



AlmereGrid

AlmereGrid, van *science Grid* naar *city Grid*

ISOC
24 januari 2005

Ad Emmen/m.m.v. Bart Buijs

© AlmereGrid

AlmereGrid wordt mede mogelijk gemaakt door een subsidie van de Europese Unie en de Provincie Flevoland en de partners: SARA, Rabobank, Capgemini en NCF.

Zoekmachine Digitale speurtechniek bekijkt alleen visuele kenmerken

Net zoiets als dit plaatje, graag

Afbeeldingen zoeken op internet of in databanken moet nog altijd op trefwoord of datum. Speuren op basis van beeldkenmerken zou veel handiger zijn. Vind foto's van koeien aan de waterkant. Met een donkere wolkenlucht. Door **Ben van Raaij**

Wie digitaal fotografeert, kent het probleem: hoe hou je het overzicht in die uitdijende bestanden vakantiefoto's, en hoe vind je een bepaalde foto terug? Catalogus-software op basis van namen, data of trefwoorden is niet ideaal, zeker als je onderscheid moet maken tussen 54 keer kleine Fleur met schepje op het strand.

Beeldbanken als Corbis of Getty Images en internetzoekmachines als Google hebben hetzelfde probleem in het groot: een exponentieel groeiende massa foto's, video-opnamen en ander beeldmateriaal en ontoereikende technieken om die goed te ontsluiten.

Want alle zoeksystemen zijn nu aangewezen op bijschriften en aftitelingen - tekst. Wil dat meer opleveren dan toevalstreffers, dan moeten namen, trefwoorden of omschrijvingen kloppen en de archivering op orde zijn.

Het zou dus mooi zijn om direct en gericht op beeld te kunnen zoeken. Aan zulke technieken voor het automatisch ontsluiten van multimediacollecties (*multimedia retrieval*) wordt hard gewerkt. Dr. ir. Thijs Westerveld van het Centrum voor Wiskunde en Informati-

ca in Amsterdam ontwierp een eigen prototype, waarop hij donderdag aan de Universiteit Twente promoveerde.

Westervelds systeem zoekt op beeldkenmerken (visuele inhoud) in plaats van bestandsnamen, trefwoorden of beschrijvingen (tekstuele context). En dat levert veel meer zoekmogelijkheden op. In theorie kun je er afbeeldingen mee zoeken waar een gewone zoekmachine niks mee kan, zeg koeien aan de waterkant onder een dreigende wolkenlucht.

