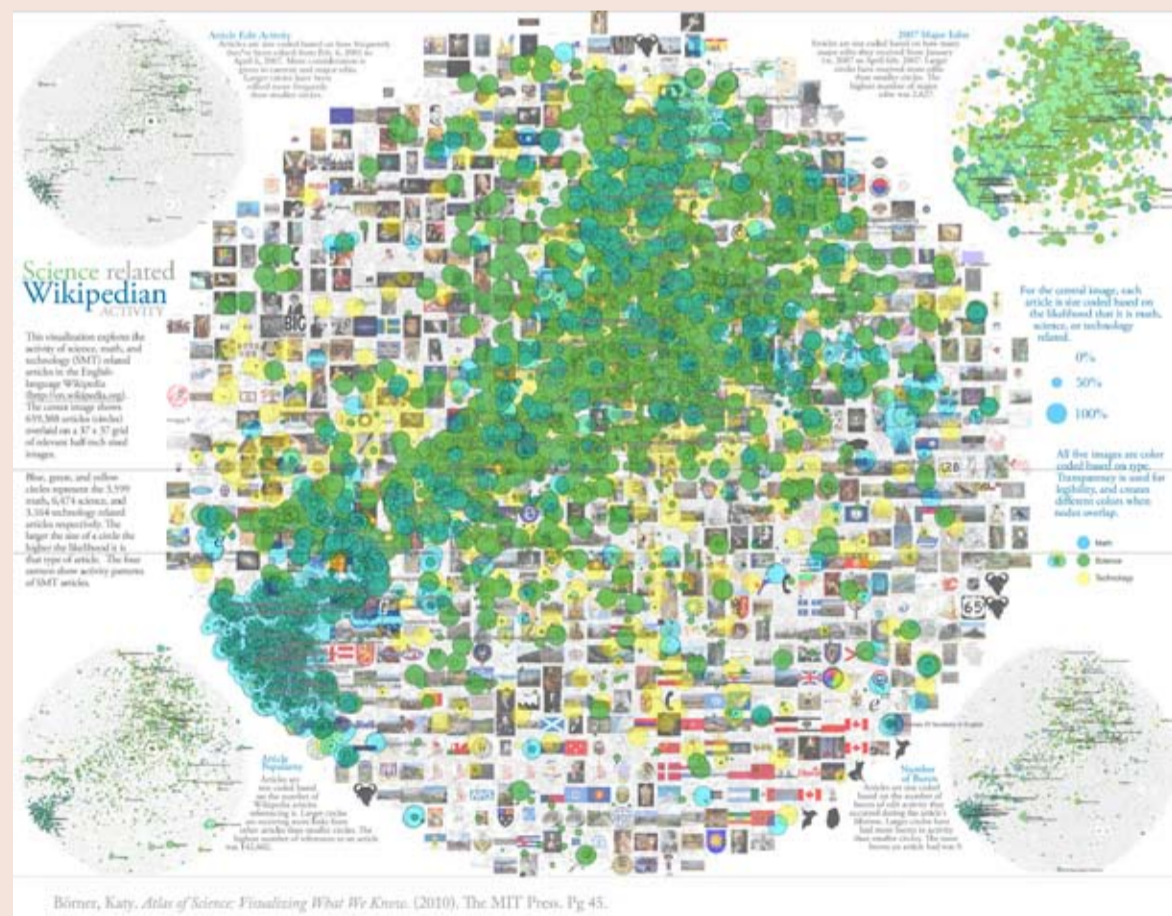
### Arbeidsvraag- en aanbodpanels



Begin maart zijn de *waves* van 2006 en 2008 van het Arbeidsaanbodpanel beschikbaar gekomen. Ook zijn onlangs de data van het Arbeidsvraagpanel 1991-2008 door het Sociaal en Cultureel Planbureau opengesteld voor onderzoekers.

De enquête van het Arbeidsaanbodpanel wordt sinds 1986 elke twee jaar afgenomen bij 4.500 personen in meer dan tweeduizend huishoudens. Er zijn inmiddels dertien metingen voor onderzoekers beschikbaar. Het Arbeidsvraagpanel verzamelt data over de vraag naar arbeid door organisaties zoals industriële bedrijven, handelsondernemingen onderwijsinstellingen, overheidsinstanties, zorg- en welzijnsinstellingen, etc. Beide bestanden kunnen direct en kosteloos worden gedownload uit DANS EASY. <https://easy.dans.knaw.nl>

## Wetenschappelijke activiteit op Wikipedia



Wikipedia staat bekend als de vrije encyclopedie met een hoog betrouwbaarheidsgehalte. Dat er ook veel wetenschappelijke activiteit is op Wikipedia is veel minder bekend. In het boek *Atlas of Science* van Katy Börner staat deze afbeelding die de (Engelstalige) wetenschappelijke

activiteit binnen Wikipedia in beeld brengt. Getoond worden 659.388 artikelen op een achtergrond van gekoppelde relevante afbeeldingen. Elke cirkel is een artikel. De kleuren blauw, groen en geel representeren respectievelijk de wiskundige (3.599 stuks), de wetenschappelijke (6.474

stuks) en technologische artikelen (3.164 stuks). Hoe groter de cirkel, des te groter de kans dat het om een wetenschappelijk artikel uit de betreffende categorie gaat. Bij een heel kleine kans wordt geen kleur toegekend en valt de cirkel bijna weg. Zie ook de rubriek Gelezen (p.8).

<http://scimaps.org/atlas/part2.html> (nummer 44-45 'Timeline 2006-2007' de twaalfde afbeelding)