

Bedankt voor het downloaden van dit artikel. De artikelen uit de (online)tijdschriften van Uitgeverij Boom zijn auteursrechtelijk beschermd. U kunt er natuurlijk uit citeren (voorzien van een bronvermelding) maar voor reproductie in welke vorm dan ook moet toestemming aan de uitgever worden gevraagd.

# Boom

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch door fotokopieën, opnamen of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van kopieën uit deze uitgave is toegestaan op grond van artikelen 16h t/m 16m Auteurswet 1912 jo. Besluit van 27 november 2002, Stb 575, dient men de daarvoor wettelijk verschuldigde vergoeding te voldoen aan de Stichting Reprorecht te Hoofddorp (postbus 3060, 2130 KB, [www.reprorecht.nl](http://www.reprorecht.nl)) of contact op te nemen met de uitgever voor het treffen van een rechtstreekse regeling in de zin van art. 16l, vijfde lid, Auteurswet 1912.

Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16, Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten, postbus 3060, 2130 KB Hoofddorp, [www.cedar.nl/pro](http://www.cedar.nl/pro)).

*No part of this book may be reproduced in any way whatsoever without the written permission of the publisher.*

[info@boomamsterdam.nl](mailto:info@boomamsterdam.nl)  
[www.boomuitgeversamsterdam.nl](http://www.boomuitgeversamsterdam.nl)

## COLUMN

| In deze rubriek laten redactieleden en lezers van KWALON hun licht schijnen over actuele ontwikkelingen binnen het kwalitatief onderzoek. Bijdragen kunnen worden aangeleverd bij Fred Wester: [f.wester@maw.ru.nl](mailto:f.wester@maw.ru.nl).

# Ieder vogeltje zingt ... over Big Data en QDA-software

Jeanine Evers\*

Stel, je zit voor je (promotie)onderzoek interviewmateriaal te coderen in een programma voor kwalitatieve analyse. Het valt je op dat Piet over de interventie die hij ondergaat, dingen zegt die een soort tijdsverloop van het resultaat indiceren, maar ook tegenstrijdigheid. Voorts meen je je te herinneren dat je iets dergelijks ook bij andere respondenten hebt gehoord. Wat zou het nu toch mooi zijn als je al deze uitspraken aan elkaar kunt koppelen, om zo het tijdsverloop en de contradicties inzichtelijk te maken! Helaas lukt dat niet in dit programma, daarvoor heb je een ander programma nodig, maar die analyseprogramma's praten niet met elkaar, je kunt de data niet overbrengen!?!)

Niet getreurd, want juist hiervoor hebben wij REFI opgericht, ofwel: *The Rotterdam Exchange Format Initiative*. Dit is een groep softwareontwikkelaars plus faciliterende coördinatiegroep die ontstond na de internationale KWALON-conferentie van 2016 (zie verslag van Moerman en Boeije in KWALON 63, 21(3), p. 65-68). Hoe het zover kwam? Ik riep destijds ontwikkelaars van Qualitative Data Analysis Software (QDAS) op om te gaan werken aan een zogenaamde 'interoperability standard', oftewel een uitwisselingsprotocol waarmee je gegevens uit het ene pakket kunt overhevelen naar het andere pakket. Dit deed ik onder andere vanuit irritatie over het regelmatig imiteren van features van de concurrent zonder dat dit leidde tot gelijkwaardige functionaliteit. En tot mijn grote verrassing zeiden zij ter plekke 'JA' op dat idee! Snel daarna was REFI gevormd en gingen we aan de slag. Met geld van verschillende instanties uit Nederland en Canada werd een forum ontwikkeld, werden diverse 'live' en online meetings georganiseerd en gingen de programmeurs aan het werk om uit te dokteren hoe hun verschillende softwarepakketten al dan niet op elkaar konden aansluiten. Een soms moeizaam, maar vooral heel interessant proces om te observeren; je bent en blijft tenslotte antropoloog ...

\* Jeanine Evers is directeur/eigenaar van Evers Research & training (voormalig KWALON), een bedrijf dat cursussen en coaching in kwalitatief onderzoek verzorgt en onderzoek uitvoert. Daarnaast is ze voorzitter van de Rotterdam Exchange Format Initiative, een groep van analyse-software-ontwikkelaars en expert users/trainers die momenteel samenwerken aan een uitwisselingsprotocol voor analysesoftware, en bestuurslid van KWALON.

In maart 2018 was het zover en konden we het eerste product lanceren vanuit Montreal: de Codebook Exchange Standard, oftewel: het kunnen exporteren van je codes (inclusief eventuele hiërarchie/groepering en bijbehorende annotaties) uit de ene software naar de andere software. Daarbij is ook een website gelanceerd ([www.qdasoftware.org](http://www.qdasoftware.org)), waar je meer over de groep kunt lezen en onze open source producten te downloaden zijn voor andere softwareontwikkelaars. Inmiddels is de Codebook Exchange in zeven van de acht deelnemende pakketten geïmplementeerd en heb je daar als gebruiker van die software toegang toe. In maart 2019 hopen we het ‘echte’ product te lanceren: de Project Exchange Standard, oftewel de mogelijkheid je hele project (inclusief data, codes, annotaties, memo’s en links) van de ene naar de andere software te migreren.

Een oud probleem is dan uit de wereld en nieuwe mogelijkheden dienen zich aan, denk aan het *buzz* woord van nu: *Big Data*.