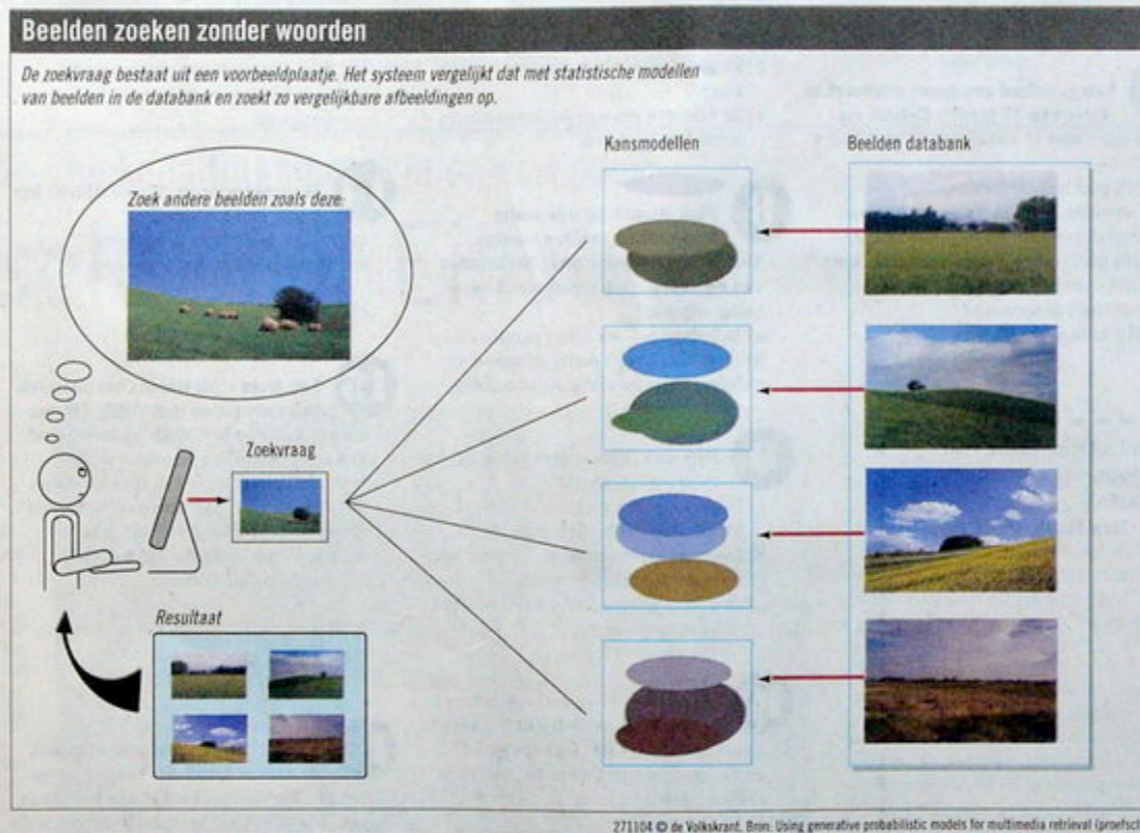
Het uitgangspunt is een voorbeeldplaatje - dat je dan wel al gevonden moet hebben - en de zoekvraag 'vind meer plaatjes zoals deze'. Westervelds programma zoekt dan in de beeldcollectie vergelijkbare foto's op.

Daartoe wordt van elk beeld in de collectie een 'generatief kansmodel' gemaakt, een statistische beschrijving van de visuele kenmerken van het beeld. Zo'n beeldmodel - een abstracte rangschikking van elementen en kleurvlakken - staat los van de voorstelling.

Westerveld: 'Je knipt zo'n plaatje in heel veel kleine stukjes. Die stukjes worden dan automatisch gegroepeerd in een aantal componenten, zeg lucht, gras en water. Dat levert voor elke component een gemiddelde kleur, positie en textuur op. Je ontleedt het plaatje dus tot enkele clusters van gemiddelden.'

En hoe gaat het zoeken dan in zijn werk? 'Als het systeem een zoekvraag krijgt, in de vorm van een voorbeeldplaatje, berekent het van alle modellen in de collectie de kans dat het voorbeeld gegenereerd kan worden uit dat model. Het systeem laat dan de beelden zien die corresponderen met de waarschijnlijkste modellen: de foto's die het meest lijken op het voorbeeld.'

Het voordeel van de automatische modellering is haar 'generiek karakter'. Het systeem speurt anders dan de



meeste beeldzoekmachines niet naar afgebakende objecten. En het is evenmin aangewezen op specifieke collecties, zoals gezichtsherkenningsoftware alleen met pasfoto's werkt. Elk soort beeld of collectie is geschikt.

Westervelds algoritme werd met succes beproefd door TRECVID, een met CNN- en ABC-nieuwsvideo's gevulde

Amerikaanse databank. De TRECVID-tests - met opgaven als: zoek shots van opstijgende vliegtuigen - zijn de Olympische Spelen in de wereld van de beeldzoekmachines. 'Mijn methode deed het zeker zo goed als andere. Maar vooralsnog zijn alle technieken slecht', zegt Westerveld bescheiden.

Die matige resultaten ontmoedigen

hem niet. 'Er is geen principiële reden waarom het zoeken op beeldkenmerken niet veel beter kan.' Dat je alle plaatjes in een beeldbank moet modelleren, is geen punt: dit is niet meer werk dan het maken van een bijschrift. Een probleem is wel dat het vergelijken van een foto met alle modellen in een collectie heel wat rekenkracht vergt.

