e science TALK

e-ScienceBriefings

Talking about e-science

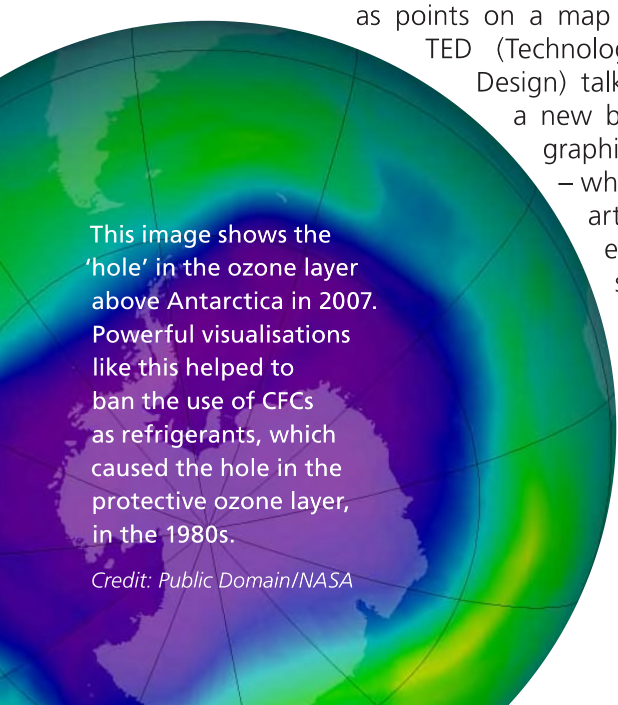
Visualisation

Information graphics, graphical information

The open data revolution, driven by a growing number of conscientious researchers and enlightened academic publishers, is making more data available to scientists and the public at large than ever before. Alone and without context, this mass of data can be a daunting deluge of numbers. Visualisation helps us not only to understand the conclusions of research, but transmits ideas across disciplines and cultural boundaries, creating a collaborative infrastructure that actually improves the quality of science being done. Visualisation may also help drive home the social and economic consequences of research, for example in understanding global environmental change.

Visualising data allows us to understand systems on a wide range of scales, from the global to the local. MAPPER – Multiscale Applications on European E-infrastructures – is a framework allowing scientists to seamlessly integrate simulations of natural phenomena, where different factors are important at different scales. "It's important that scientists should be able to zoom in and out of datasets in a coherent way," says Alfons Hoekstra, Director of MAPPER. "There are some important questions that can be explored through scientific visualisation."

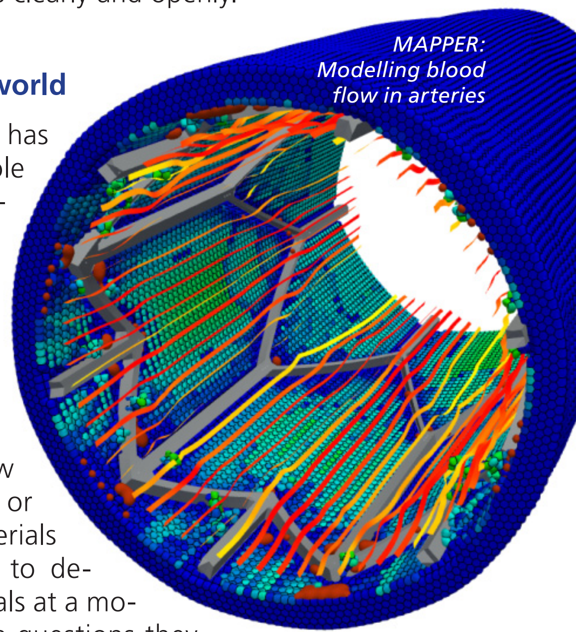
Visualising information can also connect people with scientific data more instinctively: "Visualisation is about placing data in a human context," said Jer Thorpe, data artist-in-residence at the New York Times, as he presented a display mapping out key moments in his life as points on a map for a recent TED (Technology-Education-Design) talk. As one of a new breed of info-graphics designers – whether graphic artists interested in data, or statisticians interested in communicating data visually – Thorpe plays a key role in helping the public to understand complex issues in science. Just as the free press has made political decision-making an open process in democracies, so these information journalists are making scientific data truly available to everyone, explaining research findings clearly and openly.



