

Peter Doorn, drijvende kracht in hergebruik van onderzoekgegevens:

# ‘Ik vind alles interessant, als er maar data ter beschikking komen’

HANS VAN MAANEN

Het Steinmetzarchief, het Nederlands Historisch Data Archief, het Wetenschappelijk Statistisch Agentschap, het Electronisch Depot voor de Nederlandse Archeologie: allemaal zijn ze inmiddels verenigd in DANS, Data Archiving and Networked Services. En één naam duikt in die archieven met een opmerkelijke regelmaat op: die van Peter Doorn. Wie is deze drijvende kracht in, zoals hij het zelf wat ironisch noemt, ‘de handel in tweedehands data’? Een interview.

‘Twee dingen vielen mij op toen ik met mijn proefschrift bezig was, en allerlei gegevens uit de Arbeidskrachtentelling wilde koppelen aan het Woningbehoefte-onderzoek.

Het eerste was dat het gebruik van die bestanden zo ontzettend duur was. Voor die Arbeidskrachtentelling moest een ton betaald worden – in gulden, we spreken over begin jaren tachtig, toen ook het profijtbeginsel volop speelde. Het tweede, veel belangrijker, was dat die bestanden eigenlijk amper te combineren waren. Er was geen koppeling of vergelijking van data mogelijk, omdat ze steeds hun eigen indelingen en definities hadden gekozen. Ik heb toen heel wat moeten puzzelen en programmeren om er toch uit te krijgen wat ik wilde – en ik had daar blijkbaar zo’n aardigheid in, dat het mijn vak is geworden. Ik had er in ieder geval, als sociaal-geograaf, nog voor mijn promotie meteen mijn baan bij geschiedenis in Leiden aan te danken: daar ging ik het vak ‘computertoepassingen in de geschiedenis’ doceren.’

*‘We zijn zo aan gewend dat onze privé-gegevens bekend zijn dat we onze privacy blijmoedig te grabbel gooien’*

‘De meeste historici vonden het natuurlijk maar niks, al die computers, maar er waren er ook die wel degelijk de mogelijkheden ervan inzagen. Zo konden we in 1989 beginnen met het aanleggen van het Nederlands Historisch Data Archief. Net als de Steinmetzstichting vijftienvier jaar daarvoor haalden we overal oude ponskaarten en magneetbanden met onderzoeksgegevens vandaan om die te ‘ontsluiten’.

Een van de eerste echt grote bestanden die dat Steinmetzarchief had gekregen was van de Volkstelling van 1960: dat was de eerste waarbij de computer werd ingeschakeld. Weliswaar nog met ponskaarten – het CBS moest daarvan af, en ze werden toen onder verre van optimale omstandigheden opgeslagen in een ruimte van de Universiteit van Amsterdam – dus dat heeft later nog heel wat hoofdbrekens opgeleverd, maar dat soort data was wel een prachtige basis om te beginnen.

Pas veel later hebben we digitaal reddingswerk verricht aan de bestanden van de Volkstelling van

1971. De telling van 1981 ontbreekt: er was toen zoveel verzet uit de bevolking tegen ‘Big Brother’ dat men bang was dat door het aantal weigeraars het hele onderzoek zinloos zou worden. Later is de Wet op de Volkstelling aangepast, en nu worden er geen grote officiële Volkstellingen meer gehouden, tenminste niet op de klassieke manier.’

‘Dat verzet tegen die Volkstellingen had te maken met de vrees voor privacy – maar het interessante is dat al die data waar toen zoveel zorg over bestond, in feite nu nog veel nauwkeuriger bekend zijn. Door de koppeling van bestanden. Het CBS weet letterlijk alles van je – waar je woont, wat je doet, wat je verdient, je afkomst. Een ouderwetse volkstelling is daardoor helemaal niet meer nodig.

Ik moet er wel bij zeggen dat het CBS buitengewoon zorgvuldig is met zijn bestanden – alleen met een chipkaart en een vingerafdruklezer kun je vanuit een beveiligde ruimte op een speciale pc op de faculteit bij de CBS-bestanden op individueel niveau. En je krijgt je analyseresultaten pas nadat het CBS heeft gecontroleerd dat er geen data in zitten die tot individuen herleidbaar zouden kunnen zijn.

Grappig is dat de opvatting over privacy onder de bevolking ondertussen lijkt te zijn verschoven. We zijn er blijkbaar zo aan gewend geraakt dat onze privé-gegevens bekend zijn bij de Airmiles, bij telemarketeers en bij Google, dat we onze privacy blijmoedig te grabbel gooien. Het is onvoorstelbaar wat mensen geheel vrijwillig aan persoonlijke informatie op weblogs, Facebook en andere sociale netwerken zetten. En de zorgverzekeraars kunnen op sommige community sites zo hun risicogroepen van lijders aan nare ziekten traceren, daar hebben ze het elektronisch patiëntendossier straks niet eens voor nodig.’

