

Rosemary

TOEGANGSPOORT TOT DATASCHAT

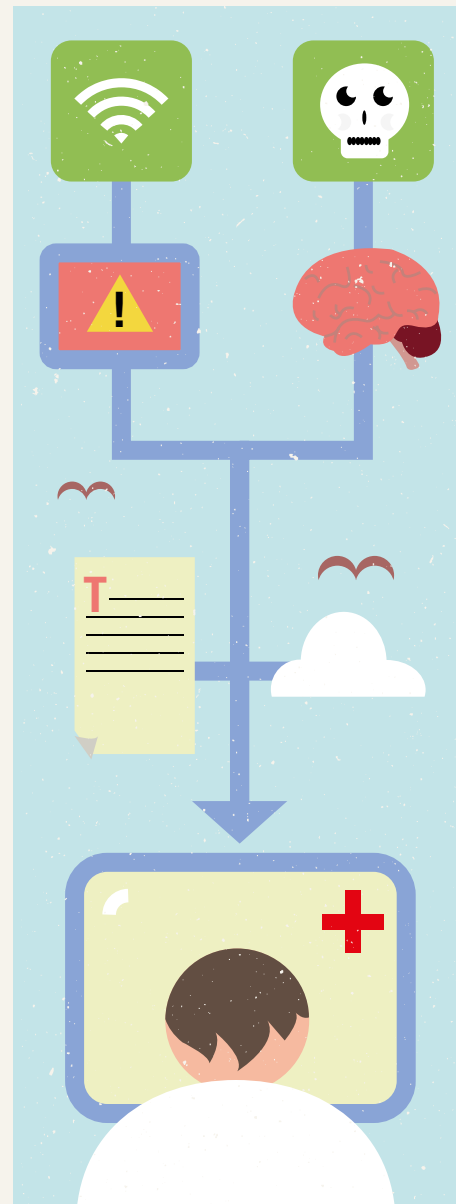
Terabytes aan data uit MRI-scans, CT-scans of DNA-analyses; ziekenhuizen krijgen steeds meer te maken met big data. Hoe ga je om met al dit soort grote files en zorg je dat er een gebruiksvriendelijke manier is om de juiste data bij de juiste persoon te krijgen? Shayan Shahand ontwikkelde tijdens zijn promotieonderzoek diverse science gateways, een soort toegangspoorten tot wetenschappelijke data, waarmee het mogelijk is deze op supercomputers eenvoudig te analyseren. De nieuwste versie, Rosemary, lijkt niet alleen nuttig voor onderzoek in het ziekenhuis, maar ook commercieel interessant.

Tekst Edda Heinsman
Illustratie Aniek Wisse

Even snel een 3D-plaatje maken van de hersenen? Shahand moet lachen. Zo'n representatie wordt opgebouwd uit meerdere MRI-scans waarbij in drie richtingen steeds een 2D-plakje van de hersenen wordt gescand. Het kost een computer 48 tot 72 uur om hier een 3D-model van te maken. Combineer je de scan met een scan waarop neutronenstralen zichtbaar zijn, kost dat al 100 uur. En dat geldt voor één proefpersoon. Dus bij een onderzoek van tweehonderd proefpersonen is dat tweehonderd keer honderd uur: 83 dagen non stop rekenen op één computer. 'Dat wil je dus niet op een gewone computer doen, maar op een supercomputer of cluster. En vanwege de ingewikkelde relatie tussen alle data, is goed databeheer noodzakelijk.'

Maar voor veel wetenschappers in het ziekenhuis is het werken met een supercomputer

niet eenvoudig. 'Toen ik begon aan dit project wilden de biomedici de supercomputers van SURFSara wel gebruiken, maar ze konden er niet mee uit te voet. Om gebruik van de supercomputer te vermijden, losten ze het



bijvoorbeeld op door een kleinere onderzoeksgroep te nemen. Mijn doel was een simpel te gebruiken systeem te bouwen, een 'toegangspoort' tot de data en supercomputers, een science gateway.'

ROSEMARY

'Eerst moest ik duidelijkheid creëren: wie heeft wat nodig, welke data moeten door welke computerprogramma's gezien worden en hoe krijg je gegevens bij de juiste persoon?' Shahand identificeerde drie pilaren die betrokken zijn: data, de verwerking daarvan en de samenwerking. Vervolgens ging hij aan de slag met het maken van een gebruiksvriendelijke toegangspoort voor big data. De verschillende prototypes die hij ontwikkelde, werden ingezet bij colleges neuroscience en gebruikt door neuro- en bio-informatica wetenschappers. Elke generatie science gateways werd beter. Inmiddels heeft hij bij de e-science groep op het Amsterdam Medisch Centrum de vijfde generatie gateway uitgebracht. Dit is de eerste versie die hij een naam heeft gegeven: Rosemary. 'In de tuin van mijn vader vroeger groeide veel rozemarijn. En ik houd heel veel van de geur bij het koken', lacht hij.

'Met Rosemary kunnen de neurowetenschappers hun onderzoek opschalen en zo statistisch betere resultaten halen. Gebruik is simpel en data toegankelijk. Onderzoekers hoeven zich geen zorgen meer te maken over waar ze hun data bewaren, hoe het over te zetten naar een andere computer. Stappen waar eerst mensen een file handmatig over moesten dragen, gaan nu automatisch.'

Rosemary is inmiddels zo ver ontwikkeld dat de gateway makkelijk aangepast kan worden en ingezet binnen uiteenlopende vakgebieden. 'Een masterstudent maakte een andere versie in een maand.' De science gateway worden gratis *open source* aangeboden, maar de ontwikkelaars worden wel betaald om het systeem aan te passen aan de voorkeuren van de gebruiker. 'Op die manier is het commercieel toch interessant.'

COMMIT/project Data2Semantics
Partner: Amsterdam Medisch Centrum

COMMIT/ is een publiek-private ICT-onderzoeksprogramma.