

# Op de virtuele zeepkist

*Een week discussie in de Open Data Speakers Corner*



# Op de virtuele zeepkist

*Een week discussie in de Open Data Speakers Corner,*

*19 - 23 oktober 2009*



2009 DANS

© Sommige rechten zijn voorbehouden

Voor deze uitgave zijn gebruiksrechten van toepassing zoals vastgelegd in de Creative Commons licentie. [Naamsvermelding–NietCommercieel–Gelijkdelen] 3.0 Nederland.

Voor de volledige tekst van deze licentie zie <http://www.creativecommons.org/licenses/by-nc-sa/3.0/nl/>

DANS - Data Archiving and Networked Services

Postbus 93067

2509 AB Den Haag

T 070 3494450

F 070 3494451

[info@dans.knaw.nl](mailto:info@dans.knaw.nl)

[www.dans.knaw.nl](http://www.dans.knaw.nl)

ISBN 978-94-90531-03-4

Redactie: M. van den Dries, M. de Groot

Vormgeving: J. Buiten

Druk: Bejo druk & print, Alkmaar

Foto Peter Doorn: Wiebe Kiestra

## Woord Vooraf

Open Access wint snel terrein in de wetenschappelijke wereld en dat is zowel logisch als goed. De mogelijkheid iedereen kosteloos kennis te laten nemen van onderzoeksresultaten mogen we niet laten liggen. Er zijn immers nog heel veel onderzoekers - bijvoorbeeld buiten de grote onderzoekinstellingen en in ontwikkelingslanden - die daarvan kunnen profiteren. Zij kunnen zich de hoge kosten van betaalde tijdschriften niet veroorloven.

Wie Open Access steunt, zal niet lang hoeven nadenken over Open Data: de vrije beschikbaarheid van onderzoeksgegevens. De voordelen zijn evident. Onderzoekers kunnen elkaars conclusies kritisch beoordelen, een gekoesterd beginsel van wetenschapsbeoefening. Ze kunnen voortbouwen op elkaars resultaten of zelfs geheel nieuwe onderzoeksvragen beantwoorden met behulp van bestaande gegevens. Dat alles zonder de veelal hoge kosten van het verzamelen van data. Toch gaat deze ontwikkeling niet hard genoeg, vermoedelijk mede doordat de voordelen voor anderen groter zijn dan voor degenen die er moeite voor moeten doen.

Voor onze eigen organisaties zijn zowel Open Access als Open Data boven elke discussie verheven. Niet voor niets is DANS, dat streeft naar vrij en duurzaam beschikbare onderzoekgegevens, een gezamenlijk instituut van beide organisaties.

DANS heeft zijn roeping in de alfa- en gammawetenschappen op inspirerende wijze gevolgd door in de Open Access Week van 19 tot 23 oktober met de 'Open Data Speakers Corner' aandacht te vragen voor het belang van vrij beschikbare onderzoeksgegevens. Dit boekje bevat de resultaten van dat initiatief: een veelzijdige en boeiende discussie die wij graag aanbevelen aan eenieder die de wetenschap een mooie toekomst toewenst.

Robbert H. Dijkgraaf  
President Koninklijke Nederlandse  
Akademie van Wetenschappen

Jos Engelen  
Voorzitter Nederlandse organisatie  
voor Wetenschappelijk Onderzoek

## Inleiding

Van 19 tot 23 oktober 2009 werd de eerste internationale Open Access Week gehouden om te pleiten voor vrij toegankelijke en voor iedereen bereikbare onderzoekresultaten. In de Nederlandse universitaire wereld vonden in die week allerlei acties en bijeenkomsten plaats.

DANS zag in de Open Access Week een kans om de aandacht te vestigen op het belang van onderzoekdata binnen het streven naar vrije beschikbaarheid, een aspect dat niet altijd spontaan in de discussies over Open Access wordt betrokken. Dat gebeurde met behulp van de Open Data Speakers Corner, een speciale website waar elke dag van de week vanuit een andere invalshoek zou worden gediscussieerd over Open Data. Op maandag werd als perspectief dat van het wetenschapsbeleid gekozen, op dinsdag dat van de datagebruiker. Op woensdag was de invalshoek van de wetenschappelijke bibliotheek het uitgangspunt, op donderdag die van de wetenschappelijke uitgeverij en op de laatste dag van de Open Access Week stond het gezichtspunt van de dataproducent centraal.

Voor elke dag van de week werd een andere auteur gevraagd om met een key-note column de discussie te openen, en enkele discussianten om op die column te reageren alvorens de vloer vrij zou zijn voor ieder die zich in de discussie wilde mengen. Zo gaf een flink aantal bestuurders, onderzoekers en betrokkenen van naam acte de présence op de virtuele zeepkist, met een geanimeerde discussie als resultaat. Dit boekje biedt daarvan een weergave in print.



# Wetenschapsbeleid





*Column van Sijbolt Noorda, voorzitter VSNU*

## Stop de digitale slegs-vir-ons apartheid!

Wie schrijft, wil gelezen worden. Door zoveel mogelijk lezers. Wie wetenschappelijke artikelen schrijft wil gelezen worden, geciteerd worden, gebruikt worden.

In het digitale tijdperk verschijnen academische publicaties grotendeels in digitale vorm. Dat bevordert de verspreiding, de snelheid en het gebruiksgemak.

Digitale bibliotheken hebben heel veel meer gebruikers dan traditionele bibliotheken leners. Wie digitaal leesbaar en vindbaar is, heeft heel veel meer kans gelezen en geciteerd te worden dan vroeger. Toch is er een grens: wie geen toegang heeft tot een welvoorzien digitale bibliotheek heeft aan al die nieuwe mogelijkheden niets.

Vandaar de roep om Open Access. Zodat resultaten van wetenschappelijk onderzoek ook buiten de kring van de bevoorrechte instituties toegankelijk zijn, gebruikt kunnen worden. Eigenlijk is de roep om Open Access hetzelfde als de negentiende-eeuwse beweging voor openbare bibliotheken wilde: wat van algemeen nut is, moet algemeen toegankelijk zijn. Het is een merkwaardig contrast, die tegenstelling tussen digitale snelheid en wereldwijde verspreiding, en de feitelijk beperkte toegang voor velen die buiten de rijke academische centra werken.

Ik vind het verbazingwekkend dat wetenschappelijke auteurs niet massaal bezorgd zijn over deze apartheid. Het lijkt wel alsof ze zich niet realiseren dat het vanzelfsprekende gebruiksgemak van hun eigen welvoorzien digitale bibliotheken voor velen beslist niet vanzelfsprekend is. Een Brits onderzoek van enkele jaren geleden liet zien dat medici buiten

de academische centra nog niet de helft van de medische literatuur konden raadplegen of downloaden.

Ook de collega's die wel bezorgd zijn, menen dat het een zaak is waaraan ze zelf niets kunnen doen. Ze denken dat het iets is voor de bibliothecarissen, voor de financiers van het onderzoek, voor de uitgevers. En wachten af.

Voor beide groepen, de zorgelozen en de machtelozen, heb ik goed nieuws. Open Access is, ook buiten de bevoorrechte instellingen, gemakkelijk te realiseren. Op de doe-het-zelf manier. Zelf je digitale artikel uploaden naar de digitale bibliotheek van je eigen universiteit of instituut. In Nederland zijn die onderling verbonden en wereldwijd goed doorzoekbaar. Je bent meteen te vinden. En bijkomend voordeel: met een simpele link naar die digitale bibliotheek of repository is het artikel meteen leesbaar in onderwijsmateriaal, werkverslagen en onderzoeksvorstellen. Geen verwijzingen meer naar teksten die alleen voor bevoorrechte collega's toegankelijk zijn.

Mag dat dan zomaar? Ja dat mag, vaak expliciet, soms na een wachttijd van zes maanden, en altijd de facto. En de uitgevers dan? Die worden goed betaald voor het oprichten en instandhouden van tijdschriften, het organiseren van de redactie ervan en van de selectie van de artikelen. Niet om af te schermen wat door wetenschappers, voor het algemeen nut en wereldwijde verspreiding, met steun van de belastingbetaler, wordt geschreven.



*Reactie van Marlene Sikkens, jurist, ministerie van Onderwijs, Cultuur en Wetenschap*

## Laat opdrachtgevers heldere afspraken maken

Waar Sybolt Noorda in zijn column aan voorbij gaat, is dat ook de toegang tot de onderliggende onderzoeksdata waarop de wetenschappelijke schrijvers hun artikelen hebben gebaseerd, goed geborgd moet zijn. Alleen dan is verificatie van publicaties mogelijk en kunnen op grond van deze bron van data weer nieuwe (wetenschappelijke) artikelen worden gebaseerd.

Deze data moeten echter wel op een zodanige manier worden aangeleverd dat hergebruik mogelijk is en daar schort het momenteel aan. Met het door DANS gerealiseerde EASY-instrument voor Open Access is een belangrijke stap gezet. Maar totdat het op de juiste wijze deponeren van onderzoeksdata een wezenlijk onderdeel is van het universitaire opleidingsprogramma, is het zeker ook de taak van universiteitsbibliotheken om onderzoekers te ondersteunen bij het uploaden van hun data en artikelen in de betreffende repositories.