Wat hebben Big Data hiermee te maken? Allereerst dat geautomatiseerde ‘text mining’ en codering op basis van computeralgoritmen, eventueel in combinatie met *Artificial Intelligence*, een toenemende rol spelen in de functionaliteit van analyse-software. Maar wat zijn eigenlijk Big Data? In ieder geval niet datgene wat wij als kwalitatief onderzoekers over het algemeen als onze ‘data’ zien: zelf verzamelde, empirische gegevens over mensen en hun leven. Door de komst van sociale media en internet is de grens tussen kwalitatief en kwantitatief komen te vervagen, omdat je zowel op sociale media als op internet snel vrij grote hoeveelheden materiaal kunt verzamelen. Op sociale media kan dat tegenwoordig vaak rechtstreeks via de hierboven genoemde QDAS, die daar dan wel vaak een tijdslimiet aan stelt, bijvoorbeeld alleen de zaken van de afgelopen week, wat weer gevolgen heeft voor de planning van je dataverzameling.

Zijn 2000 tweets dan Big Data? Of andere, grotere collecties die je op social media weg kunt halen? Het is maar de vraag of dit als ‘Big Data’ gezien kan worden. Wellicht kan de enorme bulk aan gedigitaliseerde bestanden uit de *Digital Humanities* (zie Van Dijck in KWALON 61, 21(1), p. 8-18), zoals erfgoedbestanden in de diverse musea of in de collectie van Europeana ([www.Europeana.eu](http://www.Europeana.eu)), of alle edities van een bepaalde krant via de Koninklijke Bibliotheek, eerder als Big Data gelden. Softwareontwikkelaars halen echter ook daar nog een beetje hun neus voor op. Kortom, of wij in de sociale wetenschappen snel met échte Big Data van doen zullen hebben is nog maar zeer de vraag. Waar we echter zeker wel mee van doen zullen hebben, en mogelijk in toenemende mate, is een grote en groeiende verzameling tekstuele bestanden (en in de naaste toekomst ook audiovisuele bestanden) die je onmogelijk allemaal handmatig kunt doorspitten, en waarvoor je dus geautomatiseerde ‘tools’ wilt gaan gebruiken, zoals bijvoorbeeld ‘text mining’, ‘voice recognition’ en ‘autocoding’ tools. Nou beschikten alle grotere QDAS-pakketten (ATLAS.ti, MAXQDA, NVivo) al over ‘autocoding’ en tekst-zoekfunctionaliteit, maar ‘textmining’ is van een andere orde. Bij textmining gaat de computer (op basis van algoritmes en woordenboeken) kijken naar het *samen voorkomen* van bepaalde (groepen) woorden en patronen daarin. Daarbij komt vaak ook ‘machine learning’ om de hoek kijken: de computer ‘voorspelt’ op basis van eerdere gevallen, hoe het volgende geval geïnterpreteerd moet worden. En dit voorspellen kan ‘supervised’

zijn (de mens kiest of de voorspelling in dit geval klopt) of ‘unsupervised’ (de computer doet het zelf). Sommigen zien dit als Artificial Intelligence, anderen hebben een veel nauwere definitie van Artificial Intelligence en houden het bij ‘machine learning’. Sommige QDAS-pakketten hebben zich vanaf het begin voornamelijk op ‘text mining’ gestort, zoals QDA Miner en Cassandra, andere zijn dat momenteel aan het inbouwen of hun functionaliteit ter zake aan het uitbreiden, zoals ATLAS.ti, f4analyse, NVivo of MAXQDA. Ook in de meer kwantitatieve hoek zijn er pakketten die zich bezighouden met ‘text mining’, zoals SPSS Text Analytics, KH Coder en Wordstat.

Reinoud Bosch, van huis uit econoom, politiek en sociaal wetenschapper en computertehnicus, is van mening (zie elders in dit nummer) dat wij ons nu als *kwantitatieve onderzoekers* Big Data toe moeten eigenen door *zelf* tools te programmeren om daarmee om te gaan. Dat is echter maar zeer de vraag. De tools die een kwalitatieve onderzoeker gebruikt in analysesoftware, zijn vaak afhankelijk van diens analyse-ervaring en de beschikbare tijd. De bestaande QDAS-pakketten hebben voldoende (en vaak meer) functionaliteit aan boord dan de gemiddelde onderzoeker nodig heeft. Goed programmeren is een vak dat – net zoals goed kwalitatief onderzoek – niet alleen kennis en talent, maar ook ervaring behoeft en vooral ook onderhoud. Bovendien heeft elke software een eigen ‘interface’, gebaseerd op de onderliggende ideeën en voorkeuren van de ontwikkelaar. De onderzoeker moet zich die interface steeds eigen maken.

Het lijkt me niet echt handig als we nog meer van die piepkleine softwarepakketjes krijgen, die allemaal één taakje verrichten. Dit verschijnsel zag je in de sociale wetenschappen vanaf de jaren tachtig en zie je momenteel vooral ook in *Digital Humanities*, omdat die vrij recent te maken kregen met digitale data. Een onderzoeker moet dan een hele trits van dit soort pakketjes zien te vinden, leren gebruiken en de resultaten ervan integreren om te krijgen wat zij/hij wil bereiken in de analyse. Wie heeft daar nu tijd voor of zin in?

Dan zijn we als onderzoeker toch beter af met multifunctionele generieke software die wordt ontwikkeld én onderhouden door zeer goede programmeurs? Door het bestaan van de Project Exchange Standard van REFI kan elke onderzoeker straks gebruikmaken van een hele keur aan zeer specialistische software, geprogrammeerd door ervaren, professionele programmeurs.