

# CIT neemt deel aan miljoenenproject Target

Het Centrum voor Informatie Technologie (CIT) van de RUG gaat de ondersteuning verzorgen van een uitgebreid hardwarepark in het kader van het Target-programma. Het Samenwerkingsverband Noord-Nederland (SNN) en de provincies Groningen en Drenthe hebben onlangs in totaal 16 miljoen euro toegekend aan het project Target.



**T**arget ontwikkelt methodes om zeer grote hoeveelheden data inzichtelijk te maken, bijvoorbeeld van sensornetwerken, astronomie-observatoria en medische meetgegevens. Edwin Valentijn (Sterrenkunde) en Robert Janz (CIT) zijn initiatiefnemers van het project. Target is een samenwerking tussen de Rijksuniversiteit Groningen, het Universitair Medisch Centrum Groningen (UMCG), IBM, Oracle, Task24, ASTRON en Heeii.

## Van wetenschap tot markt

De kennis en technologie van Target wordt tot nu toe ontwikkeld binnen diverse onderzoeken die gebaseerd zijn op zeer grote datastromen. Voorbeelden hiervan zijn de wetenschappelijke verwerking van DNA-profielen en ander biomedische gegevens van 165.000 deelnemers aan het Lifelines-project (UMCG), het radiosensornetwerk van LOFAR, het kosmologische 'dark energy'-project KiDS (OmegaCen) en de herkenning van 100 kilometer handgeschreven teksten van het Nationaal Archief (Scroogle, Kunstmatige Intelligentie). Doel is om de academische methodes en gereedschappen geschikt te maken voor commercieel gebruik. De verwachting is dat onder andere een aantal spin-off bedrijven zal ontstaan met hulp van de Target Holding.

## Gigantische datastroom

In de hedendaagse maatschappij neemt het gebruik van sensoren steeds meer toe, van beeldsensoren in foto- en videocamera's tot sensoren in tekstscanners, DNA-scanners, tot netwerken zoals LOFAR. Deze sensornetwerken genereren een gigantische datastroom die dient te voldoen aan de informatiebehoefte die wordt versterkt door de recent enorm gegroeide toegankelijkheid van grote hoeveelheden data via internet. In toenemende mate wordt de vraag naar informatie toegespitst op de vraag hoe

uit grote hoeveelheden data op een efficiënte manier zinvolle en betrouwbare informatie kan worden gehaald.

## Intelligente informatiesystemen

Het huidige verwerkingsmodel dat door wetenschappers gebruikt wordt en de pakketten die door commerciële IT-bedrijven worden aangeboden, kennen een groot aantal technologische beperkingen. Dit heeft vooral betrekking op schaalbaarheid en verandering van methodieken in research-omgevingen.

Target heeft een geheel nieuwe methode ontwikkeld waarmee op een efficiënte wijze toegang kan worden verkregen tot en gebruik kan worden gemaakt van grote en groeiende hoeveelheden data van verschillende structuur.

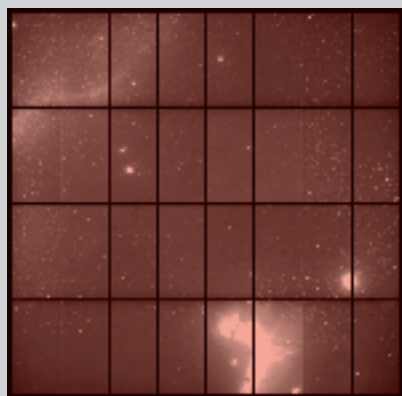
De gebruiker stelt de vraag en het systeem (re) construeert de informatie door middel van 'pulling' van de benodigde data uit de zee van gegevens en betrouwbare verwerking ervan. Dit is in tegenstelling tot de klassieke technologie die zich richt op de 'pushing' van de zee van data naar de gebruiker toe. Het idee achter Target is om alle informatie, de bewerkingstappen en de gegevens over de informatie te bewaren en met elkaar te koppelen

## GRID-technologie

Een aanzienlijk deel van de investering van Target wordt gebruikt voor de bouw van een datacentrum met een capaciteit van tenminste tientallen petabytes. Daarnaast komt er een grootschalig databasesysteem dat een centrale rol krijgt in de dataverwerking.

Het CIT, IBM, Oracle en de Target Holding zijn samen verantwoordelijk voor de gehele ICT-keten. Zij zullen de vertaalslag maken van de functionele concepten naar nieuwe ICT-architecturen en de implementatie daarvan voor grootschalige dataverwerkingomgevingen. Het onderzoek richt zich hierbij op de koppeling tussen verschillende types en gedistribueerde databases, de koppeling met de beschikbare rekenkracht en de toepassing van GRID-technologie.

De rekenfaciliteiten waar op dit moment reeds gebruik van wordt gemaakt bij het CIT bestaat uit een GRID-rekencluster dat onderdeel uitmaakt van de Europese GRID-infrastructuur EGEE, de IBM-supercomputer BlueGene/P die in het kader van het LOFAR-project van Stichting ASTRON bij het CIT staat opgesteld en het HPC-cluster bestaande uit 200 rekennodes. 



## Veel data

De astronomieonderzoeksschool Nova is in 2000 begonnen met het bouwen van een 256 megapixel CCD-camera die elke paar minuten een hemelfoto gaat maken, jarenlang voor driehonderd nachten per jaar. Dat levert gigantische hoeveelheden data op. Het zoeken naar interessante gegevens vereist daarom een nieuwe aanpak.



[www.rug.nl/target](http://www.rug.nl/target)