EGI Case Studies

engage - grow - innovate

European Grid Infrastructure

The grid computing resources provided by the EGI federation are used by scientists and researchers across Europe and beyond. The case studies are examples of how grid computing is helping their work.

http://www.egi.eu

About EGI

The **European Grid Infrastructure (EGI)** delivers integrated computing services to European researchers, driving innovation and enabling new solutions to answer the big questions of tomorrow.

EGI is a federation of 39 National Grid Initiatives (NGIs) from European countries and European Intergovernmental Research Organisations.

What does EGI do?

- Speeds research processes which are data and compute intensive
- Makes it easy for global scientific communities to collaborate
- Simplifies how data and results are shared, discovered and re-used

The research featured in these pages would not have been possible without the computing resources and services made available by the **National Grid Initiatives (NGIs)** – the resource providers of EGI.

- List of NGIs per country: http://www.egi.eu/community/ngis/

The articles presented in this brochure are a selection of the case studies published on the EGI website during the **EGI-InSPIRE project** 2010-2014.

- EGI website: http://www.egi.eu
- EGI-InSPIRE project: http://www.egi.eu/about/egi-inspire/

All **EGI Case Studies** were prepared with the collaboration with the scientists involved in the work. Many thanks to all of them for their cooperation and enthusiasm.

- The EGI Case Studies: http://www.egi.eu/case-studies/

The federation is coordinated by **EGI.eu**, a not-for-profit foundation created to manage the infrastructure on behalf of its participants.

Troops deployed to the Kunar province, Afghanistan in September 2010. Photo by Staff Sgt. Gary A. Witte. Source: Wikimedia Commons

Does combat stress have a long-term effect on attention and memory?

HOW GRID COMPUTING IS HELPING TO UNDERSTAND THE EFFECTS OF COMBAT ON SOLDIERS' BRAINS.

Veterans returning from active duty endure many challenges to readapt to a civilian life. For some there is the added complication of combat stress, which affects their memory, attention span and other cognitive functions. But for how long?

Neuroscientist Guido van Wingen and colleagues at the AMC in Amsterdam monitored a group of soldiers from before their first deployment to Afghanistan until 18 months after their return to civilian life. The idea was to investigate how combat stress affects brain areas supporting cognitive functions such as memory and attention.

The team used a grid-enabled science gateway to process and analyse brain scans from soldiers and the civilians used as controls. Thanks to a friendly user interface and the computing power of the grid, a workload of several weeks was condensed into two days.

The conclusions, published in the journal PNAS, show that combat stress impairs cognition by affecting the midbrain and its link with the prefrontal cortex. The effect is largely reversible but it could have an impact on future social and cognitive functions.

MONITORING THE TROOPS

Severe stress in war zones can lead to long-lasting problems in combat troops. The diagnosis of 'shell shock' used during the First World War has come a long way and post-traumatic stress disorder is now recognised as a major social problem.

Guido van Wingen and his colleagues at the Academic Medical Centre (AMC) in Amsterdam wanted to know how stress affects the areas of the brain that support cognitive functions. To do this, they tested 33 soldiers before they were deployed to Afghanistan.

The volunteers' brains were scanned with functional magnetic resonance (fMRI) and diffusion tensor (DTI) imaging techniques. The soldiers also completed cognitive memory and attention tests and filled questionnaires to assess psychological condition. As control, the team also scanned and tested 26 individuals in non-combat roles.

While in Afghanistan the soldiers were embedded as part of a NATO mission and exposed to combat situations, enemy fire and the risk of improvised explosive devices. The tests were repeated shortly after their return from deployment and then again 18 months later.

In total, van Wingen and his colleagues had 118 scans to analyse, which would take approximately 240 hours of computing time. This is not a lot, but "the logistics to organise so many datasets and process them is complex when using a regular computer," he said.

To make things easier, van Wingen used the e-bioinfra gateway to access SURFsara's Dutch e-science grid and analyse the DTI scans with software developed by the AMC's e-Bioscience group led by Silvia Olabarriaga.

"Silvia's group has created an incredibly user friendly interface to perform the most difficult analyses with several mouse clicks to upload the data and run the analyses," van Wingen says. "And with the computing power of the grid the analyses that would otherwise require several weeks' time are finished over the weekend."

LONG AND SHORT-TERM CONSEQUENCES

From the psychological evaluation, the team found that combat stress interfered with the soldiers' capacity for sustained attention shortly after their return from Afghanistan. The fMRI scans of their brains during a memory test revealed that this effect was caused by functional changes in the midbrain. The analysis of the DTI images showed that these were accompanied by structural alterations as well.

In the long run the soldiers returned to their baseline value for attention, which means that the influence of stress on the midbrain's function and structure is reversible.

The team also found long-term changes in the way the midbrain couples with the prefrontal cortex. "Midbrain-prefrontal cortex coupling is an index for communication between these brain regions," explains van Wingen. "This is known to be important for maintaining information in the working memory, for its evaluation and manipulation in order to make appropriate decisions."

The study, published in the *Proceedings of the National Academy of Sciences*, shows that while the effects on attention are reversible, combat-stress compromises the link between the midbrain and the prefrontal cortex. This may increase the soldiers' vulnerability to stress and have an impact on future social and cognitive functions.

Reference:
G van Wingen et al. (2012) doi:10.1073/pnas.1206330109

"Silvia's group has created an incredibly user friendly interface to perform the most difficult analyses with several mouse clicks to upload the data and run the analyses."

The DTI data analysis software was developed at the AMC and integrated into the AMC e-bioinfra science gateway described in Shahand et al. 2012. The data was first uploaded by the user to an ftp site, and then transported automatically to the grid storage and back.

The computations were managed by the MOTEUR workflow management system. When the results were ready, they were transferred back to the ftp server automatically.

A robot certificate of the VLEMED VO was used to access the grid resources. Large computation and storage capacity available for this study allowed obtaining results faster and more easily.

Shahand et al. (2012) doi: 10.1007/s10723-012-9233-4.

SCIENTISTS USE GRID COMPUTING TO CORRELATE DATA FROM MILLIONS OF CALCULATIONS AND UNVEIL THE ROCK STRUCTURE OF AN OIL FIELD UNDER THE NORTH SEA.

Making sense of seismic noise

Oil and mining companies use seismic waves to figure out the structure and type of rocks underground, so they can plan their work. Because they can't wait for natural earthquakes, they artificially induce seismic waves with explosions. This is, however, expensive and has severe environmental consequences.

Aurélien Mordret, a seismologist based in France, is working on an alternative that puts the background seismic noise to good use. Thanks to a mathematical method called cross-correlation and grid computing, it is possible to collect meaningful seismic data from the rumble of waves as they travel across the seafloor.

When applied to the Valhall oil field in the North Sea, the method revealed a structure crisscrossed with old riverbeds and the tell-tale signs of a depleted oil reservoir.

> "...the technique is very powerful to make high resolution seismic images of the subsurface (...) But it requires a huge amount of computation and this is where EGI becomes helpful."

"Everything starts with the seismic noise," says Aurélien Mordret, who did his PhD in seismology at the Institut de Physique du Globe in Paris (IPGP), France.

Seismic noise is produced when waves roll through the ocean floor. And because most of the Earth is covered by oceans, this noise is heard and registered by seismometers everywhere, even in the middle of continents.

"This whisper of the Earth was considered mostly a nuisance by seismologists because sometimes it was so loud that it could hide the observed earthquake signals," Mordret explains.

But about ten years ago, scientists discovered that the seismic noise could be transformed into a useful signal with a simple mathematics trick called cross-correlation. In other words, "we could be able to retrieve some information about the structure of the Earth by cross-correlating the seismic noise recorded at two different seismometers."

As part of his research, Mordret is applying the method to the Valhall oil field in the North Sea to investigate the structure of the rocks underneath the ocean floor.

The oil field is monitored by a network of about 2500 sensors that measure seafloor motion in three spatial directions, for example Vertical (V), North (N) and East (E).

Mordret explains what happens next:

"With the three orientations and two stations, we can compute nine correlations: VV, VE, VN, NE, NN, NV, EE, EN, EV. With the 2500 seismometers, we can make 3,123,750 combinations of station pairs. Multiplied by nine times the number of orientation pairs, we obtain 28,113,750 of correlations to compute. A crazy amount of data! This is why I needed the EGI to compute all this."

READING A WAVE

Armed with the computational power of the grid, Mordret was able to produce 3D models of the seafloor beneath the network of sensors.