This image shows the 'hole' in the ozone layer above Antarctica in 2007. Powerful visualisations like this helped to ban the use of CFCs as refrigerants, which caused the hole in the protective ozone layer, in the 1980s.

Credit: Public Domain/NASA

An eye on the biomolecular world

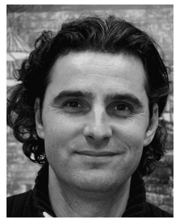


MAPPER: Modelling blood flow in arteries

Visualisation has an important role to play in making e-Science more attractive to molecular scientists. Whether they are life scientists searching for a new drug candidate, or chemists or materials scientists trying to design new materials at a molecular scale, the questions they are faced with come down to 3D shapes. The capability to render a 3D model of a molecule has been around for decades, but predicting complex 3D structures from numerical data, and calculating how that molecular structure interacts with other molecules, is computationally-intensive, and has begun to benefit from advanced e-Science technologies. Emna Harigua, a PhD student who works at the Institut Pasteur in both Paris and Tunis, studies computational modelling of a single protein in a parasite that causes leishmaniasis, a disease that can cause damage to the spleen and liver.

Using the open source software Dock, Harigua has been able to see which potential drugs should interact most strongly with her protein. She has been able to screen 85,000 drug candidates down to less than 100 that could potentially treat the disease. Her work won her the 2012 L'Oreal International Fellowship for Women in Science.

Emna Harigua, PhD candidate- Institute Pasteur, Paris & Tunis – *"Visualisation in molecular biology is very important, making e-science accessible to scientists with a variety of backgrounds, not just the highly technically literate. It enables scientists to gain a better insight into their work. In my case, visualising potential drug candidates interacting with the protein I study makes things so real. It helps me cast a critical eye over the results."*



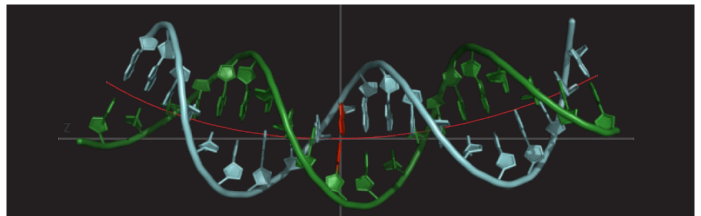
Alexandre Bonvin, WeNMR – *"Visualisation isn't just an endpoint – it is integral to how we do our research. The computer automates a lot of the process, but you still need a human eye to judge what is happening. To design a drug, you look at how the chemical groups involved in biomolecular interactions come together to rationalise your next step. Without visualisation, we would be blind."*

WeNMR: Life through the eyes of a protein

Alexandre Bonvin is Professor of Computational Structural Biology at the Bijvoet Center in Utrecht, and Coordinator for WeNMR, a project bringing collaborative, distributed computing technologies to structural biology. Bonvin uses nuclear magnetic resonance (NMR), a variant of the magnetic resonance imaging (MRI) that doctors use to scan hospital patients, to determine the distances between atoms in proteins, producing a NMR 'fingerprint'.

Proteins are large, flexible molecules that can adopt a wide variety of conformations – different shapes in 3D space – which can give rise to a vast number of subtle variations of this fingerprint. The large number of possible conformations makes figuring out what the 3D structure of the protein is very computationally-intensive, especially when only limited experimental information is available. "In order to calculate the structure, we need to do not one calculation but many tens of thousands of calculations depending on the problem," explains Bonvin. Applying grid technologies speeds up the process of turning these large tables of numbers into 3D models, enabling scientists to quickly understand what the proteins look like at minuscule scales. This is important because many medicines are based on how small drug-like molecules interact with protein-based receptors.

Henry Hocking – a postdoctoral researcher at The University of Utrecht – has been using the WeNMR infrastructure to investigate cone snail venom, a cocktail of small peptides that causes paralysis in the snail's prey. "The peptides bind to nerve cells and stop pain signals from reaching the brain," explains Hocking, "and we believe we can use the 3D models we obtain to design a powerful local anaesthetic."