*‘Er wordt zo gehamerd op de verhoging van de AOW-leeftijd, terwijl dat maar een van de mogelijke oplossingen is’*

‘Het Historisch Data Archief werd in 1995 een instituut van de KNAW, en in 1997 samen met het Steinmetzarchief onderdeel van het NIWI. Ik



WIEBE KESTRA

**Peter Doorn: ‘Van alles houden mensen gegevens bij, en die wil DANS beschikbaar maken voor analyse’.**

wilde toen eigenlijk afscheid nemen van het data-archiveren en weer onderzoek gaan doen aan de universiteit, maar nadat bij het NIWI troebelen uitbraken en de hele leiding werd vervangen, besloot ik daar toch te solliciteren als afdelingshoofd. Het NIWI was een samenvoeging van zes instituten en dat was achteraf – en volgens sommigen ook al vooraf – niet zo’n goed idee, nee. Er waren toch te veel verschillende culturen, en er ontstond niet het krachtige instituut voor wetenschappelijke informatie dat bedoeld was. Er is door het interne gesteggel veel tijd verloren gegaan, juist in een periode dat het digitaliseren een hoge vlucht nam. Er is natuurlijk wel het een en ander tot stand gebracht, maar we hebben, vind ik, toen niet de grote sprong voorwaarts gemaakt die mogelijk was. In zekere zin zijn we toen overvleugeld door de Koninklijke Bibliotheek, die heeft toen het voortouw kunnen nemen.

Toen in 2005 DANS werd gevormd, was dat na het NIWI toch wel een bevrijding. Bij DANS gaat het nu simpelweg om toegang tot data voor de wetenschap, nu en op

de lange termijn. We hebben in een paar jaar tijd een duidelijke positie opgebouwd, ook ten opzichte van de KB. Bij de KB staat de authenticiteit van de documenten voorop, wij zijn vooral geïnteresseerd in de herbruikbaarheid van de gegevens. Onderzoekers kan het meestal niet schelen in welk lettertype een dataset is gezet, of in welk handschrift een scheepsjournaal is bijgehouden, het gaat ze bijna altijd primair om de data zelf, om de informatie-inhoud. Die willen ze met moderne software te lijf kunnen.’

‘Een mooi project was dat trouwens, met die oude scheepsjournaals. Dat hebben we samen met het KNMI gedaan: alle weergegevens die in oude logboeken, van ongeveer 1750 tot 1850, waren vermeld, hebben we nageplozen. Het weer op die schepen werd elke dag heel nauwkeurig bijgehouden, en die database kun je prachtig gebruiken om het klimaat te reconstrueren voor een tijd dat er nog geen weerstations waren. Meer dan 300.000 waarnemingen zijn ingevoerd, en omdat Engeland en Spanje ook meededen, leverde dat een goed beeld van het weer op alle

oceanen. Dat is toch waar het om gaat, vind ik, met oude data nieuwe dingen doen. Er zit in die scheepsjournaals nog zoveel meer: we weten met welke snelheid ze voeren, welke routes – je kunt ze zo op Google Maps uittekenen.

Van alles houden mensen gegevens bij, en die wil DANS beschikbaar maken voor wetenschappelijke analyse. Slavenhandelaren hielden registers bij, archeologen leggen hun opgravingen digitaal vast, we zijn nu ook bezig met de psychologen – ik vind alles interessant, als er maar data ter beschikking komen. Dat is ook wel een beetje kenmerkend voor mij: ik ben een omnivoor. In mijn geografieopleiding zaten stukken sociologie, antropologie en economie; en als historisch-geograaf heb ik jarenlang onderzoek gedaan met historici en archeologen in Griekenland.’

*‘Overall haalden we oude ponskaarten en magneetbanden met onderzoeksgegevens vandaan om ze te ontsluiten’*

‘En laat ik ten slotte nog zeggen, dat het af en toe ook buitengewoon actueel kan zijn, dat onderzoek naar historische data. Neem de discussie over de verhoging van de AOW-leeftijd die de laatste tijd speelt. Theo Engelen in Nijmegen heeft, op grond van historische reeksen uit onder andere de volkstellingen, laten zien dat er in feite nog nooit zoveel mensen aan het arbeidsproces hebben deelgenomen als tegenwoordig. Vroeger, toen de bevolkingspiramide nog een echte piramide was, en zelfs toen er nog kinderarbeid was, voor 1901, moest een veel kleiner deel van de bevolking dan nu het nationaal product bij elkaar verdienen.

Zo’n inzicht krijg je alleen als je naar lange reeksen kijkt, en niet alleen naar de laatste paar jaar. En dat heeft ook politieke consequenties. Er wordt maar gehamerd op de verhoging van de AOW-leeftijd van 65 naar 67, maar dat is maar een van de mogelijke oplossingen voor het probleem. Je kunt ook meer actieven krijgen door de arbeidsparticipatie van vrouwen te verhogen – en dat lijkt mij eerlijk gezegd een veel betere politieke koers.’

Peter Doorn studeerde sociale geografie in Utrecht en promoveerde daar in 1989. Tot eind jaren negentig 1998 doceerde hij computergebruik in de geschiedenis in Leiden. Daarna werd hij afdelingshoofd van het Nederlands Instituut voor Wetenschappelijke Informatie-diensten (NIWI) en directeur van het Nederlands Historisch Data Archief. Hij was betrokken bij veel projecten op het grensgebied van geschiedenis en informatica, zoals HGIN (naar een Historisch Geografisch Informatiesysteem voor Nederland), en Life Courses in Context. Doorn nam in 2005 de leiding op zich van het nieuwe instituut DANS.