

# *Voorstellen Enlighten Your Research*

## **Table of Contents**

1. The PALFA/GBT pulsar survey project: a study of extreme physics laboratories .....	2
2. 2. A lightpath for simultaneous fMRI of social interaction	3
3. High speed light path network for the Netherlands Bioinformatics for Proteomics Platform.....	4
4. 3D Visualisatie van Complexe Netwerken.....	5
5. Lightpaths for Opportunistic Spectrum Access .....	6
6. Interdisciplinary and interinstitutional data collection for traffic simulation and travel behavior modeling .....	7
7. Gedistribueerd hoge kwaliteit audiovisueel testbed .....	9

# 1. The PALFA/GBT pulsar survey project: a study of extreme physics laboratories

## Onderzoeksvoorstel

**Titel:** The PALFA/GBT pulsar survey project: a study of extreme physics laboratories.

**Beschrijving:** Radio pulsar research provides unique opportunities for testing our most basic understanding of the nature of matter and the universe. The surface gravity of these extremely compact stars, about 1 billion times the gravity on Earth, is the largest of any object in the Universe subject to observation. The internal densities of ten times nuclear density have not existed since the Universe was about 1/1000th of a second old.

For these reasons, we have initiated several large-scale pulsar surveys that aim to discover rare objects especially suitable for their physics and astrophysics payoffs. Currently the two largest-scope projects in the world are the /Arecibo ALFA survey/ and the /Green Bank drift survey/.

Searching radio-telescope data for pulsars is very computationally intensive, currently about a thousand times real time per processor. The data from these two biggest radio telescopes in the world is therefore distributed over collaborating institutions. Both based in the Netherlands, investigators van Leeuwen and Hessels and their groups have very strong stakes in both these projects (cf. our in discoveries in Science\* this year).

The work flow for analysing radio-pulsar data requires several high-bandwidth steps:

- telescope ->
- raw data storage ->
- computing ->
- candidate database ->
- visualisation to workplace

Currently these resources are distributed over the Netherlands, with raw data storage and the candidate database in Groningen, computing in Dwingeloo and at SARA, but most person power at the Astronomical Institute in Amsterdam. Given our peak band-width needs of roughly 100MB/s (800Mbps) between raw-data storage and computing resource, we are currently having disk packs shipped to Dwingeloo, a cumbersome operation. More importantly, the limited band width available in the visualisation step has forced us to do standard preprocess all data on the database side for quick access by the investigators. This requires significant extra computing on all data, although expert investigators and students in practice single out the less than 0.1% of most-promising data for perusal. As such, a similar 100MB/s bandwidth between database and workplace would finally allow for dynamic visualisation at the workplace.

We therefore propose to further investigate having a dynamic light path setup between our four resource locations:

- 1) University of Amsterdam Astronomy department
- 2) ASTRON Astronomy department
- 3) Groningen HPC/V
- 4) SARA Amsterdam

## 2. A lightpath for simultaneous fMRI of social interaction

### Onderzoeksvoorstel

Titel: A lightpath for simultaneous fMRI of social interaction

Everyday we effortlessly interact with other people. To capture how the brain achieves this social behavior is the ultimate goal in the field of social neuroscience.

Playing the game Charades in the fMRI scanner In most experimental paradigms social processes are studied by having participants watch pictures or movies of other people while monitoring brain activity. While informative, these studies are unfortunately still far from resembling real social interactions. In order to find out which brain areas play a role in social communication, we therefore went beyond this and had romantic couples play the game Charades ('hints' in Dutch) in an fMRI scanner (Schippers, Gazzola, Goebel & Keysers, submitted). The participants took turns going in the scanner and either gestured words that were written on the screen (while their hand movements were video recorded) or tried to guess from the previously recorded video what their partner was trying to depict. Results showed that areas in the brain that are associated with mirroring other people are strongly implicated, while higher-level mentalizing areas are not.

Linking two fMRI scanners Although closer to real social interactions than previous studies, this experimental setup still lacks true interplay between people. During natural communication, we adapt our behavior to the reactions they evoke in others, an interactivity that our video-recording technique cannot provide. This drawback could be overcome by linking two fMRI scanners through a high-speed connection. This would allow two participants from cities as far apart as Groningen and Maastricht to play the game charades interactively. The participants will then communicate in real-time with each other through a video stream. To our knowledge, scanning two people at the same time while they are interacting has only been done twice (Montague et al., NeuroImage, 2002; Goebel et al., 2005, HBM Conference). The scarcity of such experiments is mainly due to difficulties finding a location where two fMRI scanners are readily available and difficulties getting a reliable connection if two video streams and scanners have to be connected over a greater distance.

