

# Zuinig op je cijfers

## GEEN VASTE REGELS

Een communicatiewetenschapper gebruikt heel andere gegevens dan een plantenonderzoeker. De universiteit heeft daarom geen gedetailleerde regels voor de beheerplannen. Er is slechts een door de WUR-bibliotheek gemaakt voorbeeldplan dat promovendi kunnen aanpassen aan hun eigen situatie. Wat er in ieder geval wel in een goed plan staat, is wie verantwoordelijk is en wie eigenaar wordt van de gegevens. Bovendien moeten onderzoekers bedenken hoe ze zorgen dat zij, of hun opvolger, de data over enkele jaren nog steeds kunnen terugvinden en begrijpen. Bestanden moeten dus een duidelijke naam en datum hebben, ze moeten logisch zijn verdeeld over mappen en computerprogramma's moeten de gebruikte bestandsformaten over tien jaar nog steeds kunnen lezen. En natuurlijk staan al deze data tijdens en na de promotie op een harde schijf waar belangstellenden ze terug kunnen vinden.

Vanaf 1 april is elke promovendus verplicht om zorgvuldig met onderzoeksdata om te gaan. Dat moet een nieuwe 'affaire Stapel' helpen voorkomen, maar het is ook in het voordeel van de onderzoeker. 'Verloren data zijn een verspilling van tijd, moeite en belastinggeld.'

tekst: Rob Ramaker / illustratie: Kito

**E**en stapeltje dvd's van ruim tien jaar oud met daarop de hardeschijfinhoud van enkele oud-onderzoekers. Dit kreeg Annemarie Patist van de WUR-bibliotheek onlangs op haar bureau. Een leerstoelgroep wilde graag de onderzoeksdata van hun oud-medewerkers veilig stellen. Op de schijfjes bleek het echter een chaos: data van verscheidene onderzoekers stonden door elkaar, net als oude en nieuwe versies. Veelal ontbraken toelichting, bron en betrouwbare tijdsaanduiding. Een onoplosbare puzzel, concludeerde Patist. De dvd's, en daarmee jaren aan data, konden in de prullenbak.

Als het goed is, behoren zulke anekdotes vanaf 1 april tot het verleden. Elke nieuwe promovendus die in Wageningen begint, moet dan een databeheerplan hebben. Dit dwingt de onderzoekers om na te denken over het zorgvuldig verzamelen, beheren en opslaan van hun data, voordat ze überhaupt beginnen met het echte werk. Niet alleen individuele onderzoekers moeten eraan geloven. Ook alle leerstoelgroepen moeten begin april een databeheerplan presenteren. De nieuwe dataregels zullen daarvoor het werk van elke Wageningse onderzoeker beïnvloeden.

### MACHTELOOS

Nieuwe regels op het gebied van databeheer hingen al langer in de lucht, maar de recente fraudeschandalen rond de wetenschappers Diederik Stapel en Dirk Smeester hebben plannen in een stroomversnelling gebracht. Vooral bij Smeester speelden data een centrale rol. De Rotterdamse marketingprofessor ontkende altijd met klem dat hij net als Stapel gegevens had verzonden. Maar bewijzen kon hij dat niet. Zijn papieren formulieren zou hij zijn

kwijtgeraakt bij een verhuizing, en de digitale databestanden waren volgens hemzelf verloren gegaan na een crash van zijn harddisk, vlak voor het integriteitsonderzoek.

'Ongeloofwaardig,' concludeerde de onderzoekscommissie knarsetandend, maar stond verder machteloos. Dat leidde tot een brede discussie over de omgang met data. Strengere regels, zo redeneerde men, maken het minder verleidelijk om te frauderen. En wanneer het toch gebeurt, wordt dit eenvoudiger opgemerkt en bewezen. Een onderzoekscommissie van de Koninklijke Nederlandse Academie der Wetenschappen (KNAW) adviseerde universiteiten vervolgens om de gedragscode rond databeheer aan te passen. De nieuwe Wageningse regels zijn daarvan het directe gevolg.