Visualisation is extremely important in biomedicine, where molecular shapes hold the key to disease and treatment

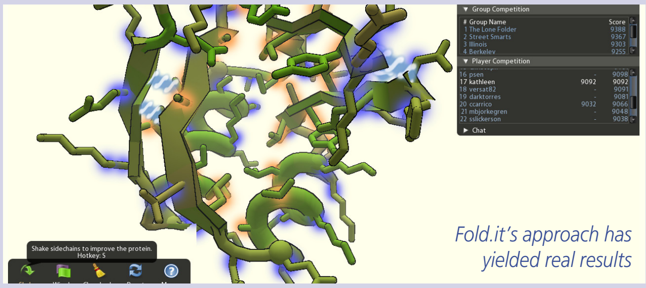
A visible change for chemists

The new wave of molecular sciences has much to gain from adopting e-Science methods in research, where chemistry meshes with materials science to produce molecular electronics, nanotechnology, novel solar power, communications, and display systems. The complexity of such systems means that modelling how they work can help to cut research times and costs dramatically. However, the interdisciplinary nature of such research means that many researchers are not fully familiar with the often physics-born technologies.

SOMA2 – a web browser-based workflow environment for graphical molecular modelling – has been developed at CSC, the Finnish IT Center for Science. "No Unix technical skills are required to access the powerful computational tools," explains Tapani Kinnunen, who has developed the gateway. "It's an intuitive and versatile visual environment that eliminates unnecessary repetitive work." The application, which has been funded by the EGI-InSPIRE project, runs on almost every web browser, and was launched in March 2012. MoSGrid, an application developed at Ludwig-Maximillians University in Garching, Munich, also aims to attract chemists, who can submit jobs to the grid – a network of computers linked together to share their processing power – without getting bogged down in the technical details. "A user can submit their job – whether it's quantum chemistry or a molecular dynamics simulation – without having to know about the middleware," explains Sonia Helles-Pawla, of MoSGrid "and then just get their results back. For a chemist, that's saving them a lot of time." Of course, the more quickly and accurately results can be obtained, the quicker the chemist can visualise their structures using 3D models.

A democratic vision for science

Since the launch of SETI@home, volunteer computing has given us all the chance to contribute to science by donating spare computer cycles. However, in much the same way as joining a political debate is more empowering than listening to one, being able to actively participate in research projects by providing observations or interpreting visual information clearly separates what is known as 'citizen science' from volunteer computing. Two of the most popular citizen science projects, fold.it and Galaxy Zoo, allow those taking part to interact with, manipulate and assess visual data. Fold.it is presented as a web browser-based game, where players score points for finding optimal conformations of protein folding. In 2011, Fold.it players helped to find the structure of M-PMV, a virus that causes AIDS in monkeys



Fold.it's approach has yielded real results

Galaxy Zoo is another highly successful citizen science project, this time allowing users the opportunity to assess deep space objects from the Sloane Digital Sky Survey. Because the information is purely visual, participants need no training. "We thought about giving people tutorials and so on," said Chris Lintott, the academic behind Galaxy Zoo in a recent interview, "but quickly saw it would be more effective – and fun – to have people get going straight away, and use the sheer volume of observers to ensure accuracy. " The project has turned some accepted theories on their heads, even identifying new types of astronomical phenomena such as Hanny's Voorwerp.

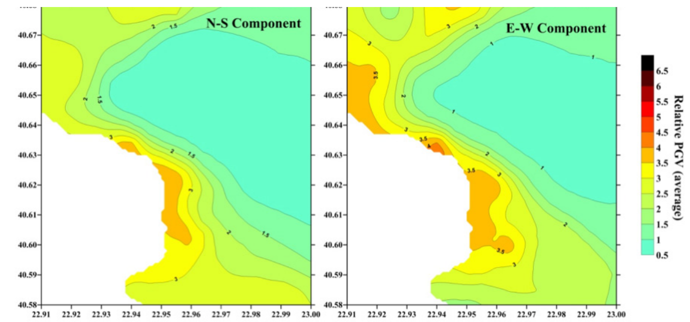
Tim Adams, The Observer New Review, 18.03.12

Mapping the Physical World

Physical processes in the real world such as water cycles, earthquakes and weather are incredibly complex, involving ecosystems with large numbers of interacting components. While DRIHMs and MAPPER are providing tools to understand hydrometeorological processes better, important work is also underway to understand seismic phenomena. Danseis, led by Hans Thybo at the University of Copenhagen, is visualising magma plumes beneath the Earth's crust using supercomputers. "It is believed there are around 30 plumes scattered around the planet," explains Thybo. "However, it isn't known for sure that they actually exist. Being able to model and visualise where they are on the globe can help scientists to better understand geological features and seismic events."

