



# Data, koffie en house

Terwijl de versneller in Genève uit staat, werken studenten en promovendi gestaag door aan de analyse van een oceaan aan meetgegevens.

Tekst: Dorine Schenk

Op CERN in Genève of op Nikhef in Nederland, voor promovendus Luuk Vermunt van het ALICE-experiment is zijn werk overal hetzelfde. Waar hij ook is, hij zit achter zijn laptop, schrijft de software die metingen gericht moet analyseren, drukt op enter en haalt koffie. 'De analyses kunnen dagen duren, dus begin je gewoon aan de volgende analyse-codes. Er is altijd meer te doen.'

Vermunt is een van de tientallen studenten, promovendi en wetenschappers van Nikhef die dagelijks data uit de grote detectoren op CERN doorwerken. Op zoek naar meer details en eventuele afwijkingen van de standaardverschijnselen. Niet starend naar metertjes op CERN, maar in lange gangen met kleine kantoren met twee of drie mensen aan een bureau. Laptops. Een dikke dataverbinding. Koffie. En veel koptelefoons voor de concentratie.

Sinds december staat de LHC-protonenversneller na ruim drie jaar meten helemaal uit voor werkzaamheden. Maar voor de analisten van LHC Run2, de meetperiode tussen de zomer van 2015 en eind 2018, is het eigenlijk business as usual. Nog niet de helft van de meetgegevens bij de verschillende experimenten is serieus doorgewerkt. 'Bij de meeste echte publicaties zijn we nog maar in 2017, dus op de helft van de beschikbare data', zegt ATLAS-wetenschapper Ivo van Vulpen van Nikhef. 'Er zit gewoon nog heel veel aan te komen.'

## Vakantiefoto's

De bijbehorende getallen zijn in elk geval indrukwekkend. De LHC produceerde in Run2 ongeveer een miljard proton-protonbotsingen per seconde. Dat resulteert in 159 zogeheten inverse femtobarn, een maat die fysici gebruiken om de hoeveelheid botsingen te meten: bijna 160 miljoen botsingen, die in principe elk een waaier aan signalen in alle onderdelen van de detectoren kunnen geven. In totaal leverde Run2 op die manier ongeveer een petabyte aan ruwe data op, vergelijkbaar met zo'n tien miljard vakantiefoto's. Zoets zou een stapel cd's van bijna 2 kilometer hoog vereisen, exclusief de doosjes.

Hoe vind je iets nieuws in zo'n oceaan aan meetgegevens uit de detectoren? Er is simpelweg niet genoeg ruimte op de harde schijven van de experimenten om alles op te slaan, vertelt Katya Govorkova,

promovenda bij het LHCb-experiment. 'Daarom selecteren automatische triggers razendsnel de gebeurtenissen waar we in geïnteresseerd zijn en gooien ze de rest weg.'

Zo'n hardware-trigger kijkt eerst of de energie van een gebeurtenis hoog genoeg is om interessant te zijn. Een software-trigger bekijkt vervolgens welke deeltjes er ontstaan en bewaart alleen de metingen met interessante gebeurtenissen. 'Na deze triggers komt er nog steeds elke seconde een dvd aan data uit elke detector', rekt Govorkova snel uit.

## Code

Het zijn die geselecteerde metingen, met informatie over de deeltjes die ontstaan zijn, hun massa, snelheid, en vervaltijd, die naar de computers van onderzoekers gaan.

Die laten daar hun eigen zelfgeschreven analyseprogramma's op los, toegesneden op de specifieke verschijnselen die ze als fysicus bestuderen. Die code is dan van tevoren al uitvoerig getest met modellen en simulaties. De uitkomst van de analyses wordt vergeleken met simulaties en voorspellingen. Doen de deeltjes wat je verwacht? De kleinste afwijking kan wijzen op iets nieuws. 'We persen alle informatie uit de metingen', vat ATLAS-promovendus Marko Stamenkovic zijn werk samen. 'En hoe meer data hoe beter. Dat maakt het zekerder dat een opvallend ding niet een gewoon meetfoutje is.'

Dat niet alle gegevens van Run2 al gebruikt worden, komt niet alleen door het immense datavolume, legt Van Vulpen uit. Meetperiodes worden altijd afgewisseld met tijd om te sleutelen. 'Om de hoogste kwaliteit informatie uit de metingen te halen, moeten er eerst veel dingen gecontroleerd en gecalibreerd worden. Dat kost tijd. Voor de jaren 2017 en 2018 zijn zulke calibraties nog in de maak.'

Maar zelfs de eerste twee jaar data van Run2 leveren handenvol werk aan tientallen promovendi en wetenschappers. Dat lijkt eenzaam werk, vertelt LHCb-onderzoeker Jacco de Vries, maar is het niet. 'Er is wekelijks overleg met groepsgenoten in het buitenland. En bijna dagelijks met mensen op dezelfde gang. We overleggen over hoe we iets gaan aanpakken of stellen vragen als we ergens niet uitkomen.'

Om afleidende geluiden buiten te sluiten werken de meeste analytici met een koptelefoon op. Wat ze luisteren? 'Van Indie en rock, tot jazz en punk', zegt Vermunt. 'Maar halverwege februari werkte ik zelfs even met carnavalsmuziek.'

Als De Vries zich moet focussen houdt hij het vaak op klassieke muziek of deep house. 'Maar ik luister ook naar classic rock.' Govorkova, een geboren Russische, vertelt de laatste tijd zelfs de Nederlandse hiphop te ontdekken.

Wat in de zee van meetgegevens precies als een echte ontdekking geldt, verschilt per analyse. Overal worden grafiekjes gemaakt met daarin een lijn met de deeltjeseigenschappen die door modellen voorspeld zijn. In diezelfde grafiek worden de geanalyseerde metingen afgebeeld. Wijken die af van het voorspelde lijntje, dan heb je iets.

## Nieuwe kennis

Maar dat is dan pas het begin, vertelt De Vries. 'Daarna moet je in de details van je analyses duiken, en elk 'wiebeltje' in de metingen begrijpen voordat je zeker bent van een ontdekking.'

De Vries: 'Saai is dat nooit, maar het kost wel veel tijd en al dat gepuzzel kan frustrerend zijn en stress geven, vooral als er deadlines voor conferenties naderen. Maar uiteindelijk geeft het wel veel voldoening om iets volledig te begrijpen.'

Analyseren is altijd spannend werk, zegt Stamenkovic. 'Eerst weet je niets, je hebt alleen metingen. Daaruit leiden we soms zomaar nieuwe kennis af.'