These models allowed him to get an idea of the way rocks are structured underground, because seismic waves travel at different speeds through different types of rock. At Valhall, Mordret was able to identify of the sediments deposited during the Quaternary and Tertiary geological periods crisscrossed by old river channels. These were highlighted by high seismic velocity areas.

"During the last glacial era, the North Sea was empty and at some points, big rivers crossed the area," he explains. "Their channels were filled by sand and gravels and later, the sea came back and covered the area with marine sediments which exhibits lower seismic velocities." The findings were published in the journal *Geophysics*.

At Valhall, the cross-correlation method is only sensitive to the first 500 meters, so the model does not reach the oil reservoir that is down at 2,500 meters. Even so, by computing the way the seismic waves propagated underground, Mordret was able to draw some conclusions about the reservoir. He reported on a cluster of cracks in the rock in a paper for the *Geophysical Research Letters*. "The cracks in the ground had a particular elliptical distribution," says Mordret. "This is due to the subsidence of the ground cause by the oil reservoir depletion at depth."

"To sum up, even if it will never totally replace the use of explosions, the seismic noise correlation technique is very powerful to make high resolution seismic images of the subsurface when it is coupled with very dense seismic network of several hundreds of stations," he concludes. "But it requires a huge amount of computation and this is where EGI becomes helpful."

The original input data amounted to 110 GBytes (compressed) and the output totalled 2.2 TBytes (uncompressed) of data. Grid storage was used as a temporary buffer. Data processing involved around 300 simultaneously running grid jobs for 3 months, distributed to more than 20 grid resource centres in five countries (France, Germany, Slovakia, the UK and the Netherlands).

The effort resulted in a dramatic reduction of the duration of the processing.

References:

A Mordret et al. (2013) doi:10.1190/geo2012-0303.1

A Mordret et al. (2013) doi:10.1002/grl.5044

An oil and gas exploration platform in the North Sea.
Image: JanChr / Wikimedia Commons.

3D liver maps using magnetic resonance imaging

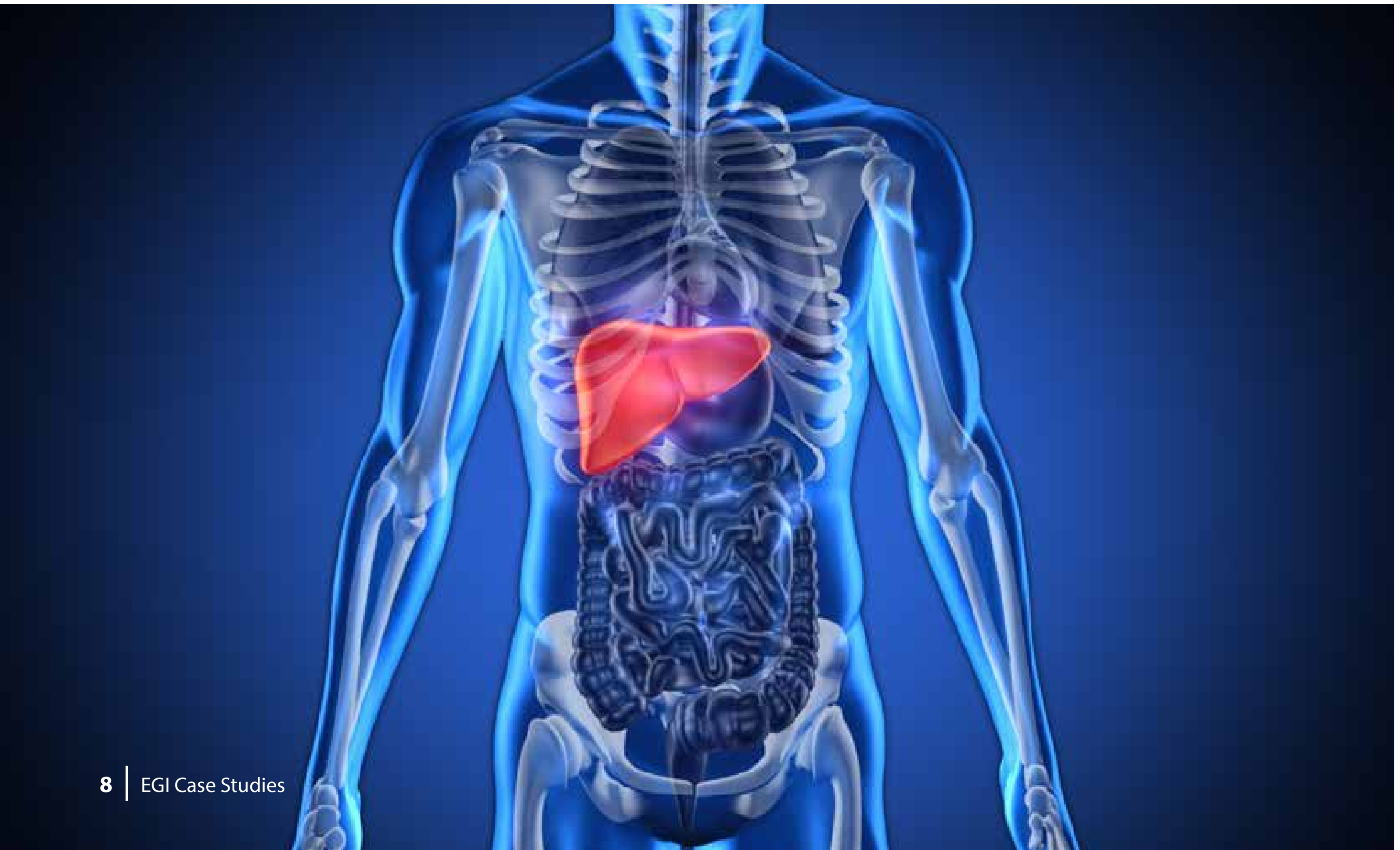
HOW GRID COMPUTING HELPS TO DEVELOP ALTERNATIVE, NON-INVASIVE TECHNIQUES TO DIAGNOSE AND MONITOR LIVER FIBROSIS.

Chronic liver disease is never good news but in about 20 per cent of the cases the condition escalates to severe inflammation and fibrosis, a kind of scarring of the liver's tissue.

When this happens, it is very important to follow the progression of the disease and monitor the onset of fibrosis to guarantee the best treatment. Liver fibrosis is usually diagnosed with a biopsy – an invasive surgical procedure to remove liver cells for lab testing.

Olivier Beuf believes that we can do better and has put together a team to develop an imaging technique that can be used in a clinical examination. Using grid computing, they developed a 3D map showing how fibrosis affects blood flow in a diseased liver.

The 3D maps will be a reliable alternative to the usual biopsy procedure and will contribute both to patient's well-being and to improve our understanding of liver disease.



AN ALTERNATIVE TO BIOPSIES?

Chronic liver disease is on the rise thanks to an increase in obesity, diabetes and metabolism disorders. "The most common cause of chronic liver disease in is non-alcoholic, fatty liver disease," says Olivier Beuf, a physicist working in the medical imaging field for over twenty years.

In about 20 per cent of the cases the disease progresses to aggressive forms, with inflammation and liver fibrosis leading to potentially fatal hypertension or even cancer.

To describe fibrosis, Beuf compares the liver to a filter. "When healthy, the filter is soft and the blood from the different vessels in the liver is flowing normally," he says. "When you have fibrosis, the filter is getting tough and starts to be blocked. Then pressure is increasing in the vessels and you get hypertension."

Current practice is to monitor fibrosis with biopsies – medical interventions where a physician removes a tiny sample of the liver for analysis. The trouble with biopsies is that they are an uncomfortable and invasive procedure with many risks for the patient.

"Biopsies are unsuitable for clinical monitoring in patients and problematic in children," explains Beuf. "Biopsies have other limitations such as cost, inter-observer variability and sampling errors."

So what could be the alternative?

LOOK, DON'T TOUCH

Beuf believes that 3D imaging of the liver is part of the solution. He teamed up with PhD student Benjamin Leporq, radiologist Frank Pilleul and a group of computer scientists to create a non-invasive imaging technique to replace biopsies as a method to diagnose and monitor liver fibrosis. The challenge was to develop an imaging method that can be used in everyday clinical examinations.

The team started by collecting a detailed 3D magnetic resonance images from the liver of six patients and one healthy person to serve as control. From these images, they extracted physiological information, which was adjusted to a mathematical model using nonlinear least-square fit to determine 3D liver maps.

"This procedure, performed on every pixel of the 3D volume is extremely long," Beuf says. With normal computing resources, each 3D map would require several weeks of processing.