Reliability, velocity and security: exactly what we need! The connection needs to be highly reliable and fast: we all know from bad phone connections how artificial social interactions become if delayed. Additionally, investigating how one partner's brain responds to the other critically depends on the temporal resolution with which data from the two scanners can be compared. Finally we need a secure connection, because the video-channel contains private information of the participants. Existing connections fail to meet these criteria but a lightpath of SURFnet would give us the reliable, high-speed and secure connection we need.

Practical considerations The project would be embedded in the PhD-project of Marleen Schippers under supervision of Prof. dr. Christian Keysers (both working at the University Medical Center Groningen) and in collaboration with Rainer Goebel from the Maastricht Brain Imaging Center. Together we have the experience and open up a new era in brain research.

### 3. High speed light path network for the Netherlands Bioinformatics for Proteomics Platform

#### Onderzoeksvoorstel

**Titel:** High speed light path network for the Netherlands Bioinformatics for Proteomics Platform

**Beschrijving:** The Netherlands Bioinformatics for Proteomics Platform (NBPP) is a joint initiative of the Netherlands Bioinformatics Center (NBIC) and the Netherlands Proteomics Center (NPC). The NBPP intends to provide a comprehensive hardware and software framework for the Dutch proteomics community by integrating tools for data management, data integration, processing and evaluation, knowledge extraction and data mining, specifically applied in proteomics experiments. The main tool in proteomics analytics is the mass spectrometer, which produces large amounts of complex data (mass spectra). Several complex bioinformatics workflows were, and will be developed in the participating laboratories of NBPP focusing on specific types of proteomics experiments. These workflows will be integrated in the common platform. The NBPP is designed to allow high throughput proteomics data-processing and evaluation, using shared resources and a parallel processing environment. Besides using the Dutch Life Science Grid (DLSG) as main hardware resource, NBPP requires high speed connections between the local data storage and processing units (clusters) to pre-process large amount of data and/or to perform a large number of short tasks. The first steps in proteomics data processing workflows is the substantial reduction (generally with a factor of 100-1000 or more) the vast amounts of raw mass spectrometric data. Therefore dedicated data extraction and integration algorithms are used. One example of such a workflow is peak quantification using diverse peak-picking algorithms followed by isotope peak clustering and deconvolution. On the other hand many data processing steps at an advance stage of proteomics bioinformatics workflows, contain large numbers of short tasks, such as parsing between diverse data formats for a large number of small files. For short jobs and large raw mass spectrometry data files, preprocessing tasks have high I/O needs for which the current GRID infrastructure is not well suited. We are applying for the lichtpad "enlighten your research" initiative in order to provide high speed connections between the hardware and storage resources of the main participants of NBPP and to obtain an optimal network infrastructure for the NBPP, where high I/O demanding or short tasks can be performed locally at the NBPP sites, and where middle and long term tasks with a reasonable I/O are executed on the DLSG. Currently a major bottleneck of high throughput proteomics experiments is the limited throughput of data processing and experimental evaluation. Proteomics projects are often delayed because the acquired data is analyzed much later by the overwhelmed bioinformatics and biostatisticians. Mass spectrometry based proteomics is still an emerging field and even more laboratories start to use high throughput proteomics profiling experiments adding even more stress on this bottleneck. Providing an integrated bioinformatics environment with optimization of processing and I/O time using lichtpad and DLSG will enable NBPP and the proteomics community to processes and evaluate their data in shorter time.

**Extra:** To date it is still difficult to exactly pinpoint where the precise location of the lichtpad connection will be. This is due to the fact that the participants in the NBPP to some part hiring the personnel and finding locations of their equipment and Grid nodes. It is to be expected that this will be resolved within the forthcoming six weeks.

## 4. 3D Visualisatie van Complexe Netwerken

### Onderzoeksvoorstel

Titel: 3D Visualisatie van Complexe Netwerken

Beschrijving: Onze maatschappij draait op complexe netwerken die zich kenmerken door grootschaligheid en een onregelmatige structuur. Voorbeelden van dergelijke netwerken zijn onze infrastructuren (bijv. het Internet), sociale netwerken, en biologische netwerken (bijv. ons brein).

