



Wat Wikileaks ons leert over bibliotheken

Wouter Schallier, LIBER - Association of European Research Libraries

DATA EN TRANSPARANTIE

Sinds de start in 2006 heeft Wikileaks duizenden en miljoenen e-mails, foto's, video's en andere data gedeeld met de wereld. Open data worden steeds belangrijker, in die mate zelfs dat we verschuiven van een informatie- naar een datamaatschappij. Het performante en vrij toegankelijke internet laat iedereen immers toe grote hoeveelheden data snel en gemakkelijk te delen. De belangrijkste motivatie is om transparantie te verschaffen in zaken van algemeen belang. We lezen op <http://wikileaks.org/About.html>: *"Publishing improves transparency, and this transparency creates a better society for all people. Better scrutiny leads to reduced corruption and stronger democracies in all society's institutions, including government, corporations and other organisations. A healthy, vibrant and inquisitive journalistic media plays a vital role in achieving these goals. [...] Scrutiny requires information. Historically, information has been costly in terms of human life, human rights and economics. As a result of technical advances particularly the internet and cryptography — the risks of conveying important information can be lowered."*

In deze paragraaf vinden we een aantal interessante elementen. Ten eerste, de referentie aan transparantie die leidt naar meer en betere democratie. Iets gelijkaardigs treffen we aan bij President Obama's verklaring in verband met zijn Open Government Initiative (<http://www.whitehouse.gov/Open>): *"My Administration is committed to creating an unprecedented level of Openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government."* Ook het Open Public Data initiative van de Europese Commissie (<http://publicdata.eu/>) vermeldt transparantie als een belangrijke drijfveer, maar beklemtoont toch ook het economisch voordeel van open data: *"In addition to increasing transparency and improving public service delivery, open data creates opportunities for businesses to build new kinds of commercial services around this new data. This is because the data is published in a way which makes it legally and technically easy for anyone to reuse for any purpose. A recent study estimates the market based on European public sector information could be worth as much as €27 billion."*

Wetenschappers ontsnappen uiteraard niet aan die roep om transparantie en openheid. Steeds meer instanties die wetenschappelijk onderzoek financieren (Europese Commissie, National Institutes of Health, Wellcome Trust, Fonds Wetenschappelijk Onderzoek Vlaanderen, om er maar enkele te noemen) vragen aan onderzoekers om hun publicaties en soms zelfs ruwe data openlijk beschikbaar te maken. Na enkele bekende gevallen van misleiding of zelfs fraude, verwacht men van wetenschappers dat zij hun onderzoek op een nog transparanter manier gaan voeren zodat te allen tijde controle mogelijk is. Ook wil men natuurlijk transparantie verschaffen rond de

besteding van deze fondsen, die doorgaan uit publieke middelen komen.

Laten we een voorbeeld nemen. Een internationaal team onderzoekers voert metingen uit rond klimaatverandering. Het resultaat is massa's ruwe data van metingen. Vervolgens gaat het team die ruwe data verwerken, organiseren, analyseren en interpreteren. Uiteindelijk leidt dat dan tot een publicatie in de vorm van een artikel in een wetenschappelijk tijdschrift. Die publicatie komt ook in een vrij toegankelijk institutioneel archief. Tot voor kort volstond dat. Tegenwoordig vraagt men steeds vaker om ook de ruwe onderzoeksdata beschikbaar te stellen.

DE KRACHT VAN SAMENWERKING

Een tweede interessant element op de pagina van Wikileaks is de vermelding van de sterk verminderde kost van informatiedeling. Dank zij het performante en toegankelijke internet, kunnen de meeste burgers heel eenvoudig en aan een erg lage kost informatie publiceren (Facebook, Twitter enzovoort) en grote bestanden delen (Dropbox, Flickr enzovoort). Dit geeft de internetgebruiker veel macht. De democratisering van het publiceren (of tenminste van informatiedeling) illustreert dat we evolveren van een informatiemaatschappij (met grote hoeveelheden informatie gepubliceerd door een relatief beperkte groep van min of meer professionele providers) naar een datamaatschappij (waar nog grotere hoeveelheden informatie gepubliceerd worden door een in principe onbeperkte groep van niet-professionele providers). De rol van traditionele informatiebemiddelaars (bibliotheken, uitgevers, journalisten) wordt steeds vaker in vraag gesteld en zal mettertijd overbodig worden tenzij zij nieuwe toegevoegde waarden weten te genereren.