So to speed things up, Beuf used EGI's grid computing services through the biomed virtual organisation. The results were, on average, 126 times faster than with conventional computing and allowed "a reliable estimation of 3D perfusion parameter maps of the whole liver in a reasonable processing time - hours compared to weeks," he says. "Our goal was unreachable without parallel or grid computing."

The results, published in the *Journal of Medical Engineering*, show that detailed 3D magnetic resonance mapping of blood flow parameters is achievable with grid computing today.

Reference:

B Leporq et al. (2013) doi:10.1155/2013/471682 [open source]

> [grid computing allows] a reliable estimation of 3D perfusion parameter maps of the whole liver in a reasonable processing time - hours compared to weeks. Our goal was unreachable without parallel or grid computing.

Olivier Beuf and his team used grid computing to perform the pixel-by-pixel analysis using a nonlinear least-square fit to the model. The processing step was executed by parallelised jobs on EGI within the biomed VO.

The whole processing operation was modelled and implemented as a grid workflow using the Gwendia language and the MOTEUR workflow engine. The interface with the grid resources was provided by the VIP web platform.

Tracking a biomarker for Alzheimer's disease

SCIENTISTS USE GRID COMPUTING TO TEST THE SOFTWARE NEEDED TO TRACK THE BIOMARKERS OF ALZHEIMER'S DISEASE.

Alzheimer's disease starts slowly and patients may not show symptoms for many years. Yet subtle physical changes will already be occurring, as their brain cells begin to die and the brain atrophies. The key to early diagnosis is to find a reliable 'biomarker' for Alzheimer's that researchers can use to monitor the disease and decide on treatments.

One useful clue is the volume of the hippocampus - the region in the brain associated with memory that starts to shrink at the onset of Alzheimer's. Many software programs have been developed to measure changes in hippocampus size from imaging scans. Now researchers at Vrije Universiteit Amsterdam have used grid computing to compare the performance of several software programs by analysing thousands of MRI scans taken from Alzheimer's patients.

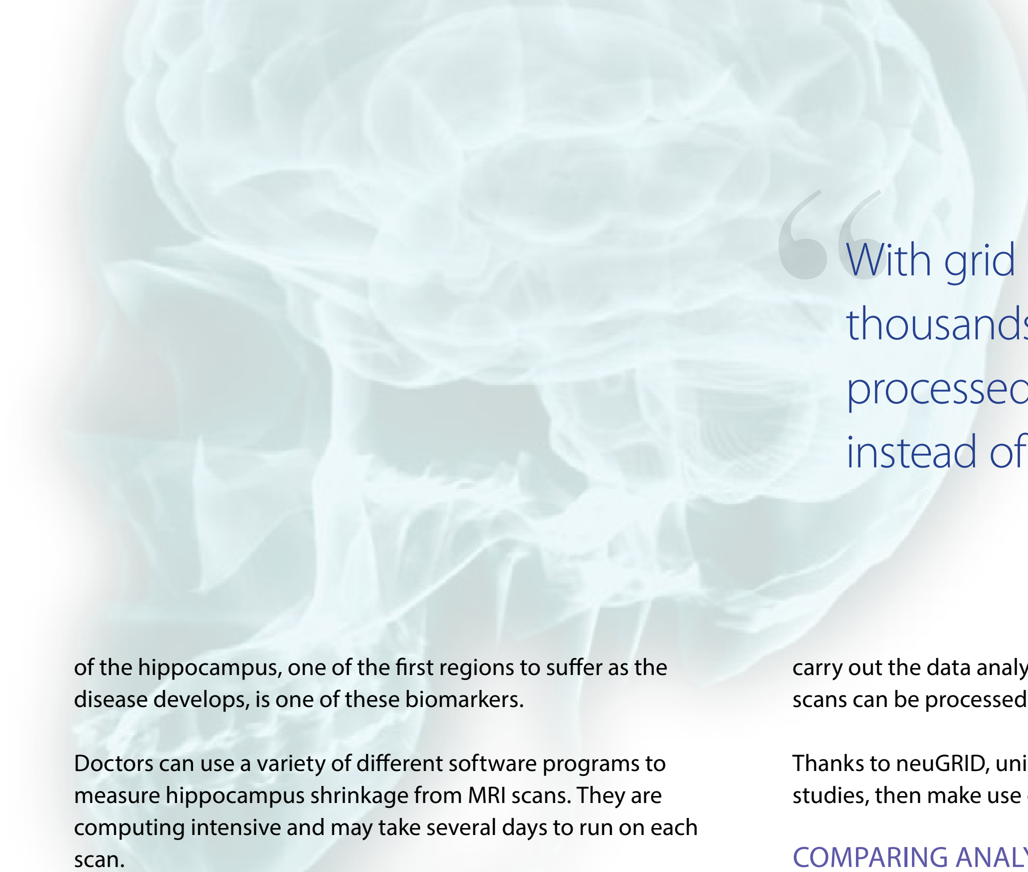
The study produced a valuable benchmark to evaluate Alzheimer's biomarkers. Better biomarkers from brain scan data open the door to earlier diagnosis, effective monitoring, and being able to quickly test new drugs for the disease.

DECIPHERING MRI DATA

Alzheimer's disease kills brain cells and causes the brain to atrophy and shrink. As the disease progresses, patients begin to experience memory loss and inability to carry out physical tasks.

The diagnosis is usually made on the basis of these symptoms and the challenge is to distinguish between the normal signs of aging and the beginning of Alzheimer's.

One way to tell normal aging from actual symptoms is to look for signs of Alzheimer's disease – the biomarkers. The volume of the hippocampus, one of the first regions to suffer as the disease develops, is one of these biomarkers.



Doctors can use a variety of different software programs to measure hippocampus shrinkage from MRI scans. They are computing intensive and may take several days to run on each scan.

Keith Cover, a physicist working at VU University Medical Centre (VUmc) in Amsterdam, tested the reliability of several software packages. This involved analysing approximately 3,300 MRI scans from over 600 patients many times each, requiring many core-years of computation.

The brain scans were collected during the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, which began in 2004. Each patient provided two back-to-back scans in every visit. This allowed the team to look at two sets of what should be identical data and check how reliable each software package is.

COMPUTING WITH NEUGRID

Software packages FSL and FreeSurfer are widely used in Alzheimer's research and consume a lot of computational power to piece together data from hundreds of patients. For this study, Cover and his team used the neuGRID - a leading European e-Infrastructure for the neuroscience research community.

Through the neuGRID platform, researchers can analyse and share brain scan datasets, use medical software tools and benefit from specialised support.

neuGRID was originally established in 2008. The EU-funded N4U project, led by Giovanni Frisoni, has expanded neuGRID through collaborations with several partners, including EGI, which have provided access to public grid computing resources via, for example, Hellasgrid, the NGI of Greece.

With neuGRID, the calculations for Cover's study were processed in weeks, instead of years.

"A modern trial could have 100 to 1,000 patients providing these scans," Cover explains. "Most laboratories have their own MRI scanners but few have their own clusters and the expertise to use them, so they don't have the computational power to carry out the data analysis. With grid computing, thousands of scans can be processed in a trivial time, days instead of years."

> "With grid computing, thousands of scans can be processed in a trivial time, days instead of years."

Thanks to neuGRID, university groups can carry out their own studies, then make use of grid resources to analyse the data.

COMPARING ANALYSES

The study was able to show that FSL and FreeSurfer analyses of hippocampal atrophy rates are similar, while noting slight differences in the current versions of the programs.

With the speed of analysis brought by grid computing, the hippocampus biomarker can be used more widely to help diagnose and monitor individual patients. The study has also demonstrated that the ADNI data provides a reliable benchmark that could be used to compare the performance of other similar software packages.

Cover is already looking towards the future: "Now that we've got this benchmark in place, we're now looking at different algorithms and hoping to see if someone has come up with something better."

Reference:
GB Frisoni et al. (2010) doi:10.1038/nrneurol.2009.215

neuGRID consists of extensive public and private resources. The grid has 6,000 CPU cores, and 20 terabyte (TB) of staging, with additional unlimited storage provided by EUDAT. It also has established connections with other e-Infrastructures such as LONI and CBRAIN.

Improving digital security with natural selection

HOW SCIENTISTS FROM THE CZECH REPUBLIC ARE TRIALLING NEW CRYPTANALYSIS METHODS WITH INSPIRATION FROM THE THEORY OF EVOLUTION AND THE HELP OF GRID COMPUTING.