Vooralsnog verwacht Westerveld het meest van een gecombineerde aanpak van zoeken op beeld én tekst. 'Je moet de statistiek van beeldmodellen combineren met die van taalmodellen. Zo kun je optimaal profiteren van de goede systemen die al voor tekstzoeken ontwikkeld zijn. Gecombineerd zoeken levert gemiddeld betere resultaten op dan zoeken op alleen beeld of tekst.'

Dat ook grote bedrijven zoeken naar manieren om beelden in digitale archieven en op internet beter te ontsluiten, bleek vorige week. Onderzoekers van Xerox Europe in Grenoble hebben een systeem ontwikkeld om *image content* te kunnen opbergen en raadplegen. Het leert alledaagse objecten zoals auto's te rubriceren en weer opsporen op basis van kenmerken of *patches*: een wiel en een koplamp betekenen een auto.

Maar dit soort beeld- of patroonherkenningstechnieken, waarbij speciale detectoren objecten als gezichten of huizen kunnen identificeren, zijn wezenlijk anders, zegt Westerveld. 'Mijn systeem werkt algemener. Ik vergelijk plaatjes rechtstreeks op het niveau van de afbeelding zelf. De inhoud van het beeld doet zoals gezegd niet terzake.'

Voorlopig staan alle technieken nog in de kinderschoenen. Zoeken in brede beeldbanken op vage omschrijvingen is nog ver weg. 'Maar gericht zoeken binnen specifieke collecties, bijvoorbeeld tennissers op het centrumcourt, moet binnen enkele jaren haalbaar zijn.'

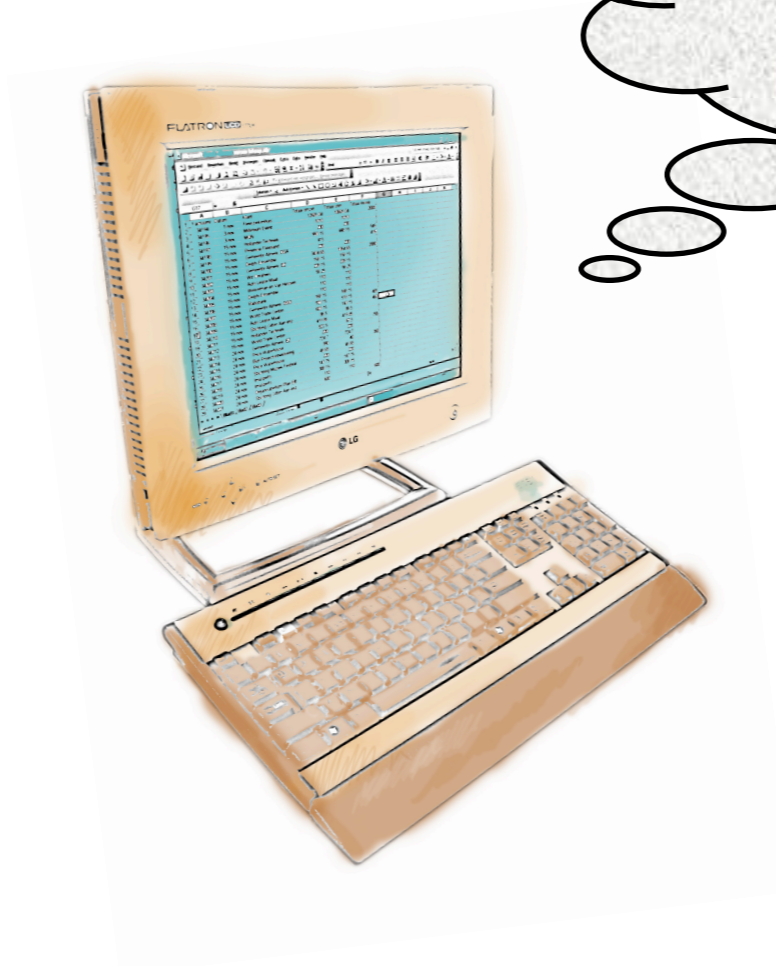
De toepassingen zijn potentieel eindeloos: medische foto's, politiefoto's, bewakingsvideo's of satellietopnamen. Veel interesse heeft niet voor niets het Amerikaanse leger, dat de TRECVID-wedstrijden sponsort. De consument zal echter nog wel even geduld moeten hebben voor zijn digitale schoenendoos met strandfoto's aan de beurt is.