### BUREAUCRATIE

Hugo Besemer, informatiespecialist bij de WUR-bibliotheek, heeft gemengde gevoelens bij het fraudeverhaal. Enerzijds is het goed dat de fraudegevallen het dataprobleem op de agenda hebben gezet. Anderzijds moeten de nieuwe regels volgens hem niet gezien worden als een teken van wantrouwen, of als wetenschappertje pesten. Een goed dataplan is volgens hem juist een kleine investering die zichzelf ruim terugbetaalt: 'Je maakt het echt voor jezelf.' Wetenschappers met een goed plan werken volgens hem meer gestructureerd, en zij kunnen hun gegevens – jaren na het aanmaken – sneller terugvinden en begrijpen. Verder vermoedt Besemer dat zorgvuldig omspringen met gegevens promovendi zal helpen om hun proefschrift sneller te schrijven. 'Hun informatie is soms zo rommelig dat ze hun onderzoek gewoon niet op papier krijgen.'

Om de onderzoekers te helpen bij het maken van hun



beheerplan, organiseert de bibliotheek sinds enkele jaren cursussen datamanagement. De deelnemers van de eerste edities blijken overwegend positief. 'De cursus dwingt je na te denken hoe je je data logisch en overzichtelijk kan opslaan,' zegt Lucie Vermeulen, promovendus bij Milieusysteemanalyse. Een andere promovendus die 'enkele tientallen terabytes' verwacht te gaan verzamelen, realiseerde zich pas tijdens de cursus hoe uitdagend het is om zoveel gegevens veilig te bewaren, vooral op de lange termijn. Aangezien hij al vroeg tegen dit probleem aanliep, heeft hij nu veel tijd om na te denken over oplossingen. Dit is precies het actieve denkwerk dat Besemer wil aanmoedigen: 'Mensen zeggen vaak tijdens de cursus: "Hier heb ik nog nooit over nagedacht".'

#### AUTEURS ONVINDBAAR

En dat nadenken blijkt hard nodig. Een recent onderzoek in *Current Biology* laat bijvoorbeeld zien dat wetenschappers geen goed idee hebben hoe ze hun ruwe gegevens veilig stellen voor de lange termijn. De paper bekeek voor 516 ecologieartikelen die tussen 1991 en 2011 werden gepubliceerd, of de data nu nog te vinden zijn. Voor de twee voorafgaande jaren lukte dat nog wel, maar met elk jaar dat ze verder teruggingen in de tijd daalde de kans op suc-

ces. Uiteindelijk bleek dat voor slechts 20 procent van de artikelen uit het begin van de jaren negentig nog data te vinden was. En over de hele linie lukte het slechts om met 40 procent van alle auteurs in contact te komen.

Het verdwijnen van deze data betekent een enorm verspilling van tijd, moeite en belastinggeld. Dataverlies is echter vooral zonde voor de wetenschap. Gegevens kunnen namelijk een productief tweede leven leiden. Zo is het voor de betrouwbaarheid belangrijk dat wetenschappers de analyses van hun collega's herhalen op zoek naar fouten. Verder verschijnen met het verstrijken van de tijd nieuwe analysetechnieken. Deze zorgen dat uit dezelfde gegevens, soms samengenomen met die van andere experimenten, scherpere of nieuwe conclusies zijn te destilleren. En tenslotte kan de wetenschap allerlei nieuwe vragen stellen aan oude data. Zo worden de weergegevens die meteorologen al honderden jaren verzamelen tegenwoordig gebruikt om klimaatverandering te bestuderen. En met stokoude data over jachtvangsten in Canada ontdekten ecologen hoe de populatiegrootte van prooi- en roofdieren elkaar beïnvloeden. Als data verstoofd in boodschappentassen of floppy's, wordt ze nooit een tweede leven gegund, maar onder het nieuwe regime zal dat niet meer gebeuren. ③

**De wetenschap kan allerlei nieuwe vragen stellen aan oude data**