Hiding private data is a major concern in today's digital world. We need to protect information from prying eyes and one way of doing that is by encryption with algorithms that scramble the data when it's sent and reassemble it after a safe arrival.

Good encryption algorithms make the processed information look as much as possible like random data, with few flaws so-called 'distinguishers'. Finding such flaws in algorithm design is essential for security. Yet computer programs can only look for as much as they are programmed to do. Human cryptanalysts are the best choice, but this is labour-intensive work. So how could a cryptanalysis program be made more like a human?

One tactic would be to give a programme the power to evolve.

Vashek Matyas and colleagues colleagues Petr Svenda and Martin Ukrop from Masaryk University have now trialled a form of cryptanalysis that uses genetic programming and the grid computing resources provided by Metacentrum, the Czech National Grid Initiative.

HOW DOES EVOLUTION COME INTO COMPUTER SCIENCE?

The 'genetic programming' approach borrows concepts from evolutionary theory and turns 'survival of the fittest' into a method to test algorithms.

The method uses programmed, hardware-like circuits as tests. These circuits take the encrypted information as input and test it. If a circuit succeeds in identifying a possible distinguisher, the circuit is selected and used again in the next round of testing. In between rounds, the evolutionary algorithms managing the process come in and tweak the circuits and their connections to form slightly different programmes.

And over many rounds the end result is a collection of 'surviving' circuits that is very efficient at finding distinguishers. In effect, their system keeps rewriting itself.

The team tested their approach on seven ciphers from the ESTREAM competition, organised by ECRYPT, and compared results to a standard cryptanalysis computer program, the NIST statistical testing suite. Overall they used 1,000 cores and consumed 30,000 CPU hours through the grid computing services provided by Metacentrum.

The system was able to find distinguishers in the ciphers, and performed almost as well as the NIST suite. The evolutionary algorithm was more efficient, once the learning evolution phase is over. Once a successful system is made, the algorithm is able to work on extremely short sequences, only 16 bytes long. This is a huge advantage over traditional statistical cryptanalysis programmes which need at least several megabytes of data.

And what was the benefit of grid computing for this work? "Grid computing allows us to execute experiment significantly faster, which in turn allows us to stop non-perspective runs, change controlling parameters of genetic programming and run it again," said Matyas.

> "Grid computing allows us to execute experiment significantly faster, which in turn allows us to stop non-perspective runs, change controlling parameters of genetic programming and run it again."

Reference:
P Svenda et al. (2014) doi:10.1007/978-3-662-44788-8 17

The team ran 50 jobs to evaluate 24 different cryptographic algorithms. Each job was assigned to 20 cores and ran for 24 hours. The total CPU consumption for the project was around 30,000 hours equivalent to almost 3.5 years of computing time.

Each job was represented as single virtual machine with MS Windows Server 2008, with a BOINC client responsible for distribution of actual computational tasks inside. For the execution of hardware-like circuits, the team used a software called EACirc bundled with GaLib library for genetic programming.

The team acknowledges Metacentrum's user support: "Because we use BOINC architecture as an upper layer, we need to have network connectivity from every virtual machine to our BOINC server positioned outside Metacentrum. The guys there helped us a lot a very quickly setup virtual network," said Petr Svenda.

On the double:
metabolic rates accelerate evolution

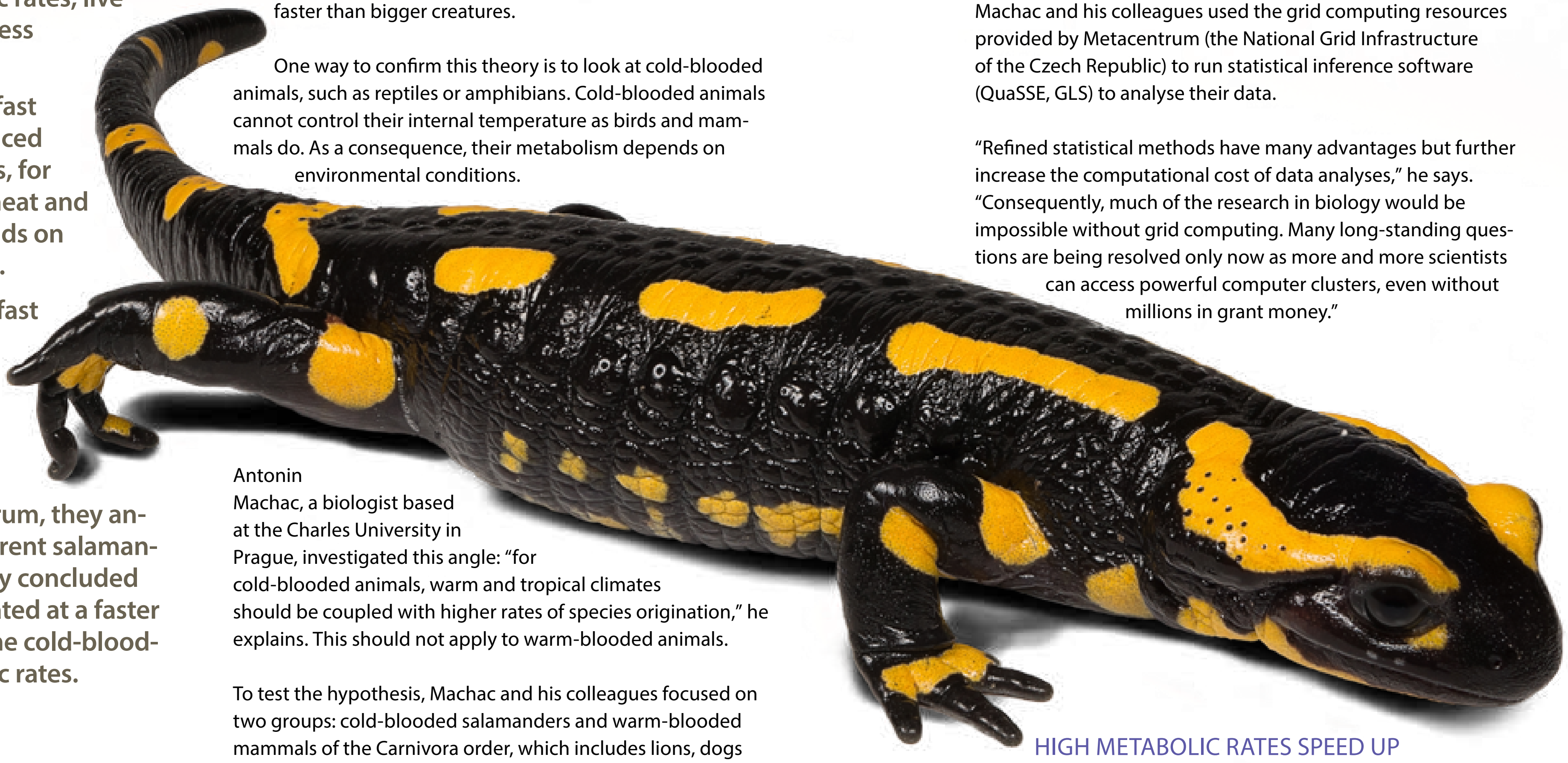
HOW GRID COMPUTING HELPS TO SHOW HOW NEW SPECIES OF COLD-BLOODED ANIMALS APPEAR FASTER IN WARMER CLIMATES.

Some animals live their lives faster than others. These differences are captured by a measure known as metabolic rate. Mice and small birds have high metabolism and live highly active but short lives. Whales and elephants, on the other hand, have lower metabolic rates, live longer and slower, and produce less offspring.

An animal's metabolic rate, how fast and long they live, can be influenced by climate. Cold-blooded animals, for example, cannot regulate body heat and have a metabolic rate that depends on their environment's temperature.

But does this also influence how fast new species appear?

Antonin Machac and colleagues from the Czech Republic, looked into the history of salamander evolution. Using grid computing resources provided by Metacentrum, they analysed data related to many different salamander species across the world. They concluded that new species were being created at a faster rate in tropical climates, where the cold-blooded animals have higher metabolic rates.



> "They [Metacentrum] organise a number of seminars which give researchers all the necessary information how to submit and run their jobs."

BLOOD, TEMPERATURE AND THE EVOLUTION OF NEW SPECIES

Biologists have long thought that fast metabolism accelerates mutations at the DNA level. More mutations lead to more opportunities for new species to evolve. In practice this means that species of animals with high metabolism should evolve faster than bigger creatures.

One way to confirm this theory is to look at cold-blooded animals, such as reptiles or amphibians. Cold-blooded animals cannot control their internal temperature as birds and mammals do. As a consequence, their metabolism depends on environmental conditions.