Uitgangspunt

Ongebruikte
computercapaciteit benutten
voor wetenschappelijk
onderzoek

Mogelijk door

Breedband in
Almere





AlmereGrid

 Wie

 Wat

 Hoe

 Wanneer

 Waarom



Wie?

Stichting AlmereGrid

Bestuur

Wim van der Sanden
Marcel Kolder
Jan Seijdel

Directie/projectleiding

Ad Emmen*

Secretariaat

Leslie Versweyveld

Communicatie

Bart Buijs*
Nancy de Koning

Beheer

Marien Vlug*

* *Initiatiefnemers*



Wie?

Adviesraad

Anwar Osseyran (SARA)
Harrie Vollaard (Rabofacet)
Dolf Zantinge (UNET)
Ron Tolido (Capgemini)
Tonny Triezenberg (IIE)

Technology Board

Jan Jacobs (Rabofacet)
Bram Dees (Capgemini)
Axel Berg (SARA)
Cor van der Struijf (IBM)

Technology Providers (Partners)

Rabofacet
Oracle
Engage Technology
UNET
IBM
Capgemini
LogicaCMG
Alca
Foundry Networks
SARA

NCF

Stage Providers

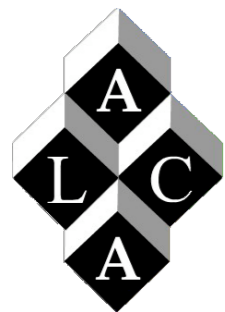
IIE – member survey study
RIVO – grid comparison study



Wie?



Rabobank





AlmereGrid



Wat?

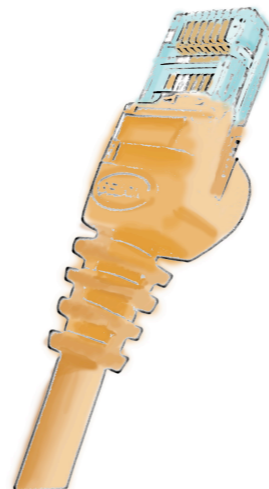
*Een decentraal heterogeen
computational grid,*

*waarop complexe berekeningen kunnen
worden uitgevoerd.*

Wat?

‘leden’

+200 processoren test fase
gezamenlijke rekenkracht



Almere Fiber Pilot
2200 adressen

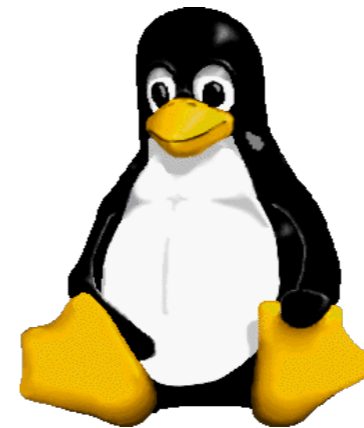
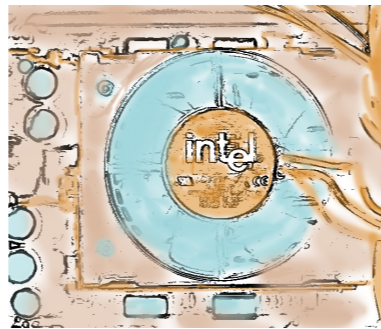