Hellasgrid: Seismic studies

Researchers based in the Geophysical Laboratory at the University of Thessaloniki are modelling the propagation of seismic waves as they might occur in a number of different scenarios, treating the Earth as a viscous, elastic resonator and taking account of local soil structures. Using Hellasgrid and the European Grid Infrastructure (EGI), the energy map they produced closely matches with a real map showing areas of intensive damage during the 4 July 1978 earthquake.

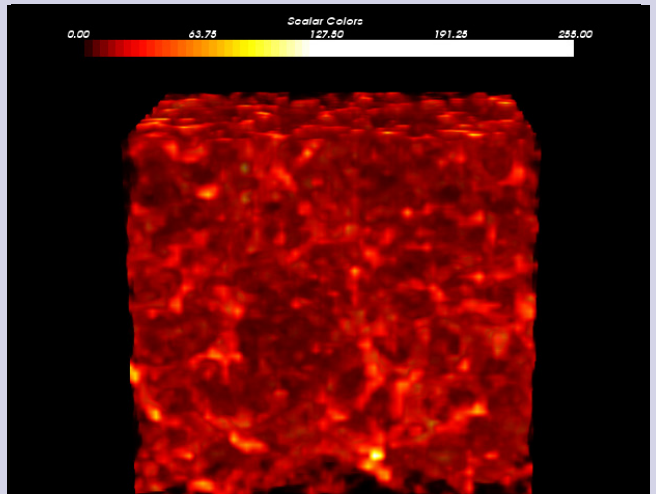


Modelling seismic waves in Thessaloniki

"It is quite a good correlation," explains Andreas Skarlatoudis, who developed the computational model for his PhD work. "It shows the model works." Accurate models like this could be used to predict which locations are likely to see high levels of damage in other earthquake-prone regions, and either strengthen buildings or avoid building certain structures, such as nuclear power plants, there entirely. Visualisations like this can help us to prepare for the worst, and can even inform how we react to data as a society, something that numbers alone often fail to achieve.

VisIVo: See the Universe on your iPad

"Being able to perform large scale simulations of the cosmos has traditionally been reliant on the availability of supercomputers," says Alessandro Costa, a technologist at the Astrophysical Observatory of Catania (INAF), "because most computational cosmology applications were developed for them. But supercomputers are often not readily available, and cannot be dedicated to a single project. This reduces the access time scientists have to be able to carry out their research."

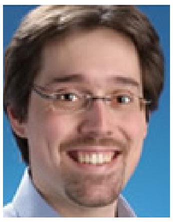


Galactic structures on the grid

The solution, executing calculations in parallel on many CPUs, has not always been easy to achieve. "Running these applications in parallel on a cluster has also been problematic, because it requires very low latency to be efficient," explains Costa. Latency is the time it takes to move data around in between different machines in a cluster – essentially, wasted time. "However, in recent years grid infrastructures have begun to offer latencies comparable with supercomputers, making grid suitable for cosmological simulations."

Costa is a developer working on VisIVo, a visualisation and analysis application for accessing observational and theoretical astrophysical data. It provides a customised application-specific gateway to SCI-BUS (Scientific gateway Based User Support), an FP7-funded project that provides seamless access to major European distributed computing infrastructures, including clusters, supercomputers, grids and clouds. For VisIVo, the software running 'behind the scenes' on the grid is FLY, a program running on gLite middleware that can calculate the forces between hundreds of millions of stars to understand the effects of gravity between them. The VisIVo application itself runs on Apple iPads and iPhones, placing easy access to the grid (which does the calculation-intensive work) within the grasp of many researchers. "This is a fast way to access SCI-BUS," concludes Costa, "and of course this technology could be used by other projects."