Antonin Machac, a biologist based at the Charles University in Prague, investigated this angle: "for cold-blooded animals, warm and tropical climates should be coupled with higher rates of species origination," he explains. This should not apply to warm-blooded animals.

To test the hypothesis, Machac and his colleagues focused on two groups: cold-blooded salamanders and warm-blooded mammals of the Carnivora order, which includes lions, dogs and bears, for example.

"Salamanders and carnivorans are great model systems because both groups are very well studied, we know how individual species within either group are related and where they occur geographically," Machac says.

First, the team collected data on evolutionary relationships, geographic distributions, and preferred environmental temperature for several hundred salamanders and carnivorans.

With this information, they estimated environmental temperate in which ancestors of current species presumably lived. And finally, statistical inference revealed if salamanders from warm climates generate more species over time than other groups.

THE HELP OF GRID COMPUTING

Machac and his colleagues used the grid computing resources provided by Metacentrum (the National Grid Infrastructure of the Czech Republic) to run statistical inference software (QuaSSE, GLS) to analyse their data.

"Refined statistical methods have many advantages but further increase the computational cost of data analyses," he says. "Consequently, much of the research in biology would be impossible without grid computing. Many long-standing questions are being resolved only now as more and more scientists can access powerful computer clusters, even without millions in grant money."

> "Refined statistical methods have many advantages but further increase the computational cost of data analyses. Consequently, much of the research in biology would be impossible without grid computing. Many long-standing questions are being resolved only now as more and more scientists can access powerful computer clusters, even without millions in grant money."

HIGH METABOLIC RATES SPEED UP EVOLUTION IN WARMER CLIMATES

The results, published in the *Journal of Evolutionary Biology*, show that new salamander species appear more often in tropical climates. This means that the higher metabolic rate of tropical salamanders is probably responsible for accelerated speciation rates.

"Even more striking," Machac says, "was that carnivorans showed the exact opposite pattern with higher speciation rates at higher, colder latitudes." This is because in warm-blooded animals, metabolic rates are higher in cold climates, where the body has to spend more energy to keep internal temperature stable.

Taken together, the conclusions suggest that temperature has a strong effect on how fast new species appear – but this effect will be different for different types of animal. New species of cold-blooded animals appear more often in tropical climates, whereas warm-blooded animals evolve faster in the cold. This might prove as an important insight into how the diversity of animal life originated.

Reference:
A Machac et al. (2012) doi:10.1111/jeb.12019

Image: Didier Descouens / Wikimedia Commons

Designing better antibiotics

HOW GRID COMPUTING IS HELPING TO DEVELOP ANTIBIOTICS WITH LESS SIDE EFFECTS BUT EQUALLY POWERFUL AGAINST FUNGI.

Diseases caused by fungi are a real risk for people burdened with weak immune systems, for example after organ transplants or long chemotherapy treatments. Fungi can do all kinds of damage, from causing pneumonia in the lungs to attacking the brain with vicious types of meningitis, or triggering life-threatening infections.

The antibiotic Amphotericin B – abbreviated as AmB – has been the drug of choice to fight fungal infections for the past 50 years. It's brutally efficient, killing a broad spectrum of fungal agents and active against all known multidrug resistant strains.

The catch is that AmB is toxic to the human body and it can cause organ damage in patients, especially in the kidneys.

Anna Neumann, based at the University of Technology in Gdansk, Poland, has been working on this problem and used grid computing to open the way for an upgraded version of AmB, with all the efficiency of the original but fewer side effects.

"We know how AmB works on a cellular level – that it acts on a cell membrane, and forms some kind of permeable structures, most probably channels, in it," explains Neumann. The channels built by the AmB allows the cell contents to leak out eventually leading to cell death.

The problem is that AmB is not very discerning and attacks the human cells together with the infectious fungi cells it's supposed to kill. This is because there is not much difference between the way this antibiotic interacts with fungal and human cell membranes.

The keys to solving this problem are the compounds known as sterols – the gatekeepers of cell membranes. Sterols control the physical and chemical properties of the membrane, for example how permeable they are. The sterol in fungal cells is called ergosterol; mammals have a different type, called cholesterol.

AmB has a slight preference to attach itself to membranes containing ergosterol (and hence kill the fungal cells), but this affinity is not strong and it explains why the antibiotic also attacks human cells: it sometimes can't tell the difference between them.

So, learning more about how AmB connects to the two types of cells and their sterols, how the antibiotic enters the cell membrane and how the channels are formed, will help create safer AmB varieties. Anna analysed the problem with molecular dynamics simulations – computer models designed to mimic the physical movements of atoms and molecules.

According to the results published in the *Journal of the American Chemical Society*, the difference in AmB's affinity for ergosterols and cholesterols is partially due to energy levels. It's easier, in terms of energy, for AmB to interact with the rigid and elongated molecular geometry of the ergosterol than with the cholesterol. In other words, AmB needs more energy to combine with human cells than with fungal cells and it is usually the lower energy option that wins out.

These conclusions, together with further analysis, will allow Anna to propose a way to make the AmB molecule more likely to attach itself to fungal cells. "That would affect AmB's activity – making it more selective for fungal cells and hence less toxic," Neumann concludes.

Reference:
A Neumann et al. (2010) doi:10.1021/ja1074344

Molecular dynamics models are very useful for describing the behaviour of atoms and molecules and their interactions, but are also very demanding in terms of computing power. For her research, Anna Neumann accessed the computing resources provided by the Polish National Grid Initiative to process the molecular dynamic simulations. She used 24 computing cores for each grid job that was submitted, adding up to a total of five million CPU hours.

Are comets born in asteroid collisions?

The tell-tail of the P/2006 VW139 main-belt comet was discovered by the Pan-STARRS1 telescope. This picture is a long-exposure image taken by the New Technology Telescope at La Silla Observatory in Chile. (Credit: Sam Duddy and Stephen Lowry, University of Kent)

HOW GRID COMPUTING HELPS ASTRONOMERS TO EXPLAIN THE ORIGINS OF MAIN-BELT COMETS.

P/2006 VW139 was just another asteroid orbiting the Sun between Mars and Jupiter until 2011, when astronomers from Hawaii spotted its comet-like tail. This was unexpected: the orbit of main-belt asteroids keeps them closer to the Sun at all times, which means that any ice they might originally have had to form a tail should have evaporated a long time ago.

So how do you explain the tail of these so-called main-belt comets? One theory suggests that they could be the remnants of a larger body smashed to pieces by a recent impact. And the way to prove it is to look for examples of an asteroid family that shares their origin with a main-belt comet.

Serbian astronomer Bojan Novakovic, from the University of Belgrade, worked on this problem with colleagues Henry Hsieh and Alberto Cellino. Their goal was to see if it could be assigned to a family of asteroids.

The search started in the AstDys database, a catalogue of 398,841 asteroids, using the hierarchical clustering method and the cut-off distance parameter to sort through the candidates. This narrowed down the possibilities to 24 asteroids.

To investigate the orbital characteristics of potential family members, Novakovic needed to go through a large number of calculations, which were computationally expensive and would take months to be done on a typical PC. Using resources provided by AEGIS, the Serbian NGI, the team was able to identify the asteroid family related to P/2006 VW139 and prove that they are all the products of a recent impact.

The conclusions suggest that 7.5 million years ago, a large asteroid crashed with another object and disintegrated into smaller pieces: P/2006 VW139 and its family members. P/2006 VW139 is special because it inherited the deeply buried ice of the parent asteroid. It's this ice that vaporises producing the characteristic tail of a comet.

Reference:
B. Novakovic et al. (2012) doi: 10.1111/j.1365-2966.2012.21329.x

> "The grid computing provided by PL-Grid was "essential to complete the project in a reasonable time," said Ewa Łokas"

Is the Andromeda II galaxy the result of a merger?

HOW SCIENTISTS FROM POLAND USED GRID COMPUTING TO FIND AN EXPLANATION FOR THE ORIGINS OF A GALAXY THAT FITS ASTRONOMICAL OBSERVATIONS.

Galaxies are massive systems of millions of stars and their planets, stellar dust and dark matter bound together by gravity. Scientists have a good understanding about how galaxies evolve and there are several accepted theories to explain how they are formed, based on information collected by telescopes.

But sometimes astronomical observations can be puzzling and the origins of Andromeda II, a dwarf spheroidal galaxy orbiting the massive Andromeda galaxy, remained a mystery.