Open Access is overigens niet alleen van belang voor de wetenschappelijke wereld. Ook opdrachtgevers die onderzoek opdragen ten behoeve van hun beleidsontwikkeling, willen dat onderzoekers kunnen voortborduren op reeds beschikbare onderzoeksdata. Ook daarvoor is nodig dat onderzoekers zorgen dat de desbetreffende databanken (vaker en beter) gevuld worden en dat de opdrachtgevers bij het formuleren van onderzoeksopdrachten hier (vaker) heldere afspraken over maken.

Gaat dit vanzelf? Helaas niet. Zijn er stappen gezet die het proces versnellen? Jazeker!

OCW is bezig met een beleidsnota Wetenschappelijke Informatievoorziening, die begin 2010 naar de Tweede Kamer wordt gestuurd. De

toegankelijkheid van data en publicaties en de daarmee onlosmakelijk verbonden digitale duurzaamheid zijn hierin de pijlers.

De rijksoverheid heeft voor beleidsgericht onderzoek de Model onderzoeksovereenkomst ARVODI-2008 vastgesteld, waarin het deponeren door onderzoekers van onderzoeksgegevens bij DANS verplicht is gesteld. Hiermee hoopt de rijksoverheid de toon te zetten voor andere opdrachtgevers en onderzoekers om het belang van Open Access te benadrukken.



*Online comment: José de Kruif*

Ik citeer: "...die onderzoek opdragen ten behoeve van hun beleidsontwikkeling..." en "De rijksoverheid heeft voor beleidsgericht onderzoek de Model Onderzoeksovereenkomst ARVODI-2008 vastgesteld, waarin het deponeren door onderzoekers van onderzoeksgegevens bij DANS verplicht is gesteld. Hiermee hoopt de rijksoverheid de toon te zetten voor andere opdrachtgevers en onderzoekers om het belang van Open Access te benadrukken."

Ik voel wel wat mevrouw Sikkens wil zeggen, maar is dit beleidsproza nu echt de manier om een discussie in gang te zetten? Welk publiek hoopt mevrouw Sikkens te bereiken? Als onderzoeker-met-data word ik hier al helemaal schichtig van. Het riekt niet naar makkelijk data deponeren, maar naar knarsetandend formulieren invullen. En hoezo gaat een modelcontract met de aansprekende naam: ARVODI-2008 de toon zetten voor andere opdrachtgevers? Zouden die nu en masse de tekst van het modelcontract opvragen? Ik waag het te betwijfelen, maar hopelijk heb ik het mis en ben ik gewoon te zuur op de maandagochtend.

Dus niet getreurd, ik blijf de zeepkist (leuk initiatief!) de komende week natuurlijk volgen en zal verslag doen op mijn blog (<http://cms.let.uu.nl/ogcictonderzoek/>) als het wat aansprekender gaat worden. Tot die tijd ga ik mijn collega's niet lastig vallen met ARVODI-2008.

De vraag is ook: wat is dringender? De uitgevers aanpakken die ervoor zorgen dat een enorm deel van het budget van universiteitsbibliotheken opgaat aan abonnementen? Of data deponeren? Of is het allebei even dringend?

*Online comment: Herman Katteler*

De toon die de rijksoverheid heeft gezet met de vaststelling van de voor onderzoekers verbindende voorwaarden (ARVODI) is nu juist contraproductief als het gaat om het bewerkstelligen van een positieve sfeer rond Open Access. De onderzoeker/het onderzoeksinstituut is uitgekleeft door de bepalingen rond de intellectuele eigendomsrechten (artikel 23): 'Alle intellectuele eigendomsrechten die kunnen of zullen kunnen worden uitgeoefend... berusten bij de Opdrachtgever.' Enzovoort.

Voorwaar een dictaat van de rijksoverheid dat de leveranciers van inzichten buiten spel plaatst. Laat DANS zich vooral niet (uitsluitend) aan de zijde van de rijksoverheid scharen.

*Reactie van Ron Dekker, directeur Instituten, Financiën en Infrastructuur NWO*

## Als het kan, doe het dan!

Open Access als doe-het-zelf methode. Makkelijker kunnen we het niet maken. Of toch wel?

Noorda legt het probleem van toegang tot publicaties bij de onderzoeker neer: die hoeft slechts zijn paper in de universitaire repository te deponeren en de hele wereld kan er kennis van nemen.

Een klein detail is dat die versie vaak niet de eindversie is zoals die in het tijdschrift verschijnt. Dat is iets met copyright, wat de onderzoeker zonder zorgen afstaat aan de uitgever. Maar door deze versieverschillen loopt hij kans dat het aantal citaties verwatert. En vervolgens wordt hij – en met hem zijn faculteit en universiteit – minder goed beoordeeld omdat de impactscores en de citaties te wensen overlaten.

Moeten we dan maar geen Open Access stimuleren? Integendeel!

Mijn credo is: Open Access: als het kan, doe het dan!

Maar de zorgelozen moeten meer bewust worden gemaakt en de machtelozen moeten handvatten krijgen om makkelijker publiek te publiceren. Dat kan bijvoorbeeld door onderzoekers erop te wijzen dat ze hun copyright niet zo maar moeten weggeven, maar vervangen door een licentieovereenkomst (die aanklikbaar is op ieders bureaublad – inderdaad, makkelijker kunnen we het niet maken).

En dan nog zijn we er niet: lezers, en zeker niet-ingewijden, moeten worden geholpen om hun weg te vinden in de vakliteratuur. Bijvoorbeeld met samenvattingen voor leken, zoals patiëntenorganisaties nu al doen.

Dat toelichten waar het over gaat is nog meer van belang voor data, die net als publicaties ook vrij

toegankelijk zouden moeten zijn.

Om hergebruik van dure databestanden mogelijk te maken, is een eerste vereiste dat goed wordt beschreven waar de data over gaan, hoe de structuur van het bestand is, etc. DANS geeft bijvoorbeeld een kwaliteitszegel op deze metadata af, zodat gebruikers weten dat ze met een goed product werken.

Voor databestanden is ook belangrijk dat er een gebruiksaanwijzing mee komt, een toelichting over do's en don'ts. Hoe mis dit anders kan gaan wil ik toelichten aan de hand van een voorbeeld: in data werd er een positief verband gevonden tussen 'bovenin een flat wonen' en 'gezondheid'. Nader onderzoek wees echter uit dat er een verbindende variabele was: 'jonge leeftijd'. Gelukkig maar, want anders had Nederland vol gestaan met torenflats.



*Online comment: Ekkehard Mochmann, Keulen*

“Ik wil gelezen worden!”

Geen twijfel mogelijk, ik ben voor OA en ik onderschrijf tevens de roep om toegang tot data. Op dit moment heb ik slechts twee bedenkingen om het proces in stand te kunnen houden. Ik houd het kort want:

1. “Iedereen wil gelezen worden.” Ja! Velen willen begrepen worden, en sommigen zullen impact hebben! De klassieken behandelden kwesties van realiteit en werden gelezen door mensen die richting zochten, dus het ging niet alleen om wetenschappelijke concurrentie en impactfactoren, maar ook over een impact hebben op de maatschappij.

2. Als je op het punt staat om Open Access te implementeren, denk dan aan de exponentieël groeiende bergen en ravijnen die overstromen met informatie en data als iedereen zijn producten dumpt. Zoals de Romeinen al zeiden: als je begint te handelen, doe het dan met wijsheid en houd het einde in gedachten. Misschien waren ze een beetje te beperkend: we zouden zeker graag meer documenten en gegevens uit hun tijd willen hebben, of was het misschien dat ze niet genoeg aandacht gaven aan lange termijn opslag, selectiviteit en interoperabiliteit, en te weinig investeerden in metadata?

Ondersteun Open Access, investeer en houd in gedachten: tijd voor lezen en voor data analyses is een kostbaar goed.

*Reactie van Peter Doorn, directeur DANS*

## Open Access: data delen tussen droom en daad

In zijn column tegen de slegs-vir-ons apartheid betoont Sybolt Noorda zich een ware Nelson Mandela van de Open Access. Hij roept alle onderzoekers in Nederland (en daarbuiten) op om hun digitale publicaties te uploaden naar de digitale bibliotheek van hun eigen universiteit of instituut. Dat kunnen ze gemakkelijk zelf doen, en bevordert het adagium dat Multatuli al van de daken schreeuwde: ik wil gelezen worden.

Tegen Noorda's oproep wil ik niets inbrengen. Ik wil hem zelfs uitbreiden. Wetenschappers aller landen, laad niet alleen uw publicaties op, maar doe dat ook met de databestanden waarop u uw onderzoek gebaseerd heeft. Dat kan net zo gemakkelijk, en ook op de doe-het-zelf manier: in het Elektronisch Archief EASY van DANS.