bedrijven



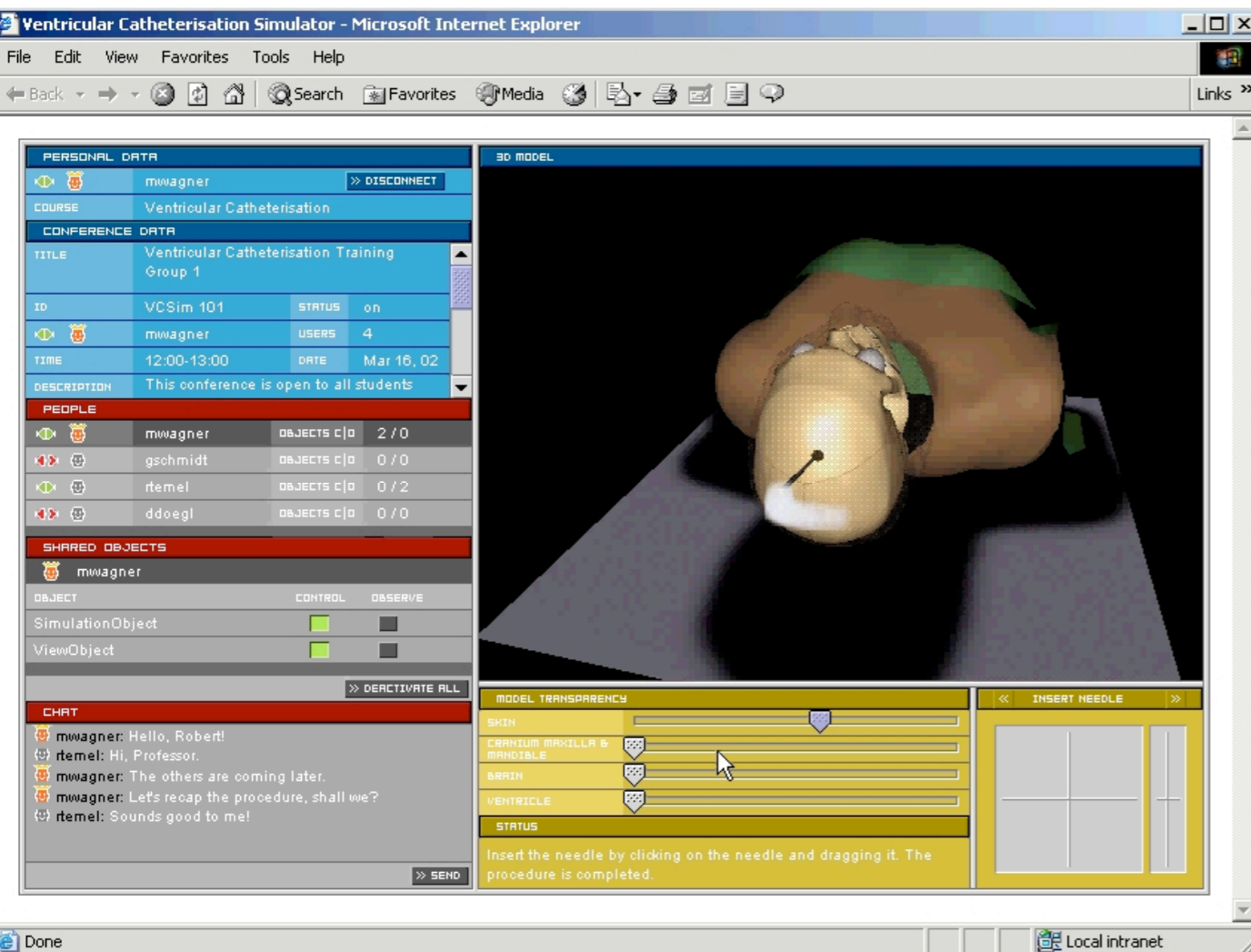
bewoners

Wat?



resources van leden

Wat?



Rekenwerk

Medisch onderzoek

Luchtvaarttechniek

Imageprocessing

Rendering

Interessant door breedband

“Pleasantly Parallel”
toepassingen met veel data



AlmereGrid

Hoe?