The most common explanation for the origin of dwarf spheroidal galaxies suggests that they evolve from the accretion of earlier progenitors by a larger galaxy. This model does not fit the observations for Andromeda II, especially its strong rotation signal.

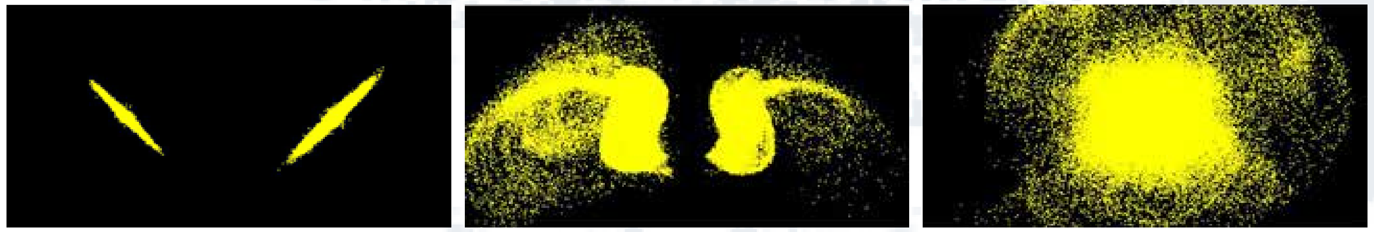
An international team led by Ewa Łokas from the Nicolaus Copernicus Astronomical Centre in Poland worked to find a way to explain the origins of Andromeda II. Using the grid computing resources provided by PL-Grid, who represents Poland in EGI, the team simulated how a merger between two galaxies would take place considering parameters such as the relative velocity of the merging dwarfs, their masses, the structure of their disks and dark matter haloes.

Łokas and her colleagues used grid computing to run about 20 simulations, each taking on average 400 CPU hours (the equivalent of about 16 days) to complete. The grid computing provided by PL-Grid was essential to complete the project in a reasonable time.

They found that the simulations that best reproduced the observed data for Andromeda II are the ones that considered a merger between two disk-shaped galaxies.

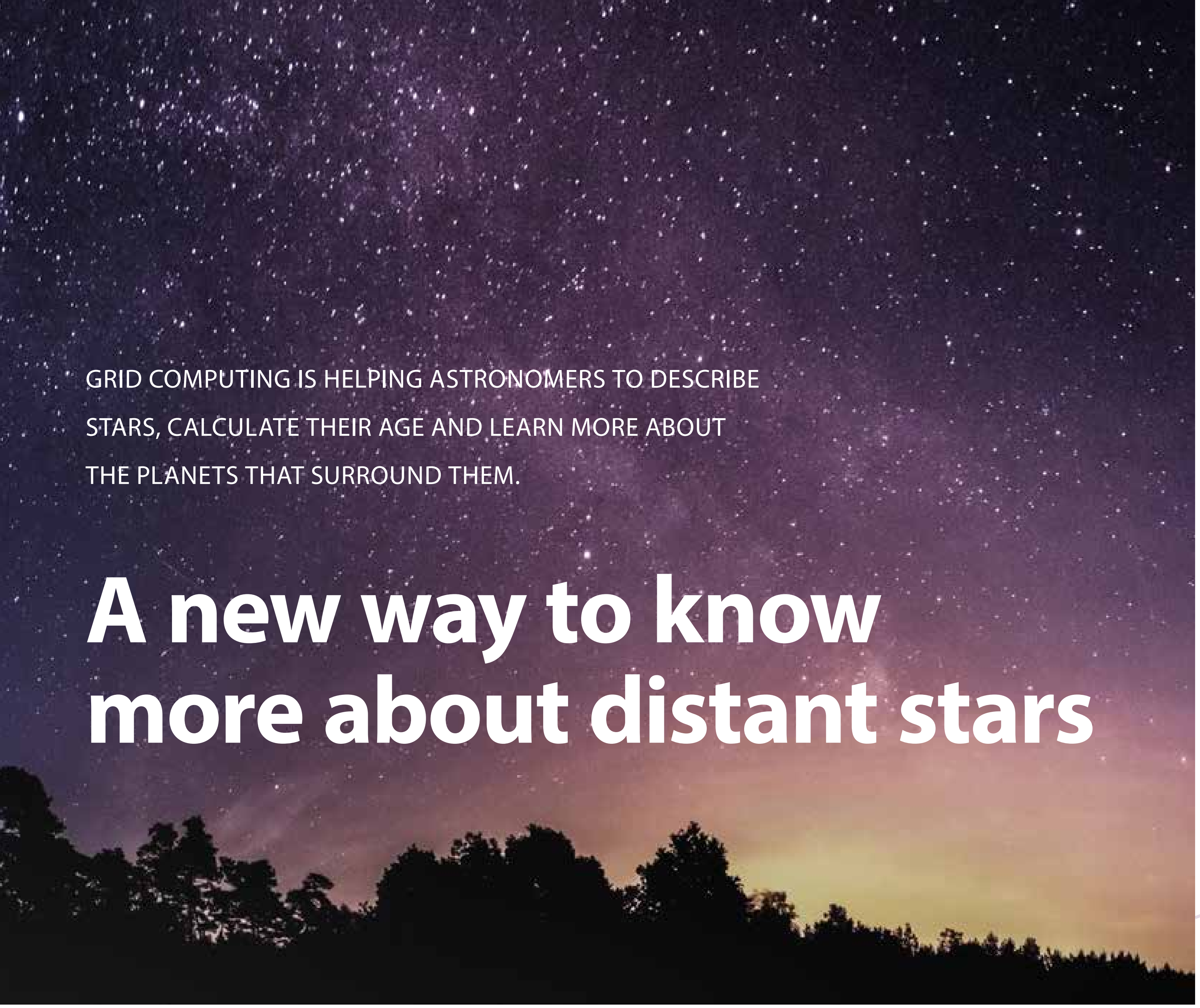
Reference:
EL Lokas et al. (2014) doi: 10.1093/mnrasl/slu128

The three stages of the evolution of the Andromeda II galaxy: the two disky dwarf galaxies (only their stellar components) before the merger, after the first fly-by and at the end (final merged product). Image: Ewa Łokas



GRID COMPUTING IS HELPING ASTRONOMERS TO DESCRIBE STARS, CALCULATE THEIR AGE AND LEARN MORE ABOUT THE PLANETS THAT SURROUND THEM.

A new way to know more about distant stars



Starry sky near Brandenburg an der Havel (Germany), close to midnight.
Image: Mathias Krumbholz / Wikimedia Commons

If there is something all cultures are and have been interested about is the Universe. How big is it? What is its origin? How are stars born and where do planets come from?

Finding the answers can be difficult from the limited observation point that is Earth. Scientists need to find clever ways of interpreting the information gathered by telescopes and asteroseismology is one of these tricks. With asteroseismology, astronomers can look into how acoustic and gravity waves propagate inside stars to have an idea of how big, or how hot the star is.

As part of his PhD at the University of Granada, Spain, astronomer Antonio García Hernández looked at two Delta Scuti stars observed by the CoRoT satellite. Delta Scuti stars can be three times the size of the Sun and rotate so fast that they have a flattened shape, instead of being perfectly spherical. An example is Altair, part of the Aquila constellation and one of the stars in the Summer Triangle.

García Hernández analysed the oscillations in frequency observed by the satellite on the Delta Scuti stars. He discovered a pattern in the frequency spectra of both stars than could be compared with the models and tried to see if this pattern could give us some information about the star.

Using grid computing resources provided by IBERGRID, García Hernández and his colleagues found that the frequency oscillations in the Delta Scuti stars can be used to predict the star's density. This will allow scientists to describe the internal structure and the age of stars, and will also be useful to learn more about the planets that surround them. The added bonus of the method is that it also allows for an accurate estimation of the age of the planetary system.

Reference:
JC Suaréz et al. (2014) doi:10.1051/0004-6361/201322270

HOW GRID COMPUTING HELPED TO DISCOVER TWO NEW PULSARS

Radio telescope LOFAR finds its first pulsars

The International LOFAR Telescope is a radio telescope centred in the Netherlands and spread across Europe. The telescope consists of a network of thousands of individual dipole antennas, connected over a fast network to a central supercomputer.