Het belang van complexe netwerken voor onze maatschappij onderschrijft de noodzaak van onderzoek naar de structuur en eigenschappen van complexe netwerken, welke tegenwoordig nog onvoldoende worden begrepen. Aan de TUDelft hebben wij verschillende kenmerken van complexe netwerken analytisch in kaart gebracht. Wij hebben programmatuur ontwikkeld, die gegeven een netwerk kan uitrekenen wat de diameter, "betweenness", spectrale waarden, etc. zijn. We hebben een scala aan parameters, maar allen geven één of meerdere waarden terug. Voor een beter begrip van complexe netwerken zullen we netwerkeigenschappen moeten kunnen visualiseren (en bij voorkeur in 3D, zodat we de netwerken kunnen draaien en manipuleren). Een dergelijke koppeling tussen analyse en visualisatie zal ons bijvoorbeeld kunnen helpen bij:

- Het observeren hoe een complex netwerk groeit (gegeven opgeslagen historische data)
- Het aantonen waar zwakke plekken zitten in een netwerk (na analyse door onze programmatuur)
- Het netwerk daadwerkelijk in perspectief plaatsen (tegenwoordig worden netwerken vaak "platgeslagen" waardoor we informatie verliezen)
- Etc.

Door de grootte van complexe netwerken gecombineerd met de 3D visualisatie, worden grote verkeersstromen gegenereerd (in één richting als we enkel de data binnenhalen en bewerken/weergeven, of in twee richtingen als de bewerking/visualisatie real-time bij beide partijen wordt weergegeven). Middels het gebruik van lichtpaden kunnen we in staat worden gesteld om real-time data over complexe netwerken binnen te halen, te visualiseren, te analyseren, en te versturen.

Eén van de meest complexe, maar tevens ook meest fascinerende, netwerken dragen we met ons mee: namelijk onze hersenen. Onze hersenen bevatten grofweg honderd miljard neuronen (knooppunten) en een veelvoud aan verbindingen. En ondanks deze complexiteit is ons brein een efficiënt en robuust netwerk waar we veel van kunnen leren.

Momenteel werken wij met de groep van Prof. Stam (VUmc) samen om hersennetwerken in kaart te brengen. In het kader van dit voorstel zal dit onze focus zijn. De groep van Prof. Stam heeft een schat aan data, die we ook nog eens 3D willen weergeven. Wij hebben kennis van netwerken en tools om er iets over te zeggen. Wij zouden daarom graag een lichtpad tussen TUDelft en VUmc opzetten om hersennetwerken te kunnen analyseren en visualiseren. Mocht deze "case study" succesvol zijn, dan kunnen in de toekomst met onze programmatuur ook andere complexe netwerken gevisualiseerd en geanalyseerd worden (middels lichtpaden naar en tussen andere instituten).

## 5. Lightpaths for Opportunistic Spectrum Access

### Onderzoeksvoorstel

Titel: Lightpaths for Opportunistic Spectrum Access

Beschrijving: To sustain the growth of the wireless networks and services enough radio spectrum needs to be available. However due to the strict regulation in the spectrum management there are no new free slots available in the spectrum charts - all of them are assigned to existing services. Many of them are used sporadically, like pager channels, but no one except for those who acquire the license are allowed to utilize the empty frequency slices. Thus if the radio regulatory agencies allowed to exploit the voids in spectrum by the third parties, the spectrum shortage problem would be of no issue. Such an idea is in the eye of many researchers around the world and is called Opportunistic Spectrum Access (OSA) (or more commonly but less accurately as Cognitive Radio).

One of the biggest problems associated with OSA is the issue of reliable spectrum opportunity. That is because of propagation conditions and noise nodes looking for spectrum opportunity can never be sure of the actual status of the licensee. And also because of the temporal variations in spectrum occupancy (licensed users come and go) secondary nodes need to scan frequencies very rapidly. Now, to model the spectrum occupancy statistics a simple model is assumed. However still little is known on how to describe spectrum occupancy for wide frequency ranges (say from 100 kHz to 10 GHz) given different time and location. The biggest problem is the lack of meaningful spectrum measurements from wide locations and long time durations (different days and times of the day) Therefore we plan to perform the following experiment.

Whole Netherlands is covered by a network of approx. 12 antennas of Agentschap Telekom (located in Heerhugowaard, Wijdmeren, Hoek van Holland, Groningen, 't Harde, Axel, Hoogeveen, Sittard, Breda, Eindhoven, Hengelo and Nijmegen) which are used to monitor whether spectrum licensees are properly exploited. We want to connect each of these measurement points by and collect measurement data for further analysis from Agentschap Telecom location to TU Delft and TNO. Given the massive amount of data that we need to transmit (we aim at Giga-samples speed for the spectrum frequency range we are interested in) we need huge bandwidth and lots of processing and storage power. It also requires to Agentschap Telekom become connected with SurfNet network.