What is middleware? : A number of general purpose software libraries must be installed on a grid so that all the machines in the gird 'speak the same language'.



Nicola Rebora, DRIHMS, hydrometerology study – *"Visual records from citizen scientists, such as videos, are extremely important in helping us to develop the hydrological models we use for DRIHMs. Indeed, we start with the requirements of the citizen scientists, and provide them with computational power, which is a complementary approach to volunteer computing. Their data can be exploited, empowering them to understand flash floods in their own area."*

Hellasgrid: air quality

In the Department of Meteorology and Climatology and the Laboratory of Atmospheric physics, researchers are using Hellasgrid and EGI to model the effects of climate change on air quality. Focusing on ozone, a gas that at low levels can react with other pollutants to create a toxic blanket of smog in cities, lead researcher Eleni Katragkou found that levels could increase drastically by the end of the century. "Until the 2040s, the increase in ozone will be less than one part per billion (ppb)," she explains, "But by the 2090s, it could increase by 6 ppb, especially in south-west Europe."

Human Mapping

Maps have a special place in data visualisation, and are probably our oldest method to represent information in an abstract form – one that transcends cultural and linguistic boundaries. London physician John Snow's 1854 map of the Soho cholera epidemic, which identified a focal point around which cases clustered – a contaminated water pump – spurred the local council to take action, promptly shutting the pump off and stemming the cholera cases. In this sense, maps are empowering. But they also have the potential to be democratising, both in the sense of making data understandable, and also helping minority groups to find solutions to problems that concern them.

Louise Francis, Mapping for Change, University College London – *"Mapping for Change works with communities across the UK to enable them to collect their own data. We then use mapping technologies to visualise that data. We've helped a community in Deptford, South East London, record air quality around a scrap metal yard close to where they lived, using nitrogen dioxide badges. Collecting data like this can empower citizens with their campaigning activities."*

Mapping our environment is the first step to understanding how to solve the problems we face, from disease to deforestation

Mapping for Change: Extreme Citizen Science

Muki Haklay, Louise Francis and Claire Ellul of University College London have been empowering communities in the UK and Africa by persuading the public to record their environment using visual means. Using simple measuring tools and GPS devices, projects such as Mapping for Change have brought real differences to communities in inner-city areas of London. As Professor of Geographical Information Science, Haklay wants to go further, transcending linguistic and literacy boundaries.

At the Citizen Cyberscience summit in London in February 2012, he discussed Extreme Citizen Science (ExCites), which aims to empower people regardless of their literacy. Funded by the UK Engineering and Physical Sciences Research Council, the project puts adaptable scientific tools in the hands of people in remote regions.

One project in the Congo basin of Cameroon, enables a forest community to collect data about their own environment using a highly visual, pictorial interface. "They can then analyse it and understand different changes that are happening in their area, monitoring aspects that they care about. It's a more interesting project now than when we first started." Using the pictorial interface, the forest community have mapped out resources and sacred sites important to them, helping to coordinate conservation efforts in areas where forestry companies also have an interest. Efforts like this could prove as important as John Snow's cholera map, enabling people to identify both problems and solutions for their changing environment.

Ultimately, visualisation is about taking tables of numbers, and representing them in a form that makes sense to us as human beings.

For more information:

- ***www.flowingdata.com***
- ***www.sci-bus.eu***
- ***www.drihm.eu***
- ***www.mapper-project.eu***
- ***www.mosgrid.de***
- ***www.hellasgrid.gr***
- ***www.csc.fi/english/pages/soma***
- ***www.communitymaps.org.uk***

Scan this QR code into your smart phone for more on this e-ScienceBriefing

EGI : www.egi.eu
Real Time Monitor: rtm.hep.ph.ic.ac.uk
iSGTW: www.isgtw.org
e-ScienceTalk: www.e-sciencetalk.org
email: info@e-sciencetalk.org

Books:

***Edward Tufte** – The Visual display of quantitative information*

***Michael Nielsen** – Reinventing Discovery: A New Era of Networked Science*

e-ScienceTalk is co-funded by the EC under FP7
INFSO-RI-260733