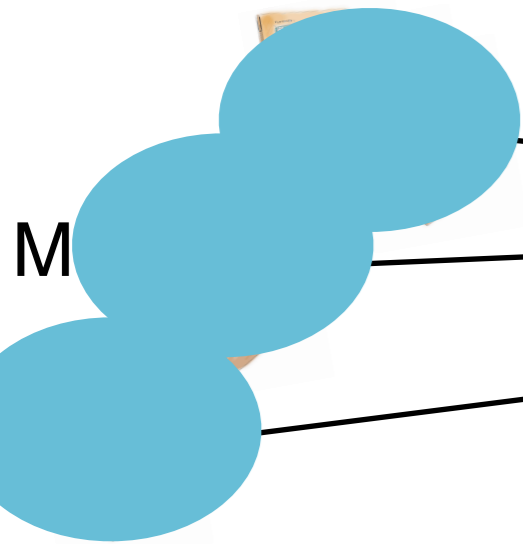
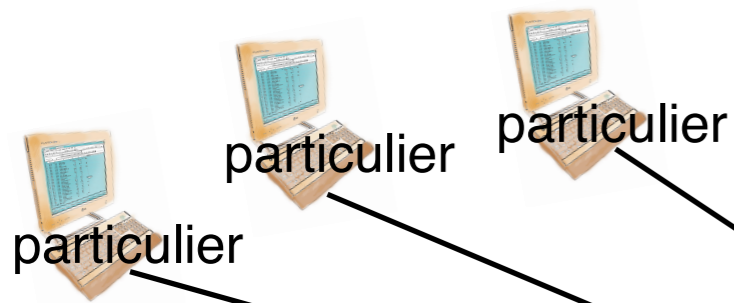
Leden stellen ongebruikte computercapaciteit ter beschikking.



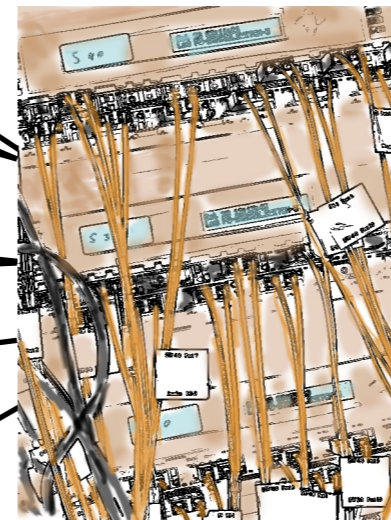


Hoe?

*De AlmereGrid Exchange
verdeelt taken over de leden*



leden







onderzoekers



Hoe?

Werkwijze

-  Open project (voor partners)
-  Per onderwerp relevante experts van technology partners erbij betrekken
-  Studenten RIVIO en IIE doen activiteiten op meer technisch en communicatiegebied
-  Verslaglegging en rapportage



AlmereGrid

Wanneer?

Tijdspad: van Scientific Grid naar CityGrid

Tentative road map








Scientific Grid	Start-up phase	Test phase	Production Grid	Expansion phase	
Community Grid			Start-up phase	Test phase	Production Grid



Wanneer? Nu:






Survey gridsoftware

-  Eigenschappen desktop Gridsoftware, client- en serverside
-  Stabiliteit, efficiency, encryptie, makkelijk te integreren met o.a. AlmereGrid
-  Benchmarks: metingen over o.a. breedband
-  Eerste circa 10 door AlmereGrid. Daarna testprocedure beschikbaar
-  Uitvoering: RIVIO studenten



Wanneer? Nu:

Survey MKB en particulieren

-  Nulmeting: aansluiting netwerk, bekendheid etc.
-  Identificeren van verwachtingen die positief gebruikt kunnen worden
-  Identificatie van mogelijke barrières
-  Opbouw relatie met MKB
-  Uitvoering: studenten IIE

Wanneer?



*Testgrid Rabotoren/
IIE/bedrijven/SARA*

Testen infrastructuur

Basis Grid software

Communicatie



AlmereGrid



Waarom?

Omdat het kan in Almere

Almere Fiber Pilot (100 mb/s)

SARA

UNET

Rabobank

Almere Kennisstad

Initiatiefnemers



Waarom?

