

Statistiek en sushi

De Rotterdamse aio Maria Fleischmann bezocht de Summer School in Ann Arbor.

Deze zomer ben ik twee keer vier weken in Ann Arbor, Michigan in de VS geweest. Ann Arbor is een klein stadje dichtbij Detroit en zou niet echt bekend zijn als er niet twee dingen waren: de University of Michigan met al zijn sportverenigingen en het ICPSR summer program voor kwantitatieve methoden.

Voor mij, aio bij de afdeling Sociologie aan de Erasmus Universiteit Rotterdam, waren de cursussen in statistische methoden de reden voor dit verblijf. In mijn dissertatie houd ik me bezig met arbeidsmarkttransities. Het paper dat ik recentelijk heb geschreven gaat over de beslissing om met pensioen te gaan in 21 Europese landen. Ik kijk daarbij naar individuele kenmerken, maar ook naar landenkenmerken.

ICPSR staat voor Inter-university Consortium for Political and Social Research, maar de letters zouden voor mij ook kunnen staan voor:

International

Sinds de jaren zestig wordt dit *summer program* aangeboden en het staat hoog aangeschreven. Mensen uit alle continenten komen naar Ann Arbor om een methodische opleiding te volgen en als het over statistiek gaat begrijp je elkaar snel. Ver-



schillen in alledaagse dingen leiden dan juist tot spannende discussies, en je leert meteen nog meer.

Classes

De programmaleiders ontraden het om meer dan twee workshops tegelijk te volgen. Maar het is erg moeilijk om je tot twee cursussen te beperken als je eigenlijk iedere cursus wel spannend vindt. En op school of op de universiteit heb je nog nooit zulke goede docenten gehad! Ik koos voor vier workshops: Regression II en Maximum Likelihood Estimation (MLE) in de eerste vier weken, en Regression III en Advanced Methods of MLE in het tweede gedeelte. In Regression II ging het om lineaire modellen, in Regression III onder meer om Model Specification en Outlying Observations. Het is erg nuttig om tot een betere modelspecificatie te komen of om te leren wat je moet doen met onverwachte observaties. Voor mij zal dat een rol spelen in bijvoorbeeld landenkenmerken die extreme waarden hebben. MLE was erg nuttig om survival analyse te doen. Met betrekking tot de pensioensleeftijd zou je kunnen vragen welke individuele kenmerken tot een jongere pensioensleeftijd leiden.



Picknick

Na een week hard werken moet je even wat anders doen. Iedere twee weken organiseert het ICPSR een picknick. De bedoeling is om iedereen bij elkaar te krijgen en een middag te kletsen, te netwerken of samen te sporten. Volleybal, baseball, basketbal of voetbal, het is er alle-

maal en docenten en studenten genieten van een middagje niets statistiek-achtigs doen.



Sushi

Eten in de VS was een onderwerp waarover je het met iedereen kon hebben. Met de supermarkten ver buiten de stad en wij allemaal zonder auto. Gelukkig is het restaurant-aanbod in Ann Arbor erg groot en er waren veel alternatieven naast burgers en sandwiches: Sushi, Koreaans, Veggie. The sky is the limit en het is goed betaalbaar!



Reis

Tussen de cursussen en opdrachten door blijven dan maar enkele dagen over om nog wat van het land te zien. Een auto huren is goedkoop, maar ook de trein en de bus brengen je naar Chicago of Detroit. Wil je geen steden meer om je heen, dan kun je naar het Noorden van Michigan, naar de Sleeping Bear Dunes en Lake Michigan. Alleen, je merkt het al, aan je dissertatie kom je tijdens deze weken dus echt niet meer toe.

Het was een ontzettend leuke, leerzame en belevingsvolle zomer. Ik ben erg blij dat ik met de financiële steun van DANS en mijn begeleiders naar Ann Arbor kon gaan.

Maria Fleischmann

GELEZEN

David Giaretta [et al.]: *Advanced Digital Preservation*. Springer, 2011, ISBN 978 3 642 16808 6, €99,95

In dit boek legt Giaretta uit waarom gangbare technieken voor digitale duurzaamheid (migratie, emulatie) niet werken voor wetenschappelijke onderzoeksdata, die vaak bestaan uit complexe digitale objecten als databases en softwareprogramma's. In plaats daarvan ontwikkelt Giaretta met zijn team een techniek die draait om het netwerk van representation information - ieder digitaal object moet alle informatie in zich bergen die ooit nodig kan zijn om het object te hergebruiken. Naast deze les digitale duurzaamheid voor gevorderden bevat het boek veel basisinformatie over digitale duurzaamheid en kan het ook goed op een minder technisch niveau worden gelezen, bijvoorbeeld door managers van repositories en databibliothecarissen.