Toch hebben nogal wat onderzoekers twijfel, of misschien wel huiver, bij het delen van hun onderzoeksdata. Gelezen wil iedereen wel worden, maar geheranalyseerd? Daar staat in het algemeen geen premie op: het kost tijd en moeite om een dataset netjes te documenteren; iemand anders kan over de gegevens publiceren die met veel moeite door jou verzameld zijn. Erger nog, iemand kan een fout in je data vinden, of in de door jou toegepaste analyse. Of die ander begrijpt de ingewikkelde manipulaties niet die nodig zijn om verantwoord met de data te werken. Of zij verwijst helemaal niet eens naar jouw onderzoek. En ten slotte, een publicatie telt op je CV en in het lijstje dat je jaarlijks moet opgeven, maar een databestand?

Bij DANS hebben we alle argumenten tegen het delen van data al vaak gehoord. We zijn het er niet altijd mee eens, maar we hebben er wel begrip voor. Immers, tussen droom en daad staan

wetten in de weg en praktische bezwaren. Daarom kunnen onderzoekers die een bestand uploaden in EASY zelf bepalen onder welke voorwaarden anderen aan hun data kunnen komen. Open toegang is één mogelijkheid, maar het is ook mogelijk dat een embargo-periode wordt vastgelegd, of dat de datagebruiker eerst toestemming moet krijgen van de eigenaar.

Er zijn verschillende mogelijkheden om de bereidheid om data te delen te bevorderen. Het kan bijvoorbeeld een voorwaarde zijn voordat subsidie wordt verstrekt, zoals bij veel NWO-subsidies. Maar ik denk dat het heel goed zou zijn als digitale producten zouden meetellen op het publicatielijstje. Databestanden en zelf ontwikkelde software, voor zover beschikbaar in een open repository, zouden daar in ieder geval bij moeten horen.





# De datagebruiker





Column door Jan Luiten van Zanden, economisch historicus, Internationaal Instituut voor Sociale Geschiedenis

## De lange weg naar 'echte' Open Access<sup>1</sup>

Onlangs probeerde ik data los te krijgen van een van de grote namen in mijn vak, economische geschiedenis. Ik zal geen namen noemen maar het is echt een kanjer, die heel veel vernieuwend, kwantitatief onderzoek heeft gedaan. Voor een vergelijking op wereldschaal kon ik zijn data heel goed gebruiken, en het lukte me niet te reproduceren wat er door hem gedaan was. Het bleek dat de data er niet meer waren. Na acht jaar gooide hij alles weg; dat deed hij met zijn belastinggegevens ook zo.

Zo is het lot van de onderzoeker die de wereld afstroopt op zoek naar data. Data om de geschiedenis van de ongelijkheid op wereldschaal te schrijven – om uit te zoeken waarom sommige landen arm zijn en andere rijk. Economisch-historici stellen in toenemende mate vragen op deze schaal, en hebben grote moeite om de gegevens bij elkaar te krijgen waarmee ze gefundeerde antwoorden kunnen geven. Verhalen hierover zijn er natuurlijk voldoende in omloop, maar het is juist de vraag hoe we van deze verhalen komen naar de systematische toetsing van sociaalwetenschappelijke theorieën op dit terrein.

Collectief, als groep, weten we heel veel over de wereldeconomie. Er zijn specialisten die alles weten over India in de negentiende eeuw, of China rond het jaar 1000, die soms zelfs bereid zijn om de best denkbare schattingen te maken van de belangrijkste sociaal-economische en demografische grootheden. Maar om inzicht te krijgen in wat er in de hele wereld gebeurde moet deze informatie op een of andere manier

gepoold worden. Hoeveel lag bijvoorbeeld in de bloeiperiode van de Song zo'n duizend jaar geleden China voor op de rest van de wereld? Of was juist het Midden-Oosten, opgebloeid na de verovering door de Islam, het centrum van de toenmalige wereldeconomie?

Er is een middel ontwikkeld om dit poolen mogelijk te maken – de zogenaamde *collaboratory*. Het idee is prachtig, maar het vraagt een hele investering in samenwerking om dit op gang te brengen. Het is een kwestie van lange adem, en de kunst is om een project zo te construeren dat deelnemers al vanaf het begin resultaten kunnen boeken die er toe doen. Want je moet sleutelonderzoekers – zoals de genoemde 'kanjer' – met positieve prikkels overhalen om mee te werken; daar geloven economen nu eenmaal in.

Open Access is, dat blijkt hieruit, nog maar een eerste stap op weg naar echte samenwerking in onderzoek. De meeste collega's zijn wel bereid om data te delen – maar ze moeten daar soms tijdig aan herinnerd worden. Wetenschappelijke tijdschriften zouden hier een rol in moeten spelen, en een Data Availability Policy (DAP) moeten ontwikkelen: een artikel zou alleen gerefereerd, of alleen gepubliceerd kunnen worden, als de data beschikbaar worden gemaakt voor herhaling van de experimenten waarover bericht wordt (meestal regressies van de betreffende databestanden). Eigenlijk zou dit een normale wetenschappelijke standaard moeten zijn, maar het kost toch veel moeite om dit idee ingevoerd te krijgen – ook omdat redacties niet altijd weten hoe ze het moeten organiseren.

Een ander, even onmisbaar onderdeel van een



echte Open Access strategie is het permanent en toegankelijk bewaren van de betreffende databestanden. DANS doet daar in Nederland veel aan - al kan het nog steeds beter - maar internationaal is dit veel minder goed georganiseerd. Als je dat allemaal goed voor elkaar hebt - de collabs, de DAP en de archivering - kunnen we echt van Open Access gaan spreken. En kunnen we, nog belangrijker, veel betere antwoorden gaan geven op heel oude en talrijke nieuwe vragen. Plannen om tot zo'n infrastructuur te komen zijn overigens in voorbereiding. Kijk maar eens naar het CLIO INFRA voorstel op [www.clio-infra.eu/index.php/Main\\_Page](http://www.clio-infra.eu/index.php/Main_Page).

<sup>1</sup> Dit stukje is grotendeels gebaseerd op ideeën ontwikkeld in Tine De Moor en Jan Luiten van Zanden: 'Do ut Des: laboratories as a 'new' method for scholarly communication and cooperation for global and world history'; *Historical Methods*, 41(2)2008, pp. 67-78.

*Reactie van John Mackenzie Owen, professor of Information Science, Universiteit van Amsterdam*

## Organiseer Open Access langs disciplinaire lijnen

De vrije toegang tot onderzoeksdata is nog weinig aan de orde geweest in de discussie rond Open Access (OA). Dat heeft een aantal oorzaken. Een daarvan is dat in veel vakgebieden (zeker in de exacte en sociale wetenschappen) er al lang de praktijk bestaat dat je onderzoeksdata in een repository deponert en daarheen verwijst in je publicaties. Daar waar het deponeren van data heel gewoon is, is er weinig behoefte aan discussie.

Een tweede reden voor de de gebrekkige aandacht voor data in de context van Open Access, is dat het OA-debat nogal verzand is in een ideologische actie van mensen die zich hebben gekeerd tegen de commerciële uitgeverij die schandalig veel zouden verdienen aan informatie die voortkomt uit door belastinggeld gefinancierd onderzoek en die aan onderzoekers hun auteursrecht ontnemen om er zelf rijker van te worden. Gelukkig hebben uitgeverij zich nooit met onderzoeksdata beziggehouden, en dat moeten we vooral zo laten. Het zou heel ongelukkig zijn als uitgeverij zelf repositories zouden opzetten en van auteurs zouden eisen dat ze hun data – inclusief de rechten – aan hen afstaan.

Een nieuwe ontwikkeling, waar onder meer SURF zich voor inzet, zijn 'verrijkte publicaties' waarbij er een directe koppeling is tussen publicaties, data, software, multimedia e.d. Ook dit is een ontwikkeling die vooral in de context van Open Access moet plaatsvinden, en niet in handen moet vallen van commerciële dienstverleners.

Open Access gaat echter veelal mank aan zaken als onvolledigheid, gebrekkige kwaliteitsbewaking, slechte ontsluiting en vooral ook versnippering door tal van ongecoördineerde activiteiten. De oplossing daarvoor ligt in organisatie

langs disciplinaire (dus niet institutionele) lijnen, en op basis van internationale samenwerking (dus niet ieder land z'n eigen ding). Die weg is inderdaad nog lang.



*Reactie van Kees Aarts, hoogleraar Politicologie, Universiteit Twente*

## Altruïsme gevraagd

In 1985 rapporteerden drie economen in het *Journal of Money, Credit and Banking* over hun systematische pogingen om publicaties in voorafgaande jaargangen van dat tijdschrift te repliceren. Gevraagd om de data en analyseprogramma's, gaf tweederde van de aangeschreven auteurs niet thuis. Bij die artikelen waar replicatie wel mogelijk was, leidde dit vaak tot geheel andere resultaten. Slechts in een klein aantal gevallen werd duidelijk wat de oorzaak van deze verschillen was.