“DE ROL VAN TRADITIONELE INFORMATIEBEMIDDELAARS (BIBLIOTHEKEN, UITGEVERS, JOURNALISTEN) WORDT STEEDS VAKER IN VRAAG GESTELD EN ZAL METTERTIJD OVERBODIG WORDEN TENZIJ ZIJ NIEUWE TOEGEVOEGDE WAARDEN WETEN TE GENEREREN.”

Maar wat verstaan we eigenlijk onder data? Het rapport van de National Science Foundation (2005) onderscheidt de volgende categorieën van data: observationeel, computationeel, experimenteel en records. Een van de grootste uitdagingen van data is hun hoeveelheid: er zijn zo veel data dat het een goed georganiseerd netwerk van mensen en computers vereist om ze te analyseren binnen een redelijke tijdspanne. In wetenschappelijk

onderzoek worden de productie en analyse van enorme data-sets meestal verdeeld onder verschillende teams en computers op verschillende plaatsen ter wereld. Wikileaks werkt in feite op dezelfde manier met een enorm netwerk aan bronnen en analisten. Dit toont aan dat data niet alleen gedeeld worden omwille van transparantie maar ook omwille van efficiëntie. Data openstellen laat anderen toe data (te gebruiken en te hergebruiken) die duur of onmogelijk te reproduceren zijn. Zoals gesteld door de Europese Commissie, hebben open data bovendien het potentieel om een boost te geven aan een nieuwe economie, namelijk innovatieve diensten die gebruik maken van die vrij toegankelijke data.

DATA EN BIBLIOTHEKEN

Wat betekent dit allemaal voor bibliotheken? Ten eerste, data moeten worden verzameld, beschikbaar gesteld en bewaard door bibliotheken, omdat *“data are no longer considered as interim products to be discarded once the research reporting them is published. Rather, they have become important sources of scholarly content to be used and re-used”* (Borgman, 2008). Als data(sets) beschreven worden door adequate metadata en/of automatisch door computers kunnen verwerkt en gepubliceerd worden (cf. Van de Sompel, 2011), hebben ze zelfs het potentieel om traditionele publicaties zoals artikels en boeken volledig te vervangen. Accurate gestructureerde metadata zijn een absolute noodzaak om data (her)bruikbaar en nuttig te maken. Zonder die metadata zal een of meer van de volgende aspecten van data management worden verwaarloosd: beschrijving en identificatie, organisatie, bescherming, privacy, ethische kwesties, visualisatie, interpretatie, preservatie, persistente link tussen publicaties en data, geïntegreerd zoeken, validatie en peer review, bewaking van kwaliteit en integriteit, interoperabiliteit, controle over correct gebruik, selectie, publicatie, citatie. Onderzoek toont aan dat een van de belangrijkste redenen waarom onderzoekers geen data delen, te maken heeft met (de vrees voor) een gebrek aan controle over het juiste gebruik van hun data.

Bibliotheken moeten een actieve rol spelen in de bewustmaking rond datadeling. Behalve de vrees voor mogelijk misbruik, kunnen wetenschappers andere redenen hebben om hun data niet te delen, zoals: ze beschouwen de data als hun eigendom en willen ze ‘veilig’ lokaal bewaren op hun computer; ze hebben de indruk dat datadeling hun carrière niet verder helpt.

Bibliotheken van hun kant kunnen ook vele redenen vinden om niets te doen met data: er zijn te veel data, data zijn te complex en te duur, er zijn geen mensen noch middelen, en er zijn vele andere uitdagingen. Niettemin zijn er ook vele goede argumenten waarom bibliotheken zich wel moeten engageren voor data. Precies die positieve argumenten belichten we in ODE (Opportunities for Data Exchange, <http://www.ode-project.eu>), een project van twee jaar gefinancierd door de Europese Commissie. Datadeling vermindert het risico op

VIJF BELANGRIJKE GEVOLGEN VAN DATA VOOR BIBLIOTHEKEN EN DATA CENTRA:

Welke gevolgen heeft die evolutie naar een data-maatschappij voor bibliotheken? Is het een taak voor bibliotheken om wetenschappelijke data te verzamelen, te ontsluiten en te bewaren, net zoals ze al doen voor boeken en tijdschriften? Welke instrumenten en technieken kunnen bibliotheken aanwenden om ruwe data (her)bruikbaar en nuttig te maken?