Marloes Scholtens: *Born-digital Bijzondere collecties*. Rapport van de werkgroep Born-digital van de UKB commissie Bijzondere Collecties. Leiden, 2011

Digitaal materiaal dat van oorsprong digitaal is, en dus geen analoge tegenhanger heeft, loopt nog steeds een groot risico om verloren te gaan. Vaak is niet duidelijk wie er verantwoordelijk is voor collectiëren en voor duurzaam bewaren. In Nederland heeft de Koninklijke Bibliotheek de verantwoordelijkheid genomen voor elektronische boeken en tijdschriften, en het Nationaal Archief ontfermt zich over de bedrijfskundige archieven van de universiteiten, maar veel ander materiaal valt nog tussen wal en schip. De persoonlijke archieven van hoogleraren bijvoorbeeld, of wetenschappelijke archieven van vakgroepen. Het zeer gedegen rapport inventariseert de soorten materiaal die bedreigd worden en de internationale praktijk op dit gebied, en geeft aanbevelingen. Het rapport is beschikbaar op de NCDD-website www.ncdd.nl/.

Madeleine de Cock Buning, Barbara van Dinther, Christina H. Jeppesen-de Boer, Allard Ringnalda: *The legal status of research data in the Knowledge Exchange partner countries*. Centre for Intellectual Property Law (CIER) and Molengraaff Institute for Private Law, Utrecht University, 2011

Knowledge Exchange is a cooperative effort between four European organisations that support the use and development of Information and Communications Technologies infrastructure for higher education and research. Based on the four national strategies the joint vision of the initiative is to make a layer of scientific and scholarly content openly available and re-usable on the internet. The central aim of this report is to identify whether Intellectual Property (IP) rights can form any obstacle to sharing publicly funded research data under open access terms. To the extent that such obstacles do indeed exist, the report makes some suggestions as to how these could be resolved. www.knowledge-exchange.info/

COLUMN

Drie netelige datakwesties

Open Acces van onderzoeksdata biedt nieuwe mogelijkheden voor wetenschappelijk onderzoek. Het is echter nog maar de vraag of in het huidige academisch klimaat deze vorm van e-science wel tot zijn recht kan komen.

Er is een standaardmodel van wetenschapsbeoefening aan de Nederlandse universiteiten. Wetenschappelijk onderzoek wordt verricht door excellente onderzoekers, die samenwerken in onderzoeksgroepen en onderzoeksprogramma's en beoordeeld worden op wetenschappelijke publicaties. Onderzoeksprojecten duren vier jaar, de periode waarbinnen een beginnend onderzoeker een proefschrift kan schrijven. Wetenschappelijke carrières worden gebouwd op individuele subsidies. Alle prestatiecijfers van de Nederlandse wetenschap laten zien dat dit model zeer succesvol is.

In haar boek *Zuiver om de wetenschap* verhaalt Patricia Faasse over het ontstaan van de levenswetenschappelijke instituten van de KNAW, waaronder het Herseninstituut, het Centraal Bureau voor Schimmelcultures (het ene CBS) en het Hubrechtlaboratorium. De oprichting van deze instituten aan het begin van de 20e eeuw laat een andere stijl zien van wetenschappelijk onderzoek. Met veel geduld bouwen de eerste directeurs van de instituten collecties op van onderzoeksmateriaal. Soms zonder dat duidelijk is wat de directe waarde van het verzamelde materiaal is voor het eigen onderzoek, en vaak om andere onderzoekers op zeker moment van dienst te kunnen zijn.

Het ontstaan van de instituten weerspiegelt een taxonomische stijl van onderzoek die in de loop van de 20e eeuw steeds meer is weggedrukt. Veelzeggend is bijvoorbeeld hoe in de jaren tachtig bij bezuinigingen de taxonomie vrijwel geheel verdween uit de biologische faculteiten. En misschien wel het meest taxonomische instituut dat we in Nederland op dit moment kennen, het Centraal Bureau voor de Statistiek (dat andere CBS), komt in overzichten van de Nederlandse wetenschappelijke faciliteiten nooit voor.

Maar de taxonomie is terug: in de vorm van e-science en Open Acces van databases.

Is de organisatie van het academisch onderzoek wel toegesneden op deze nieuwe stijl van onderzoek? Ik geef u drie netelige kwesties om de gedachten hierover te prikkelen.

- Promotie-eisen: Zijn we bereid onderzoekers te laten promoveren op een mooie database? Of leren we hen de data eerst alleen zelf te gebruiken voor publicaties en deze pas openbaar te maken als de eigen carrière op de rails staat?
- Databasebeheer: Zijn we binnen universiteiten in staat om de continuïteit van databases te organiseren: de opbouw, de opslag, het onderhoud en de vernieuwing van data? Of moeten we daarvoor nieuwe para-universitaire instituten oprichten?
- En als laatste kwestie: Zijn we in staat om het eigendom en de toegang te regelen? Van wie zijn de data: van de onderzoeker, de onderzoeksinstelling, de financier of misschien wel van de data zelf (lees: patiënten, geëquipteerden, organisaties)? En wie neemt dan de telefoon op als een onderzoeker de data graag wil gebruiken voor nieuw onderzoek?

Ik vrees dat bij veel organisaties deze vragen nog nauwelijks beantwoord zijn. Ze worden overgelaten aan de onderzoekers. In het wetenschapsbeleid heet dat dan 'bottom up' en dan zijn we er voor. In de praktijk betekent het dat onderzoekers moeten aanmodderen, en dan is het jammer dat we Open Access van onderzoeksdata nog niet goed geregeld hebben.

Barend van der Meulen
hoofd afdeling Science System
Assessment van het Rathenau
Instituut.