Eerstejaars studenten krijgen te horen dat wetenschappelijke kennis altijd voorlopig is, totdat er iets beters is bedacht. Kennis die voor altijd vast staat, die immuun is voor kritiek, is niet wetenschappelijk maar dogmatisch. Replicatie en verificatie behoren tot de belangrijkste strategieën om de geldigheid en betrouwbaarheid van kennis vast te stellen.

Terecht wijst Jan Luiten van Zanden op het belang van Data Availability Policies bij tijdschriften. Zonder uitdrukkelijke aandacht voor replicatie, verificatie en meta-analyse (en dus voor het openbaar beschikbaar maken van data) dreigen de alfa- en gammawetenschappen voortdurend weer in een 'dogmatische sluimer' weg te zakken.

Het belang van collaboratories is eveneens volstrekt duidelijk. Zeker op het gebied van surveyonderzoek zijn hierbij al veel resultaten geboekt. Kijk eens op: [www.cses.org](http://www.cses.org), [www.europeansocialsurvey.org](http://www.europeansocialsurvey.org), of [www.europeanvaluesstudy.eu/](http://www.europeanvaluesstudy.eu/).

Stuk voor stuk monumenten van samenwerking, die zo snel mogelijk na de dataverzameling vrij beschikbaar zijn voor iedere onderzoeker.

De vraag is hoe onderzoekers nog beter kunnen worden geprikkeld om hun steentje bij te dragen. Het gaat om de productie van collec-

tieve goederen. De geëigende instrumenten daarvoor zijn de stok, de wortel en de preek. De stok is in handen van subsidiegevers zoals NWO (zonder data te deponeren, geen geld voor onderzoek) en van uitgevers en tijdschrift redacties (zonder data geen publicaties). De wortel wordt voorgehouden door de gemeenschap van onderzoekers en studenten, die samen met de data-archieven meerwaarde creëren uit verzamelde gegevens en daarbij hopelijk niet vergeten hun bronnen te vermelden met naam en toenaam. De preek blijft ook nodig. Zonder een cultuur waarin het documenteren en deponeren van data vanzelfsprekend is, komen we er niet. Die cultuur vereist een grote dosis altruïsme bij de onderzoekers. Zijn we bereid dat op te brengen?



*Online comment: James Stell, business development manager*

Het is dramatisch als data waar jaren aan gewerkt is worden vernietigd. Er zullen ongetwijfeld veel redenen voor zijn zoals (elektronisch) ruimtegebrek, geen goede afspraken vooraf over de te bewaren data en dit terwijl de data veelal nog nodig zijn voor andere onderzoeken zoals trendanalyses. Je hebt dan niks aan het zogenaamde 'van horen zeggen'. Je hebt harde feiten nodig. Een digitaal bewaardepot was 'vroeger' misschien een ver-van-je-bed show, nu zou het een logisch onderdeel moeten zijn van interne processen ook wel e-depot genoemd. Het is vandaag de dag gewoon betaalbaar maar de klassieker gaat hier nog steeds op: 'Onbekend maakt onbemind' en dat is doodzonde voor de toekomst van de wetenschap.

# De wetenschappelijke bibliotheek



*Column door Bas Savenije, algemeen directeur Koninklijke Bibliotheek*

## Naar een intensiever gebruik van wetenschappelijke informatie

Open Access is here to stay. De Open Access beweging heeft de afgelopen jaren momentum gekregen. Het is duidelijk dat dit de wetenschappelijke informatievoorziening in steeds sterkere mate gaat bepalen.

Een belangrijk element van Open Access is de vergroting van het gebruik van wetenschappelijke informatie. Dat maakt Open Access aantrekkelijk voor de auteurs (meer impact) en dit punt wordt dan ook naar wetenschappers vaak gebruikt als argument voor Open Access publiceren. Maar het is minstens even aantrekkelijk voor de lezer. En dan bedoel ik in het bijzonder die lezer die vanwege de hoge prijzen anders niet in staat zou zijn om de informatie tot zich te nemen.

De zichtbaarheid van Open Access tijdschriften is goed te noemen, mede dankzij de Directory of Open Access Journals (DOAJ). Veel wetenschappelijke bibliotheken hebben de Open Access tijdschriften opgenomen in hun tijdschriftenaanbod. Maar hoe zit het met repositories? Iedere zichzelf respecterende universiteit doet tegenwoordig zijn best zoveel mogelijk van de wetenschappelijke output van de instelling in de repository op te nemen en plaatst dit op de website van de universiteit. Meestal heeft de universiteitsbibliotheek hierbij een cruciale rol. NARCIS maakt de totale Nederlandse output zichtbaar. Maar als een argeloze Nederlander bijvoorbeeld gecertificeerde medische informatie wil: weet hij wel waar hij heen moet? Het is om te beginnen al een hele uitdaging voor de Nederlandse bibliotheken (wetenschappelijke en openbare) om iedere Nederlander een eenvoudige en snelle toegang te geven

tot wetenschappelijke informatie. Als men niet onder een licentie valt, zal daar meestal voor betaald moeten worden. Soms is dat veel, soms valt dat mee. Zowel de slechte vindbaarheid als de prijs vormen een belemmering voor het gebruik. Er is daarom een gerichte inspanning nodig om ervoor te zorgen dat iemand die naar wetenschappelijke informatie zoekt, in zijn zoekproces min of meer vanzelf ook de Open Access informatie krijgt. Dat kan een artikel zijn uit een Open Access tijdschrift, een kopie van dat artikel vanuit een repository, of de finale auteursversie van dat artikel wanneer de uitgever geen toestemming geeft de uitgeversversie in een repository op te nemen. Als deze zoekresultaten in een en dezelfde zoekactie beschikbaar komen als de 'betaalde toegang' publicaties, dan gaan we pas echt profiteren van de resultaten van de Open Access beweging.



*Reactie van Maria Heijne, directeur TU Delft Library*

## Aandeel van wetenschapper ligt aan de basis van elk succes

“Iedere Nederlander toegang tot wetenschappelijke informatie” vind ik een noodzakelijk streven voor de komende jaren. De Nederlandse bibliotheken moeten hier in gezamenlijkheid een rol spelen en hun inspanningen richten op goede toegankelijkheid. Open Access als ‘het’ nieuwe business model is daarin cruciaal maar voorlopig voor de bulk van hoogwaardig wetenschappelijk materiaal nog geen werkelijkheid.

Daarom moeten bibliotheken de handen vooral praktisch ineenslaan en

1. zich richten op het verzamelen, in de repositories, van alle wetenschappelijke output in hun instelling die op kwaliteit getoetst en vrij toegankelijk is. Dus niet alleen artikelen, maar ook courseware, data etc. Uiteraard moet dat wereldwijd gebeuren en moeten al die repositories met elkaar verbonden worden. In een eerdere bijdrage door John Mackenzie Owen wordt gepleit voor een verzameling langs disciplinaire lijn. Dat zou mooi zijn, maar praktisch veel lastiger te verwezenlijken.
2. hun repositories gastvrij openstellen voor andere instellingen waar producenten van wetenschappelijk materiaal werken en stimuleren dat die ook hun output via die repositories beschikbaar stellen.
3. hier inderdaad, zoals Savenije betoogt, een goed zoekstelsel onder leggen waarmee iedereen zeer gemakkelijk kan zoeken en vinden. Op zoekgebied vinden diverse ontwikkelingen plaats, zowel in de wetenschappelijke als de openbare bibliotheken. Binnen het landelijk consortium ‘Gemeenschappelijke Informatie Infrastructuur’ worden deze ontwikkelingen ook gestimuleerd en ik verwacht daar binnen korte tijd ook goede

resultaten van.

4. in goed overleg met overheid en uitgeverij komen tot ‘overgangsregelingen’ in afwachting van de (in mijn ogen onvermijdelijke) transitie naar het wereldwijde Open Access model. Ik denk dan aan aanvullende regelingen op de bestaande licentiecontracten en/of aan een mogelijk Nationaal Licentiemodel op uitgeverijbestanden waarmee toegang voor ‘iedere Nederlander’ nu al mogelijk wordt.

Maar bibliotheken kunnen nog zo hun best doen om vrij toegankelijk materiaal te verzamelen en goed toegankelijk te maken, het aandeel van de wetenschapper hierin ligt aan de basis van elk succes. Zolang zij vooral prudent met hun auteursrecht omgaan is er al veel gewonnen. Het auteursrechtinformatiepunt bij de universiteitsbibliotheek kan alle vragen daarover beantwoorden.





*Reactie van Andrew Treloar, adjunct-directeur Australian National Data Service*

## Laten we onze energie richten op alle onderzoeks-resultaten

Het lijkt er inderdaad op dat, zoals Bas Savenije het zegt, 'Open Access is here to stay'. Maar Open Access tot wat eigenlijk?

Tot nu toe heeft de Open Access beweging zich voornamelijk gericht op de wetenschappelijke publicatie als belangrijkste onderzoeksresultaat. In het verleden was dit misschien nog wel te verdedigen, toen datagestueurd onderzoek tot een klein aantal disciplines beperkt bleef. Maar alle vakgebieden zijn inmiddels data-intensief aan het worden. Er zijn genoeg voorbeelden aan te halen: de etnografie heeft grote videocollecties, de sociale wetenschappen hebben archieven vol kwalitatieve en kwantitatieve gegevens, de biologie produceert en moet toegang hebben tot gensequenties, en de linguïstiek werkt met veel teksten.