1. Onderzoeksdata die ingebed zijn in traditionele wetenschappelijke artikels: zogenaamde verrijkte publicaties vereisen dat bibliotheken en data centra meer gesofisticeerde presentatie- en preservatiemethoden ontwikkelen
2. Databestanden die als supplement bij artikels horen: bibliotheken en datacentra moeten zorgen voor een persistente link tussen artikels en databestanden; alsook zorgen voor adequate presentatie- en preservatiemethoden
3. Datasets, bewaard in datacentra en repositories, waaraan (slechts) gerefereerd wordt in artikels: er moet gestreefd worden naar een zo duidelijk mogelijke verdeling van de verantwoordelijkheid tussen de houders van de dataset en van het artikel; ook hier moet gezorgd worden voor een persistente link tussen artikels en datasets, alsook adequate presentatie- en preservatiemethoden
4. Data die op zichzelf (dus zonder enige link met een artikel) gepubliceerd zijn in zogenaamde datatijdschriften of databanken (“data publication”): bibliotheken en datacentra kunnen hier onderzoekers ondersteunen tijdens het publicatieproces; ze kunnen metadata voorzien om (her-)gebruik en preservatie mogelijk te maken
5. Data her en der bewaard op disks in de onderzoekinstelling, die niet centraal beheerd/bewaard worden: bibliotheken en datacentra kunnen ondersteuning verlenen aan onderzoekers van in het begin van het wetenschappelijk onderzoeksproces, o.m. bij het opstellen van een datamanagementplan; verder moeten ze diensten ontwikkelen die gebruiksvriendelijke datapublicatie en/of -deponering mogelijk maken

> Bron: Reilly, S., Schallier, W., Schrimpf S., Smit, E., Wilkinson, M. (2011)



zinloze duplicatie, laat heranalyse toe, biedt een belangrijke toegevoegde waarde voor traditionele publicaties, kortom: het delen van onderzoeksdata stimuleert de wetenschappelijke vooruitgang en is een vereiste voor steeds meer financierders van wetenschappelijk onderzoek (en helpt dus wel de carrière van wetenschappers vooruit). In het ODE-project bekijkt enkele vroege successen (maar ook mislukkingen als we daaruit kunnen leren) vanuit drie perspectieven: wetenschappers, uitgevers, en bibliotheken en datacentra. En we formuleren mogelijke rollen voor elk van deze betrokkenen in ons *Report on Integration of Data and Publications* (<http://www.libereurope.eu/sites/default/files/ODE-ReportOnIntegrationOfDataAndPublication.pdf>).

DE NIEUWE BIJZONDERE COLLECTIES

Data moeten in een stabiele en betrouwbare omgeving worden bewaard en bibliotheken en data centra hebben hier een belangrijke rol te spelen. Cf. CLIR (2008). Onderzoeksdata hebben zonder meer het potentieel om de nieuwe bijzondere collecties te worden van bibliotheken, aangezien zij materiaal bevatten dat uniek is of dat slechts moeilijk en tegen een hoge kost kan ge(re)produceerd worden. Het valt te verwachten dat het succes van onderzoeksinstellingen in de toekomst heel sterk

zal afhangen van hun capaciteit om onderzoeksdata te genereren en te beheren, en om eigen data te presenteren, interpreteren en linken met andere data op een innovatieve manier. Aan de andere kant zal het succes van bibliotheken en datacentra afhangen van hun vermogen om ten volle deel te worden van de onderzoeks- en onderwijsworkflows, vanaf het begin van de productie van onderzoeksdata (en dus niet slechts wanneer data gepubliceerd worden). Bibliotheken en data centra kunnen zo direct bijdragen tot een efficiënter en transparanter wetenschappelijk onderzoek. Of zoals Luce (2008) het samenvat: *“New organizational models should reflect the environments they [i.e. wetenschappelijke bibliotheken] are attempting to support, recognizing the synergy and interdependence between scholars and information pioneers. To proactively support this environment, librarians must become part of the research process—full members of the research team. To do this, library staff members need to “go native” and embed themselves among the teams they support. Clearly this will have significant implications for the library’s staffing profile and workforce skill set.”* ■■

> De presentatie van Wouter Schallier op *Informatie aan Zee* is te herbekijken op <http://www.vvbad.be/bericht/wat-wikileaks-ons-leert-over-bibliotheken-wouter-schallier>