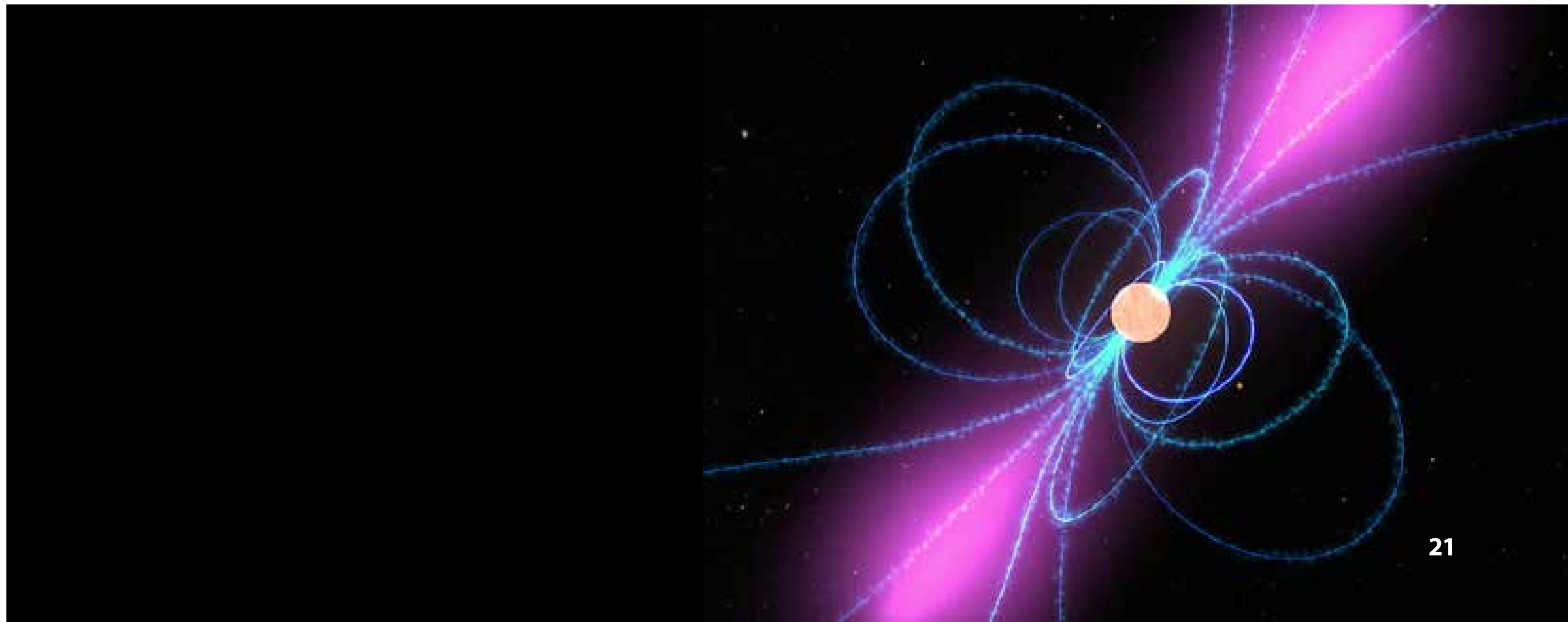
Pulsars act as cosmic lighthouses, emitting radio beams that sweep the galaxy. Their signals allow scientists to study the behaviour of gravity and matter in circumstances so extreme that they cannot be reproduced on Earth, not even in the most advanced facility. Pulsars are important because of this: they are the ultimate cosmic laboratories. So far, about 2,000 pulsars have been identified, but astronomers think there must be about 50,000 active pulsars in our galaxy.

Thijs Coenen did his PhD research as part of the LOFAR Pulsar Working Group based at the University of Amsterdam. Their work focused on discovering nearby weak pulsars using data from the LOFAR Pilot Pulsar Survey (LPPS) and the LOFAR Tied-Array Survey (LOTAS).

Processing the data from pulsar surveys proved to be a large computational task. The two pilot surveys consisted of several thousands of observations of the sky that all needed to be searched for periodic signals and single bright pulses. Using the Dutch resources operated by SURFsara, which is part of EGI, Coenen was able to search the data quickly and efficiently. In total, the team needed only a month to search through a set of 2010-2013 LOFAR images that would have occupied a single computer for more than a century.

As part of his PhD defence, Coenen was able to announce the first new pulsars discovered by LOFAR. The results validated the approach taken for the LOFAR pulsar surveys and offers promise for the full-strength pulsar survey currently underway.

Clouds of charged particles move along the pulsar's magnetic field lines (blue) and create a lighthouse-like beam of gamma rays (purple). (Image: NASA via Wikimedia Commons)



Where do African fish parasites come from?

SCIENTISTS USE GRID COMPUTING TO TRACE THE TAPEWORMS INFECTING NORTHERN AFRICAN FISH BACK TO EUROPE.

When tapeworms of the species Ligula intestinalis infect fish, the consequences are gory. The parasites grow to fill the host's body cavity, leading to behavioural changes and high mortality rates. It's unpleasant for the fish and a problem for the fish farming industry, and this is why scientists keep an eye on where the parasites live and how their geographical distribution evolves.

The tapeworms infecting North African freshwater fish are a good example of cryptic diversity: detailed genetic analysis reveals two hidden lineages within the species, one local and one invasive. The distinction is not just an academic problem, different lineages are likely to infect different types of fish. So where do the invasive parasites come from?

Jan Štefka, a biologist based in the Czech Republic, and his international team of colleagues suspected that the invasive parasites may have been introduced from Europe at this time, through already infected fish.

To test the theory, the team needed to understand the genetic variability between European and North African parasites, both between and within populations. So they collected a large, diverse dataset of the genetics of the fish parasites found in both Europe and North Africa. In the end they had hundreds of genetic samples of parasites from both sides of the Mediterranean.

The team used the grid computing services provided by Metacentrum (the Czech National Grid Infrastructure) to analyse the dataset.

The results, published in the journal *Biological Invasions*, show that the tapeworms arrived in Africa from Europe, hitching a ride with the fish introduced to improve the local lakes. Knowing where the invasive species came from, and how it got established in a new area, is vital for mitigating potential problems.

Reference:

W Bouzid et al. (2013) doi: 10.1007/s10530-013-0418-y

"Had we run the analyses on a desktop, we would probably still be running them now," said Jan Štefka.

How current flows in organic semiconductors

GRID COMPUTING IS HELPING MATERIAL SCIENTISTS TO UNDERSTAND HOW ELECTRIC CHARGES ARE TRANSMITTED THROUGH ORGANIC MATERIALS.

Organic semiconductors (such as pentacene or naphthalene) hold great promise for innovation as an alternative to silicon, the dominant material in conventional electronics. First, they are easy and cheap to produce. Second, organic semiconductors have interesting mechanical properties (they can be bent without damage, for example), while retaining their electrical characteristics.

The practical applications of organic semiconductors can be found everywhere, in solar cells, transistors or light-emitting diodes (LEDs). But the electronic properties of these devices are still lagging behind conventional materials.

Nenad Vukmirović, a material scientist based at the Institute of Physics Belgrade in Serbia, studies how electrical current flows in organic materials. His goal is to understand the relationship between current and crystalline structure and see how they can be improved.

His recent work, published in the *Physical Review Letters,* used grid computing to calculate how electrons (carrying the electrical current) interact with phonon waves (formed by oscillations within the material).

Like any other wave, phonons are characterised by their wavelength and it is possible to calculate the interaction of electrons with the different wavelengths separately. This is where the grid computing resources provided by AEGIS, the Serbian NGI, came in. The calculations were performed on separate processors for each wavelength and then assembled together to get the final result. The main advantage offered by grid computing was time-saving: using conventional computing methods, it would take too much time to finish the project.

Vukmirović and his colleagues concluded that the electron-phonon interaction in naphthalene is not strong enough to form polarons – a conceptual particle used to describe both the electron and the effect it causes in crystalline matter. This is further evidence of how efficient naphthalene can be as semiconductor.

Reference:

N Vukmirović et al. (2012) doi:10.1103/PhysRevLett.109.126407

Vukmirović submitted 100 jobs to EGI, each using 250 CPUs and lasting about 6 hours. User support was provided by AEGIS, the Serbian National Grid Infrastructure.