Een publicatie hangt in toenemende mate van de onderliggende data af en kan niet goed genoeg worden beoordeeld of vol vertrouwen als basis worden gebruikt om verder mee te werken zonder toegang te hebben tot die data.

Bovendien zijn in sommige gevallen de gegevens zelf net zo zeer een eerste klas onderzoeksresultaat als de publicatie. Maar voor wie gegevens schept en beheert, bestaan nog niet dezelfde soort beloningen als voor degenen die artikelen en papers voor congressen schrijven. Evenmin hebben we eenzelfde kwaliteit infrastructuur voor gegevens (met noemenswaardige uitzonderingen, zoals DANS) als we voor artikelen hebben, met zaken zoals repositories op het instituut en innovatieve interfaces zoals NARCIS.

Laten we, als we ons richten op de

voordelen van Open Access tot onderzoeksresultaten, dan ook niet voorbijgaan aan het belang van data. Laten we ervoor zorgen dat we onze energie richten op alle resultaten die nodig zijn om er mee verder te kunnen werken, om zo daadwerkelijk te komen tot 'een intensiever (en effectiever!) gebruik van wetenschappelijke informatie'.



*Reactie van Kurt De Belder, bibliothecaris van de Universiteit, directeur Universitaire Bibliotheken Leiden & Leiden University Press*

## Open Access, een bouwsteen voor de transformatie van wetenschappelijke communicatie

Open Access biedt inderdaad, zoals Bas Savenije stelt, het perspectief op een brede wereldwijde toegankelijkheid van wetenschappelijke informatie en een grote zichtbaarheid en impact van onderzoeksresultaten.

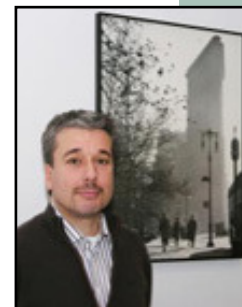
Daarnaast is Open Access ook een voorwaarde om nieuwe vormen van wetenschappelijk onderzoek mogelijk te maken. Het is onvoldoende dat digitale teksten, data en beelden alleen toegankelijk zijn voor het 'traditionele' proces van kennisacquisitie: het lezen van informatie door een mens. Daarnaast behoort dit materiaal ook toegankelijk te zijn voor de computer als 'lezer', als onderzoeksinstrument. Het digitaal materiaal dient onttrokken te worden aan de digitale silo's waarin het opgesloten ligt. Het op grote schaal kunnen indexereren, doorzoeken, koppelen, manipuleren en masseren van digitale informatie maakt nieuwe vormen van wetenschappelijk onderzoek mogelijk. Hierbij kunnen technieken zoals text- en datamining ingezet worden.

Open Access stimuleert hierdoor niet alleen een intensiever 'traditioneel' gebruik van wetenschappelijke informatie maar ondersteunt ook nieuwe digitale mogelijkheden voor wetenschappelijk onderzoek en communicatie.

Dit zijn dan ook de redenen waarom bibliotheken publicatie- en datarepositories verder uitbouwen, samen met onderzoekers en uitgevers werken aan de ontwikkeling van verrijkte publicaties, en virtuele kenniscentra opzetten waarin leden van een onderzoeksgroep kunnen samenwerken en digitale instrumenten kunnen inzetten voor hun onderzoek.

Maar Open Access was toch niet alleen hiervoor bedoeld? Open Access als publicatie- of businessmodel had/heeft ook de ambitie om wetenschappelijke informatie betaalbaar te houden. Dit is nog steeds een lonkend perspectief want de huidige businessmodellen van commerciële uitgevers met name in de STM-hoek (Science, Technology, Medicine) worden onhoudbaar voor universiteiten en onderzoekinstellingen. De jaarlijkse prijsstijgingen voor de grote tijdschriftpakketten (in de laatste tien jaar met minimaal 5% per jaar) zal in de komende jaren brede toegang tot wetenschappelijke informatie meer en meer in de weg staan vooral wanneer de mogelijkheid om binnen deze pakketten flexibele keuzes te maken sterk beperkt wordt. Een verschraving van toegang tot wetenschappelijke informatie en een verdere concentratie van marktpartijen zijn niet in het belang van wetenschappelijk onderzoek en de kenniseconomie.

Vervult Open Access de initiële belofte dat dit model juist ook een stap was om de kosten van wetenschappelijke publicaties beheersbaar te houden? Wat mij betreft is dit nog niet helemaal duidelijk. Twee recente onderzoeksrapporten (van Houghton/SURF en van het Research Information Network) die op macroniveau de totale kosten van wetenschappelijk informatie trachten te becijferen, komen globaal tot niet helemaal vergelijkbare conclusies. In het ene rapport betekent de wereldwijde implementatie van het Open Access model een significante reductie van kosten. Het andere rapport



stelt dat dit min of meer tot een nuloperatie zal leiden maar dat er wel een significante verschuiving van kosten richting onderzoeksintensieve landen en instellingen zal plaats vinden. Dit behoeft verder onderzoek. De baten van open en brede toegankelijkheid van wetenschappelijke informatie zijn duidelijk, bekeken moet worden welke partijen binnen een kenniseconomie hieraan een (financiële) bijdrage kunnen leveren.

Er zijn wat mij betreft voldoende redenen om de weg naar Open Access krachtdadig verder te zetten. Maar daarnaast dienen we ook de discussie aan te gaan over de nauwe relatie tussen wetenschappelijk uitgeven en onderzoek. Waarbij de uitgeefpraktijk in dienst hoort te staan van de wetenschap. Open Access is een belangrijke bouwsteen voor de noodzakelijke transformatie van wetenschappelijke communicatie maar is wellicht niet de steen der wijzen.

*Online comment: Laurens Mommers*

Er zijn verschillende krachten die Open Access kunnen stimuleren:

- De belangen van de maatschappij, waarvoor resultaten van met collectieve middelen betaald onderzoek drempelloos beschikbaar zouden moeten komen;
- De beperkende auteursrechtelijke arrangementen en die traditionele wetenschappelijke uitgevers opleggen aan hun auteurs;
- De hoge kosten van veel wetenschappelijke uitgaven;
- De belangen van wetenschappers om hun werk zo goed mogelijk vindbaar te maken.

De daadwerkelijke toegankelijkheid van wetenschappelijke publicaties hangt echter niet alleen af van de vindbaarheid en vrije toegang op internet, maar uiteraard ook van het vermogen van diverse 'publieken' om daadwerkelijk de inhoud tot zich te kunnen nemen. Wat dat laatste betreft ben ik een stuk minder optimistisch dan over de technische realisatie van de toegang.

In de discussie (zie bijdrage Treloar) valt op dat de toegankelijkheid van onderliggende data als obstakel wordt gezien. In het juridische domein in Nederland, waar Open Access tijdschriften nog nauwelijks van de grond zijn gekomen, lijkt een gespiegelde situatie te zijn ontstaan: de onderliggende 'data', in het juridische domein de rechtsbronnen, zijn juist wel voor een steeds groter deel vrij beschikbaar.

Mede voortvloeiend uit het beleid van de Europese Unie met betrekking tot hergebruik van overheidsinformatie zijn in Nederland belangrijke delen van Kamerstukken, wetgeving en rechtspraak gratis te raadplegen. Ook de toegankelijkheid daarvan wordt echter in niet onbelangrijke mate bepaald door de bruikbaarheid van de sites waarop ze gepubliceerd

worden en de 'leesbaarheid' van de informatie (zie over het laatste mede <https://openaccess.leidenuniv.nl/dspace/handle/1887/13525>).

Hergebruik, ook door commerciële partijen, is van groot belang om de vindbaarheid van bronnen te optimaliseren. De licentiëring van Open Access artikelen dient op dit punt niet te beperkend te zijn. Ik verwijs hierbij graag naar het initiatief tot de gratis site <http://www.liigl.nl>, dat het bedrijf waar ik werk naast mijn universitaire functie heeft geïntroduceerd om verschillende rechtsbronnen naast elkaar eenvoudig toegankelijk te maken.



# De wetenschappelijke uitgeverij



*Column door Rafael Sidi, vice-president of product development for Elsevier's ScienceDirect*

## Door middel van Open Data de wetenschappelijke innovatie versnellen

Openheid met betrekking tot onderzoeksgegevens is een snel opdoemende trend, die zich over het internet verspreidt van Best Buy's Remix tot de Britse krant de Guardian. Recent daagde Sir Tim Berners-Lee, de man achter het World Wide Web, de wetenschappelijke gemeenschap uit met een oproep tot actie: "raw data now". Hoewel zijn opmerkingen meer waren gericht op de behoefte aan het openbaar beschikbaar maken van gegevens van de overheid, krijgt het pleidooi voor een grotere transparantie van alle wetenschappelijke data steeds meer wind in de zeilen.