Leden

Betrokken bij het goede doel

Maatschappelijk verantwoord ondernemen





Kennisontwikkeling en voorsprong

On-demand gebruik pioniers

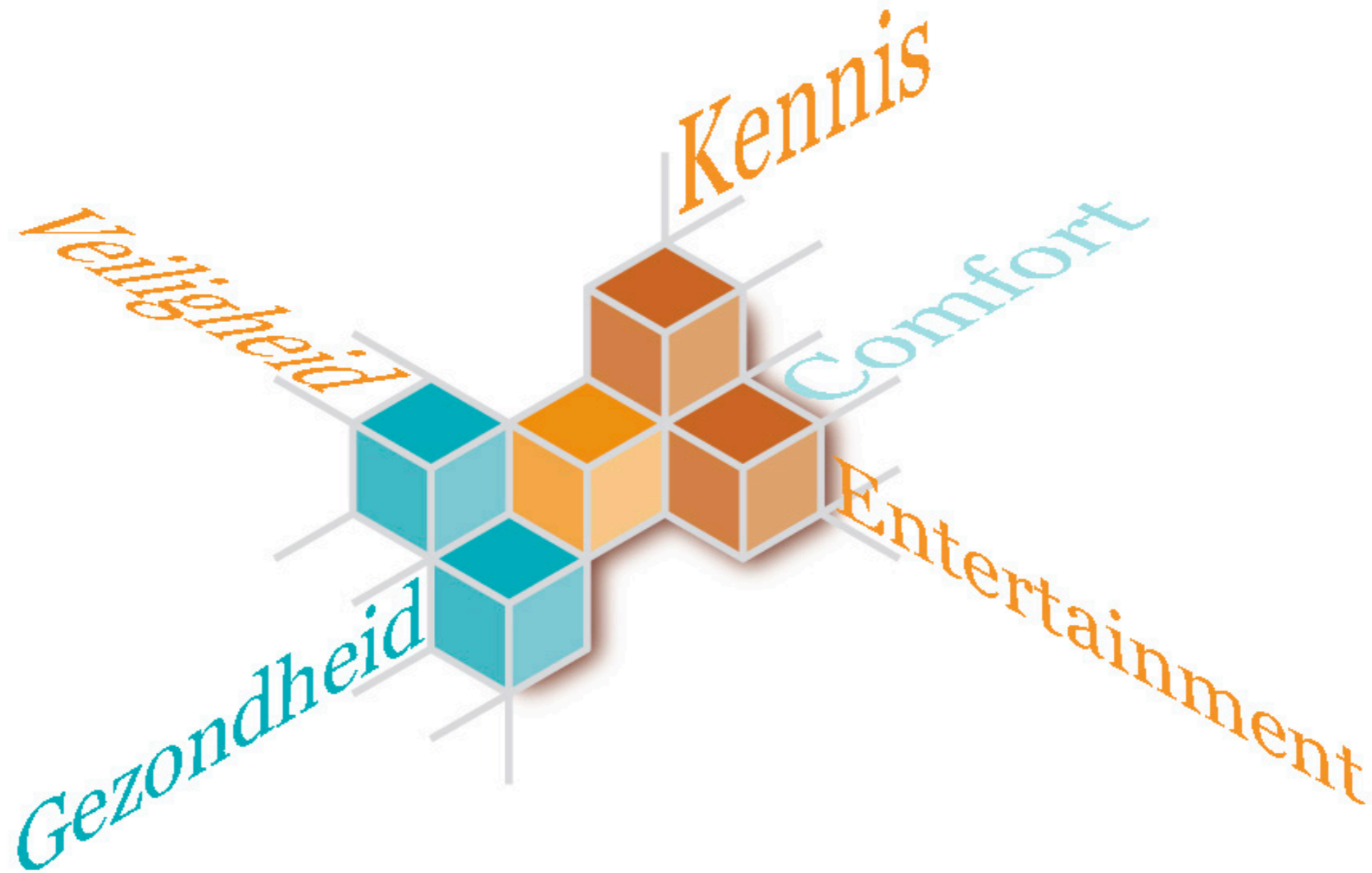


Waarom?

Visie

-  Grote publiek bekend maken met Grid
-  Waardering on-demand services
-  Grid providers ontwikkelen aanbod rekenkracht, opslag, muziek, film, literatuur, veiligheid, zorg, etc
-  Ontwikkeling community Grid: CityGrid

Waarom?



Powered by



AlmereGrid