## 6. Interdisciplinary and interinstitutional data collection for traffic simulation and travel behavior modeling

### Onderzoeksvoorstel

**Titel:** Interdisciplinary and interinstitutional data collection for traffic simulation and travel behavior modeling

**Beschrijving:** Traffic problems in the Randstad have grown to high levels given the difference in the increase of the number of cars and the available road surface: the overall volume/capacity ratios are leading to longer congested periods and longer queues, the traffic system is getting vulnerable. To study this effectively, more and new data, and fast and reliable data is needed to be able to build new generations of traffic models that are more than a global, average picture of travel behavior and traffic propagation in generalized networks.

Current traffic models are based on travel behavior models: trip generation, distribution and modal split, route choice and assignment. These behavior models comprise of mathematical functions and parameters derived from household surveys and other samples of mobility. The test of accuracy of the models is in general focusing on a fit on traffic counts. There is almost no reliable data available to test the accuracy of the individual behavior models, only the total outcome in terms of traffic volumes on major roads can be evaluated. Adjustments are made with correctional factors, which are in fact the unexplainable elements of the model, to deal with inaccuracies in the input data but also with the stochastic aspects in travel behavior. These adjustments will comprise of factors that changes the traffic volumes with up to 10% or more. To improve the models more data is needed, but data collection is already very expensive as it is, so any improvement using only traditional survey methods is unlikely. Not only the size of the samples, but also the accuracy and the types of data can be enhanced: travel times and distances are in general measured very roughly and route choice is in most cases disregarded because of the complexity of that information.

The traffic models are normally describing the average working day. For long term forecasts that is a logical thing to do, since the uncertainty in all the explanatory factors of the model is major. The social-economic basis is highly dependent on economic scenarios for the far future, scenarios that are trying to grasp developments that can change strongly. The spreading in these developments is assumed to lead to differences in traffic and transport in the range of about 20% over 20 years. This adds on to the basic technical uncertainty in traffic models due to inaccuracies and stochastic aspects in the behavior models. A sharper understanding and explanation of traffic and transport would be necessary, based on much more data, to reduce the uncertainty in forecasts and to gain substantial reliability to be able to assign effectively budgets to extreme costly infrastructural building projects.

On top of this day-to-day differences of traffic and transport are of at least the same magnitude. The working days are all quite different, even within the same week. They are even more different over the seasons of a year. Spreading of traffic volumes over the days of a work week and spreading over the seasons of traffic volumes on the same workday are again 10% or more. To capture these effects for effective traffic management, also to be able to deal with incidents and accidents, is again an extra data need. This not only makes it necessary to have more data, but also to have the data very fast, including consistency checks and evaluation against historical data.

New data sources that could overcome much of these problems are found through devices that are being taken along on trips. Mobile or cell phones and systems with GPS, e.g. navigation, provide floating car data that is filling much of the gap of lacking or inaccurate

data: exact timing and routes under changing circumstances will be available at large numbers and with considerable gain in speed. With the new kind of data models could be build that are more accurate, more detailed, more varying according to real life circumstances. Those models will be available for all traffic analysis, from operations and management to planning and impact evaluations of economic and urban development, of building infrastructure, of changing business cases and transport policy.

Our interdisciplinary approach addressing spatial and social interactions represents the frontier in traffic simulation and travel demand behavior modeling. The project develops state-of-the-art decision support for traffic flow of vehicles using dynamic data. This transportation engineering approach allows environmental impact assessment. At the same time, the project develops innovative activity-based decision support for travel demand behavior. This human behavioral geography approach allows socio-economic impact assessment.

- 1) traffic simulation on the basis of dynamic data collection - TU Delft and TU/e
- 2) environment simulation and social geography impacts - TU Delft and UvA
- 3) interactive workshops with stakeholders from municipalities, provinces and private sector, using results from (1) & (2) – ITC, TU Delft and UvA
- 4) interactive stated preference survey of travel behavior in laboratory setting, using results from (1) & (2) - TU/e and UvA

A special element is HD videoconferencing, building on experiences developed step-by-step in previous projects at the UvA. Namely, it would open up exciting possibilities for research, for education and for the interaction between both, if in parallel with videoconferencing there could be a dedicated channel for simultaneous data-sharing and high resolution interactive visualization of computer simulations at a distance. This could be of direct use for discussing geographical information system data and simulations, for traffic flows as well as environment and social geography impacts. Synergy with SARA for high performance computing aspects in an inter-institutional virtual circuit would also be important. At the same time however, such a development could also be potentially interesting for various other data-intensive fields of research whereby there is benefit from discussing real-time simulations face-to-face at a distance. As such the implication of the development cou

ld have a number of possible spillover and multiplier effects.