De democratisering van ruwe data bevat beslist de belofte van versnelde innovatie – het onderzoekers mogelijk maken om voort te bouwen op het werk van vakgenoten over de hele wereld, en gegevens te hergebruiken en herordenen en zo verdere doorbraken en ontdekkingen mogelijk te maken. Voor een voorbeeld van hoe het idee van Open Data de wetenschap voortstuwt, hoef je niet lang te zoeken. Na de lancering in 1990 stelde het Menselijk Genoom Project (HGP) zijn databanken voor het publiek open. Als gevolg daarvan slaagde het HGP er in om in 2003 de sequentie van alle basenparen in het menselijke genoom ruim binnen het budget voor elkaar te hebben, daarbij meer dan twee jaar op de planning voorlopend. Een vergelijkbaar voorbeeld in de commerciële wereld is NextBio dat toegang geeft tot een beheerde en gecorrleerde databank met experimentele data van bronnen als de Gene Expression Omnibus (GEO), ArrayExpress en caBIG.

Terwijl overheidsinstellingen in Groot-Brittannië, de Verenigde Staten en Australië momenteel aan kop lopen in de Open Data beweging, wint

het concept overall in de wetenschappelijke gemeenschap aan vaart. Nu onderzoek steeds meer multidisciplinair wordt en er ook steeds meer wordt samengewerkt, zal de beschikbaarheid van ruwe gegevens zich ontpoppen als cruciale component in het aanwakkeren van wetenschappelijke ontdekkingen. Makkelijke toegang tot de datasets is echter niet genoeg. Nog steeds zijn bewezen strategieën voor het creëren van zinvolle koppelingen tussen data nodig. Hoe meer connecties en relaties er kunnen worden aangebracht tussen openbare datasets, hoe steviger de funderingen voor wetenschappers om lagen aan te brengen, uit te breiden en erop voort te bouwen. Bovendien kan het hogere aantal koppelingen helpen om een licht te doen schijnen op innovatieve toepassingen die mogelijk niet zichtbaar zijn als de data afzonderlijk worden onderzocht. Als gegevens eenmaal kunnen worden vermengd met andere onderzoeksinformatie, zal dit tot veel wetenschappelijke innovaties leiden.

Een andere uitdaging van Open Data is de aarzeling bij wetenschappers om hun met pijn en moeite verkregen data te delen met concurrerende onderzoekers. Er begint een dualiteit zichtbaar te worden in de wetenschappelijke gemeenschap van meer samenwerking in een omgeving waarin de concurrentiedruk steeds hoger wordt. Velen worstelen om de juiste balans te vinden. Kunnen beloningen als academische 'credits' voor wetenschappers worden gemaakt om het delen van data aan te moedigen?

De beweging in de richting van Open Data kan worden opgevat als een bron van efficiëntie voor wetenschappers en commercie. Nu ze niet opnieuw het wiel moeten uitvinden en datasets



moeten repliceren die andere onderzoekers al hebben opgebouwd, kunnen wetenschappers hun voordeel doen met deze openbaar toegankelijke bouwstenen en sneller toekomen aan het ontdekkende stadium van hun werk. Er moeten nog veel uitdagingen worden overwonnen, maar de voordelen van het openstellen van wetenschappelijke data – meer samenwerking en een hogere onderzoeksefficiëntie, evenals de mogelijkheid om te onthullen wat eerder binnen nog niet gekoppelde informatie verborgen kan zijn gebleven – zullen de wetenschap doen versnellen, waardoor open data iets wordt dat we allemaal zouden moeten omarmen.



*Reactie van Gert Oostindie, directeur Koninklijk Instituut voor Taal-, Land- en Volkenkunde*

## Een alternatief is er anno 2010 niet meer

Overtuigend, Rafael Sidi's pleidooi voor democratisering raw data. Toch zijn er, zoals hij aangeeft, wel problemen. Deels zijn die van methodologische aard, maar dat is geen principiële kwestie. Een ernstiger belemmering ligt bij onderzoekers, die bang zijn dat hun collega's er met hun al in een vroeg stadium Open Access gepubliceerde data ruwe onderzoeksgegevens materiaal vandoor gaan. Dat probleem zie ik niet zo snel verdwijnen, ook niet als er iets slims wordt bedacht in de sfeer van puntentelling.

Mij is gevraagd deze kwestie van een andere kant te bekijken, die van de wetenschappelijke uitgeverij. Ter verduidelijking: mijn instituut beweegt zich qua discipline op het vlak van de humaniora, qua aandachtsgebied op de voormalige Nederlandse koloniën en hun geografische omgeving, alsmede op postkoloniale migranten in Nederland. De zegeningen van IT voor het beheer van onze immense collecties is evident. Onze onderzoekers profiteren enorm van de gedigitaliseerde onderzoeksomgeving en dragen daar het hunne toe bij.

Maar wij zijn ook een uitgever van wetenschappelijk werk: geen raw data, maar het eindresultaat van onderzoek van wetenschappers vanuit de hele wereld. Wetenschappelijke publicaties in de humaniora: vrijwel per definitie moet daar geld bij, zeker in een niche zoals de onze. Maar dan nog behoort er een redelijke balans te zijn tussen de kosten van het publiceren en de baten van de verkoop. Wij besloten onze boekenseries gewoon commercieel te blijven uitgeven. Jaren geleden gingen wij wel onze twee prestigieuze Engelstalige tijdschriften digitaliseren.

Die zijn nu (ruim 150 respectievelijk 90 jaargangen) online te lezen, maar nog altijd met een moving wall van twee jaar. Nu bereiden wij ook de laatste stap voor, naar volledig Open Access. Dat gaat ons geld kosten: derving van inkomsten uit abonnementen. Elders in de organisatie zal dat pijn doen. Lang heeft ons dat weerhouden van die stap, maar een alternatief is er anno 2010 niet meer. De volgende uitdaging: optimale zichtbaarheid in de digitale wereld.



*Reactie van Saskia C.J. de Vries, directeur Amsterdam University Press, coördinator Open Access Publishing in European Networks*

## Een pleidooi voor Open Data is net zo goed een pleidooi voor Open Access publicaties

Rafael Sidi stelt terecht in zijn column dat grotere transparantie van alle wetenschappelijke data de innovatie van onderzoek zal versnellen. Maar zijn argumenten voor Open Data zijn even geldig als pleidooi voor Open Access van onderzoeksresultaten, en het is heel logisch om niet alleen voor Open Data maar ook voor Open Access van wetenschappelijke publicaties te pleiten. Dit geldt voor publicaties in STM (science, technology and medical studies) maar misschien nog wel meer voor uitgaven in de geestes- en sociale wetenschappen. De weerstand tegen Open Access - zowel van data als van onderzoeksresultaten (d.w.z. publicaties) - komt zowel van de kant van onderzoekers als van uitgevers. Onderzoekers zijn bevreesd dat ze met het openstellen van hun onderzoeksgegevens en resultaten de concurrentie in de kaart spelen. Maar doorbouwen op data en analyses daarvan vormt nou net de kern van wetenschap. En voor de geestes- en sociale wetenschappen zijn de resultaten van onderzoek, dat wil zeggen de publicaties, net zo cruciaal als de datasets zelf. Het is dus beslist niet productief om deze niet vrij en open toegankelijk te maken. Uitgevers maken zich met name zorgen om verlies van inkomsten. De problemen die ontstaan met het vrij toegankelijk maken van gegevens zijn veelal te wijten aan het feit dat deze een commerciële waarde voor de uitgevers vertegenwoordigen. Maar als uitgevers voet bij stuk houden en hun content om financiële redenen blijven afschermen, dan zullen andere bedrijven - als gevolg van de steeds geavanceerdere middelen voor technologische verspreiding - met de uitgevers gaan concurreren en uiteindelijk zou

de wetenschappelijke wereld de traditionele uitgevers wel eens niet meer nodig kunnen hebben. Als uitgevers dus sleutelfiguren in de academische en wetenschappelijke wereld willen blijven, en onder andere de controle op kwaliteit (peer-review) blijven organiseren, wat de kern is van wat ze doen, dan moeten ze moedig en vastberaden de weg van Open Access inslaan, van data en publicaties.



*Reactie door Theo Mulder, directeur Onderzoek en Instituten KNAW*

## Openheid is niet hetzelfde als bruikbaarheid

Ik ben verbaasd over de statements van Sidi. Zijn pleidooi voor toegang tot de data is zoveel anders dan de opvatting van de uitgevers over de toegang tot het eindresultaat van die data, de publicaties. Hierover bestaat veel minder eensgezindheid, hetgeen zich uit in een onderhandelingsproces dat met de snelheid van een vermoeide slak in de richting van een overeenstemming schuifelt (zie ook de column van Sybolt Noorda). Toch kunnen we een pleidooi voor openheid van data niet isoleren van openheid van publicaties.