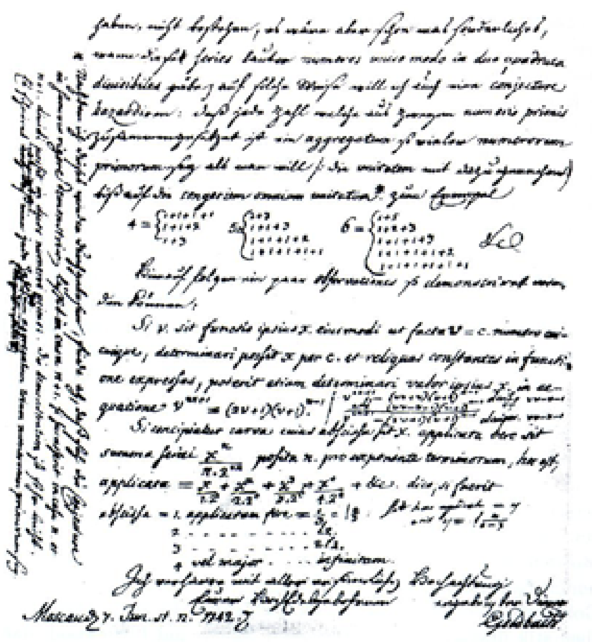
HOW GRID COMPUTING IS HELPING MATHEMATICIANS TO TACKLE A 200-YEAR OLD PUZZLE.

Cracking Goldbach's Conjecture

Goldbach's Conjecture says that every even number larger than four is the sum of two odd prime numbers, for example 60 = 43+17.

It sounds easy enough to understand – unfortunately it's not easy to prove and despite attempts since the 1800s and a 1 million dollar reward that went unclaimed in the 1990s, a definite proof is still elusive. But even though scientists can't prove it, they have been busy verifying if it holds for every case tested, or if there is a counter-example.

In a nutshell, the verification of a conjecture means testing if it holds for every case tested, or if there is a counter-example. For Goldbach's Conjecture, verification means checking even numbers to see if they are indeed the sum of two odd primes.



An excerpt of the letter where Goldbach enunciates his conjecture to Euler.

> "For this project, Silvio Pardi submitted 173,816 jobs to EGI adding up to 869,080 tasks, consuming 1 CPU-hour each. This means that in a normal laptop, Silvio would have needed about 99 years to do the same thing he did in 7 months."

It's not as definite as a proof, because numbers are infinite, but it can help mathematicians to understand the problem better. And the more verified numbers, the better.

Completing the task by hand is a fantasy, with a laptop it would take generations of scientists and even with supercomputers we are looking to a timescale of years.

It was only when a team of Portuguese and Italian enthusiasts embraced the power of grid computing provided by the Italian Grid that the verification process gained momentum. To put it into context, with access to HPC facilities the team was able to check an average of 1×10^{14} numbers per day. With grid computing, this could be completed in minutes. The verification limit of 4×10^{18} was reached on April 4, 2012, well ahead of the initial schedule.

HOW SCIENTISTS IN SERBIA ARE CONTRIBUTING TO OUR UNDERSTANDING OF BOSE-EINSTEIN CONDENSATES AND HOW THIS MAY LEAD TO A NEW COMPUTING PARADIGM.

Getting closer to quantum computing with the grid

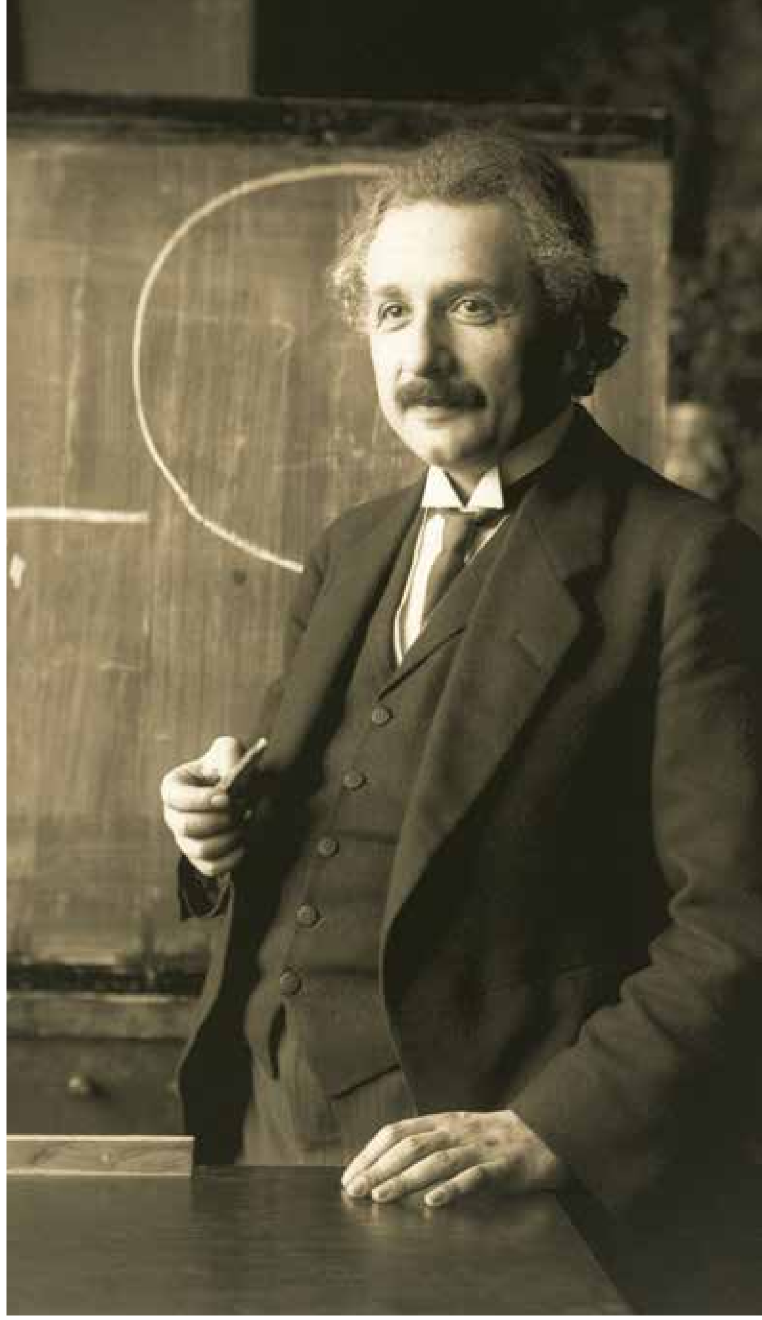
In 1924 Albert Einstein and his colleague Satyendra Nath Bose created a theory of what happens when matter is cooled to nearly absolute zero. At this low temperature, they said, atoms would gather together to form almost one single entity, an ultracold quantum gas now called a Bose-Einstein Condensate (BEC). We had to wait until 1995 to see the first BEC created in a lab.

Ultracold quantum gases hold huge promise because they have properties unlike any type of matter. Their potential applications are many and include interferometry-based gyroscopes, nanolitography, and an entirely new computing paradigm: quantum computers.

Antun Balaz and his colleagues based at the Institute of Physics Belgrade in Serbia, have been looking more into BECs and other ultracold quantum gases work as physical systems.

Numerical simulations of 2D disk-shaped BECs or 3D setups require a substantial amount of computing power to process the model and analyse the results. Serbia's National Grid Initiative, AEGIS, provides the muscle to perform massive numerical simulations of BEC systems.

From these simulations, Antun and his colleagues can find numerical values for several key physical properties of BEC systems. The numerical values can in turn be used to create an analytical method that produces results which can be compared to the original figures. The process is repeated many times until they obtain an efficient numerical simulation.



Albert Einstein 1921 by F Schmutzer / Wikimedia Commons

> "Since 2010 the team of two senior researchers and five PhD students has consumed more than 10 million hours of CPU time to publish 12 papers in international journals."

Contents

Does combat stress have a long-term effect on attention and memory? 4

Making sense of seismic noise 6

3D liver maps using magnetic resonance imaging 8

Tracking a biomarker for Alzheimer's disease 10

Improving digital security with natural selection 12

On the double: metabolic rates accelerate evolution 14

Designing better antibiotics 16

Are comets born in asteroid collisions? 18

Is the Andromeda II galaxy the result of a merger? 19

A new way to know more about distant stars 20

Radio telescope LOFAR finds its first pulsars 21

Where do African fish parasites come from? 22

How current flows in organic semiconductors 23

Cracking Goldbach's Conjecture 24

Getting closer to quantum computing with the grid 25

Colophon

Editing: Sara Coelho (EGI.eu)

Design: 5UP, www.5up.nl

The EGI-InSPIRE project (Integrated Sustainable Pan-European Infrastructure for Researchers in Europe) started on 1 May 2010, co-funded by the European Commission (contract number: RI-261323) for four years, as a collaborative effort involving more than 50 institutions in over 40 countries. Its mission is to establish a sustainable European Grid Infrastructure (EGI).

http://www.egi.eu