## 7. Gedistribueerd hoge kwaliteit audiovisueel testbed

### Onderzoeksvoorstel

Titel: Gedistribueerd hoge kwaliteit audiovisueel testbed

Beschrijving : Momenteel vindt wereldwijd onderzoek plaats naar de productie, transmissie en vertoning van audiovisueel materiaal van extreem hoge kwaliteit. Dit gebeurt in het kader van CineGrid. In Nederland bestaat CineGrid Amsterdam waarin twaalf partijen met deze technologie experimenteren.

Waag Society, UvA, NFTA, Thomson en SARA doen onderzoek naar het gebruik van netwerken voor het transporteren van audiovisuele data. Het gaat om beeldresoluties van 8M pixels en hoger; minimaal 4096 x 2160 pixels (4K). Dit levert grote bestanden op (5.5 TB voor twee uur materiaal zonder geluid) die zo in (bijna) real-time moeten worden getransporteerd. De datastroom is 7.6 Gbps voor ongecomprimeerde transmissie. Door state-of-the art compressie wordt de datastroom minder dan 1Gbps.

Lichtpaden maken het mogelijk om opname, opslag, processing en vertoning uit elkaar te halen. Zo kan optimaal gebruik worden gemaakt van dure hardware die zich op verschillende plekken bevindt. Dit stelt specifieke eisen aan het netwerk: hoge en gegarandeerde bandbreedte, kleine constante vertraging en een veilige verbinding. Dynamische lichtpaden zijn essentieel voor audiovisuele toepassingen. Productie, transmissie en vertoning duren enkele uren tot enkele dagen. Het is niet realistisch om voor kort gebruik fysieke verbindingen aan te leggen of te verleggen; een dynamisch lichtpad biedt uitkomst.

De System en Netwerk Engineering (SNE) groep bij de UvA zorgt voor de wetenschappelijk kwaliteit van het onderzoek. SNE heeft expertise in (optische) netwerkprotocollen en het ontwerpen van (optische) netwerk architecturen. Dr. Paola Grosso coördineert het onderzoek.

Waag Society, medialab en trekker van CineGrid Amsterdam, ontwikkelt een (semi-) permanent podium voor audiovisueel 4K materiaal in Pakhuis de Zwijger. In de Waag bevindt zich een HD studio. De UvA beschikt over 64 TB (Sun Thumpers) opslagcapaciteit waarmee een grote hoeveelheid data real-time kan worden ontvangen en verzonden. De NFTA ontwikkelt audiovisueel materiaal. Thomson Technicolor beschikt over "deep storage" voor langdurige opslag van audiovisueel materiaal. SARA bevindt zich op het knooppunt van de internationale glasvezelnetwerken waarmee lichtpaden kunnen worden uitgebreid. Daarnaast beschikt SARA over twee tiled displays en een 29-node OptIPuter cluster voor rendering.

Via Enlighten Your Research wordt de apparatuur en de mankracht op deze plekken verbonden. Daarmee ontstaat een uniek testbed waarmee onderzoek wordt gedaan.

Wetenschappelijke vragen die worden geadresseerd:

- ontwikkeling van geschikte zoekengine en metadata tbv. een portal
- capaciteiten voor real-time streaming en real-time ophalen van data uit opslag
- dynamische koppeling van sender nodes naar lichtpaden
- Quality of Service
- open codecs en transmissie

Maatschappelijke en economische vragen die worden geadresseerd:

- online samenwerken met & aan grote hoeveelheid data
- efficiënt benutten van dure gedistribueerde apparatuur
- nieuwe toepassingen van extreem hoge kwaliteit audiovisueel materiaal

Extra:

Naast de vijf genoemde punten zien we mogelijke uitbereidingen:

- Media Plaza in Utrecht

- SARA, locatie Almere
- SURFnet, Utrecht
- Via the NetherLight Exchange in Amsterdam:
  - o KEIO University, Tokyo
  - o UCSD, San Diego
  - o I2Cat, Barcelona
  - o Cesnet, Praag
  - o KTH, Stockholm
  - o PSNC in Poznan, Polen