Openheid van data zal er op termijn zeker komen. Nu al is het bij sommige tijdschriften in de levenswetenschappen verplicht om de data openbaar te maken. In andere domeinen staat men echter nog niet te springen om volledige openheid en wordt het bezit van de data gezien als een onderdeel van de concurrentiestrijd. Overigens, daar waar sprake is van onderling vertrouwen op basis van samenwerking, worden data gedeeld. Het stimuleren van samenwerking is om die reden dus al belangrijk.

Belangrijk is verder dat openheid nog niet hetzelfde is als bruikbaarheid. Dit is bij een instituut als DANS (een KNAW-NWO instituut), waar ze bij voortduring aan dataverrijking doen, maar al te goed bekend.

Lees ik nu door de regels heen de gedachte dat uitgevers ook datadepots kunnen opzetten (zoals bv Nature doet)? Zou dat wat zijn voor Elsevier? Dit brengt gemengde gedachten teweeg. Data behoren openbaar te zijn, maar wil je ze bij een commerciële uitgever opslaan of toch liever bij DANS?



*Reactie van Tine de Moor, assistent-professor Universiteit Utrecht en co-founder van het Open Access e-journal International Journal of the Commons*

## Wat we zelf doen, doen we beter?

Als wetenschapper interesseert mij eigenlijk maar één ding: een snelle verspreiding van hoogstaande wetenschappelijke onderzoeksresultaten. Zowel van onderzoeksdata of 'raw data', als van de verwerkte, geïnterpreteerde versie of in de vorm van een publicatie; zowel van collega's als van mezelf. Een zo kort mogelijke weg van kennisproducent naar kennisconsument. Snelheid van communicatie en topkwaliteit, daar wordt wetenschap beter van. Wetenschap bedrijft je niet voor het geld, zeker niet als je economisch historica bent. Wanneer commerciële bedrijven pleidooien gaan voeren voor Open Access, gaan bij mij dan ook meteen de voelsprietten rechtstaan.

De laatste jaren zijn de door hen aangeboden abonnementen op wetenschappelijke tijdschriftenbundels zo goed als onbetaalbaar geworden voor universitaire bibliotheken. Dus, denkt men dan: laten we de financiële last dan maar elders leggen. Open Access wordt door velen verstaan als free access en wanneer het op wetenschap aankomt zou het dat ook moeten zijn. De meeste wetenschappelijke resultaten komen tot stand met overheids gelden, en de resultaten dienen ook ten goede te komen aan de maatschappij. Maar Open Access is binnen de uitgeverijwereld vaak 'open-yes-but access': gratis voor de consument van de kennis maar niet voor de producent. Page charges van drieduizend euro per artikel en meer zijn normaal binnen de Open Access journals. Dit corrupteert de publicatie van wetenschappelijke gegevens: geen geld, geen publicatie. En dat is geen goede zaak voor de wetenschap.

Naast het feit dat we als belastingbetalers zelf recht zouden moeten hebben op gratis toegang tot wetenschappelijke resultaten, is er ook het sim-

pele gegeven dat zowel snelheid als kwaliteit meer baat hebben bij een voor alle partijen zo goed als gratis model dan bij een betalend model. Laat ons wel wezen: het zijn de wetenschappers zelf die ervoor zorgen dat wetenschappelijke journals naar een hoogstaand niveau getild worden, niet de uitgeverijen. In mijn vakgebied word je niet betaald voor peer reviews van artikelen, en dat moeten we ook zo houden. Uitgeverijen dragen vooral bij tot betere looks van een wetenschappelijk tijdschrift, maar ook dat is tegenwoordig geen onoverkomelijk probleem meer; zoiets kan je makkelijk outsourcen.

We kunnen ons de vraag stellen of intermediairen als commerciële uitgeverijen op het vlak van wetenschappelijke tijdschriften nog wel zo nodig zijn. De huidige fase is er een van transitie, die de wetenschap uiteindelijk ten goede kan komen. Uitgeverijen zijn een beetje de typisten van deze generatie: inmiddels zijn er al tal van gratis tools op de markt die wetenschapper in staat stellen om nagenoeg alles zelf te doen, voor veel minder geld. Een echt Open Access tijdschrift – waarbij noch de producent noch de consument betaalt - oprichten en 'runnen' kost in dit elektronisch tijdperk nog weinig geld. Met tienduizend euro per jaar kom je al een heel eind – meerdere issues per jaar -, terwijl ik voor dat geld amper drie artikelen gepubliceerd krijg in een page charge-journal. Met een eerder bescheiden bijdrage van elke auteur die enkel en alleen de werkelijke kosten dekt en dus vele malen kleiner is dan wat commerciële uitgeverijen vragen, kunnen kosten voor lay-out en promotie – voor zover echte kwaliteit zichzelf



niet promoot- gedekt worden. De open source software voor journal management doet tegenwoordig niet onder qua gebruiksvriendelijkheid voor de meeste commerciële pakketten. Het is tijd voor een nieuwe wind in het wereldje van academisch publiceren. Een die waait in de richting van Total Open Access. Kortom: wat we zelf doen, doen we beter, sneller en vooral goedkoper.

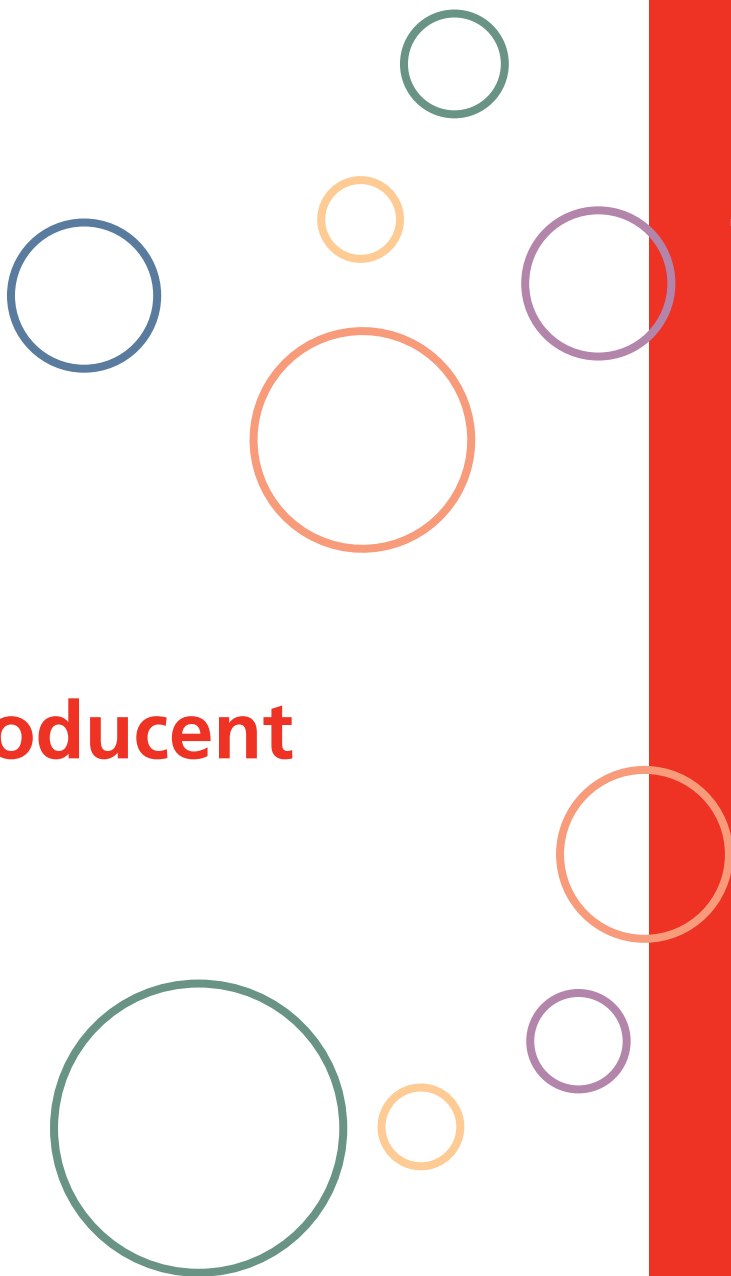
*Online comment: D.W. van Bekkum*

Wat is de citatie index van uw tijdschrift? Het overgrote deel van de onderzoekers in mijn gebied (biomedisch onderzoek) vermijdt publiceren in Open Access tijdschriften omdat hun universitaire carrière afhangt van het aantal publicaties vermenigvuldigd met de C.I. Is de universiteit van Utrecht een uitzondering?

*Online comment: Tine de Moor*

U hebt gelijk, de citatie index is voor onderzoekers vaak al een reden om bepaalde tijdschriften te ontwijken. 'The International Journal of the Commons' heeft nog geen citatie index, gewoon omdat het een aantal jaren duurt om die te krijgen, en het tijdschrift pas in 2007 gestart is. Bovendien is het in de meeste gevallen tegenwoordig niet de citatie index waar naar wordt gekeken maar het aantal artikelen dat is gepubliceerd in peer-reviewed tijdschriften (zowel bij carrièrestappen als voor aanvragen van grote projecten). En dit is zwaar peer-reviewed. Het hangt echter waarschijnlijk ook af van het onderzoeksgebied: in mijn onderzoeksgebied - economische geschiedenis - worden citatie indexes niet gebruikt. Elders wellicht wel, maar dan nog. Maar om een andere manier van meten aan te reiken: na slechts twee jaar hebben we een impact factor van 0,79. Een impact factor van boven de 1,0 is voor een tijdschrift zeer respectabel. Meer artikelen zijn meer dan 2000 maal gedownload. En dat is wat ik als universitair onderzoeker wil. Gelezen worden.

# De dataproducent



*Column van Marcel Das, CentERdata, Universiteit van Tilburg*

## Meer aandacht voor snelle en professionele datadiseminatie

Het belang van kwalitatief hoogstaande data-infrastructuren wordt vaak genoeg onderkend. Dataverzamelaars genieten minder aandacht en respect dan de onderzoekers die met de onderzoeksvragen op de proppen komen, de analyse uitvoeren en daarover rapporteren. Nog minder dan de dataverzameling zelf, wordt het belang ingezien van snelle en professionele datadiseminatie. Met name data die zijn verzameld met publieke gelden dienen meteen voor een zo breed mogelijk publiek toegankelijk te worden gemaakt. Uiteraard gelden hier voorschriften en regels die de privacy van de respondent moeten respecteren en die het vertrouwelijk omgaan met de data garanderen. Echter, het is moreel onverantwoord als de verzamelde informatie gefinancierd met publieke gelden slechts toegankelijk blijft voor een enkel individu. Een voorwaarde voor het verkrijgen van subsidie voor het verzamelen van gegevens moet dan ook zijn dat deze gegevens binnen enkele maanden voor andere onderzoekers en voor beleidsmakers beschikbaar komen.

Datzelfde geldt voor bestaande registerbestanden bij de overheid en publieke instellingen. Ook deze bestanden moeten snel en toegankelijk beschikbaar worden gesteld voor wetenschappelijk- en beleidsonderzoek. Bovendien, het alsmaar verzamelen van gegevens die elders reeds beschikbaar zijn is inefficiënt en vergt een onnodige investering van publieke gelden. Uiteraard geldt ook hier weer de voorwaarde van het respecteren van privacy en het garanderen van het vertrouwelijk omgaan met de data, maar daar kunnen prima afspraken over worden gemaakt.

De eis dat gegevens snel beschikbaar moeten

komen, betekent overigens niet dat dit zonder veel kosten en moeite mogelijk is. Het dissemineren van data gaat verder dan het plaatsen van een link naar een dataset op het internet. Gebruikers van de data moeten inzicht verkrijgen in de manier waarop de gegevensverzameling tot stand is gekomen, wat er precies in de dataset aan variabelen te vinden is, en welke vragen ten grondslag liggen aan deze variabelen. Subsidieverstrekkers die wat mij betreft als voorwaarde moeten stellen dat de gegevens snel voor een breed publiek beschikbaar moeten komen, moeten dan ook niet aarzelen om voor het dissemineren van gegevens een serieus budget beschikbaar te stellen.

Het door NWO gefinancierde MESS project met als kernonderdeel een representatief panel van Nederlandse huishoudens (het LISS panel) stelt alle verzamelde data zo snel mogelijk beschikbaar aan iedere onderzoeker die als doel heeft data te gebruiken voor wetenschappelijk, beleids- en maatschappelijk relevant onderzoek ([www.lissdata.nl](http://www.lissdata.nl)). Hopelijk volgen meerdere onderzoeks- en overheidsinstellingen dit initiatief. Ik juich het van harte toe!



*Reactie door Kees de Zeeuw, manager product- en procesinnovatie GEO, Kadaster*

## **Alleen goed onderhoud kan dataset bruikbaar houden**

De oproep van Marcel Das voor vrije beschikbaarheid van data is natuurlijk plausibel. Toch vind ik de suggestie dat het een kwestie is van aandacht, mentaliteit en budget als onderdeel van verstrekte subsidie niet voldoende. Ik heb twee opmerkingen hierbij.

Datasets (en zeker de registerbestanden van de overheid) zijn alleen maar interessant als deze niet alleen beschikbaar worden gesteld, maar vervolgens ook goed onderhouden worden. Actualiteit, betrouwbaarheid en mogelijkheid tot terugmelding op data zijn essentiële onderdelen van datasets: een dataset moet 'levend' zijn. Leven kost geld. Dat betekent dat er langetermijn budget gereserveerd moet worden, of inkomsten gegenereerd moeten worden met de data zelf.

In de praktijk blijkt dat afspraken hierover het moeilijkste onderdeel zijn van beschikbaarstelling van data. Goede metadata (beschrijvingen van de dataset) zijn inderdaad een vast onderdeel van een bruikbare dataset (zie bijvoorbeeld het Nationaal Geo Register voor beschikbare geo-informatie in Nederland, ([www.nationaalgeoregister.nl](http://www.nationaalgeoregister.nl))). Toch vraagt ook dit om een vorm van beheer en onderhoud. Het internet stroomt anders vol met een overmaat aan vergelijkbare data, waarvan niemand meer aan kan geven welke dataset er nu waarvoor gebruikt dient te worden. Ook binnen het wetenschappelijk onderzoek lijkt mij dit geen wenselijke situatie.





*Reactie van Jan Donner, directeur Koninklijk Instituut voor de Tropen*

## Therapieën

Wij leven in een global village. Een pleidooi voor snelle en professionele datadiseminatie spreekt aan, maar is wel erg vanuit de Nederlandse situatie geredeneerd. De werkelijkheid is echter ook, dat men elders in de wereld al jaloers is op wat wij hebben.

Voor het Koninklijk Instituut voor de Tropen (KIT) is werken op wereldschaal dagelijks werk. Activiteiten op tal van terreinen worden vanuit Amsterdam ondernomen of ondersteund. Snelle toegang tot de juiste data is soms zelfs een zaak van leven en dood. In sommige landen in ontwikkeling is de capaciteit van de nationale provider vergelijkbaar met die van de servers op het KIT. Weg is de snelheid!

In die situatie is het lastig om in veel bestanden te moeten zoeken. Om traag van het ene naar het andere bestand te moeten zappen is een bezoeking. Dan is een repository, waarin verschillende bestanden bijeen zijn gebracht, nuttig en het betekent tijdwinst.

Een dergelijke repository is [www.search4dev.nl](http://www.search4dev.nl), een online bibliotheek voor digitale publicaties van Nederlandse ontwikkelingsorganisaties. Ook uit beheersoogpunt betekent een dergelijke repository een besparing: de gezamenlijke, uitbreidbare, digitale en toegankelijke opslag van veel kleine bibliotheken in een grote, die van overal ter wereld benaderbaar is.

Databestanden moeten er wel zijn. Iemand moet ze aanleggen. Epidemiologische gegevens over heel wat landen en aandoeningen ontbreken of worden niet systematisch bijgehouden. We kennen zelfs zogenaamde 'verwaarloosde ziekten' – leptospirose, chagas of leishmaniasis om er een paar te noemen. Deze ziekten worden niet

verwaarloosd omdat ze niet belangrijk zijn – het tegendeel is het geval. Maar we weten niet hoeveel miljoenen(!) mensen er aan lijden. Op basis van goede epidemiologische data en hun verspreiding zou het niet moeilijk zijn om te besluiten, dat we onmiddellijk moeten investeren in goede therapieën!



*Reactie van Frans Hoeve, Centrum voor Beleidsstatistiek, CBS*

## Wie data van anderen gebruikt, brengt zelf ook data in

Marcel Das vraagt aandacht voor professionele en snelle datadissemiatioe. Niemand kan daar tegen zijn, toch? Dat zou je tenminste mogen verwachten. Toch is er in brede kring nog een opvatting die er op neerkomt dat het gebruiken van data van anderen prachtig, maar het zelf ter beschikking stellen een heel andere zaak is. Er ligt nog een mooie taak om onderzoekers er van te overtuigen dat het delen van data uiteindelijk voor iedereen profijtelijk is.

Bij het CBS is het al een aantal jaren mogelijk om gebruik te maken van databestanden op individueel niveau (microbestanden). Sinds enkele jaren kan dat ook via een remote access verbinding, waardoor onderzoekers vanaf de eigen werkplek analyses kunnen uitvoeren.

Hier is veel belangstelling voor, ook al omdat het mogelijk is zelf databestanden in te brengen en die op individueel niveau te koppelen met CBS bestanden. Maar het wordt nog veel aantrekkelijker als je bedenkt dat dit ook met databestanden van anderen mogelijk is. Om uitwisseling te bevorderen zou je er aan kunnen denken een stelregel in te voeren dat bestanden van anderen te gebruiken zijn als je zelf iets inbrengt.

Als je daarnaast nog bedenkt dat uitvoering van het DISS-project (Data Infrastructure for the Social Sciences) betekent dat deze voorziening voor een belangrijke groep onderzoekers vrijwel gratis te benutten zou zijn, dan lonken er wel hele mooie perspectieven.





**OPEN  ACCESS**

