Inspired

ISSUE 15
APRIL 2014

news from the EGI community

(Matti Mattila / wikicommons)

TOP STORIES

MORE

European Grid Infrastructure

www.egi.eu

This Issue

In this issue...

> Franck Michel introduces VAPOR, a new set of tools to make VOs easier and cheaper to run

> Jan Bot writes about the experience of creating a MOOC from scratch

> Gergely Sipos unveils the EGI Engagement Strategy

> Anna Slater writes about the use of grid computing to design a toxic chemical-eating bacteria

> Neasan O'Neill interviews astronomer Thijs Coenen

> Sergio Andreozzi asks if Open Science can boost EGI's impact

> Javier Jiménez writes about a new iteration of the EGI Solutions

> Kitti Varga and Akos Ajnal introduce DataAvenue

Please send your comments, feedback and suggestions to: sara.coelho@egi.eu Thanks!

The Lutheran Cathedral is just across the Helsinki University Main Building, where the Community Forum will be held. (Illustration: Matti Mattila / wikicommons)

Lightning talks in Helsinki

It's almost time for the EGI Community Forum 2014 (19-23 May)

The final preparations for the EGI Community Forum 2014 in Helsinki (19-23 May) are well under way.

The timetable is available online at: *http://go.egi.eu/CF2014*

The list of keynote speakers is finalised and we have abstracts and more information about the speakers on the conference website: *http://go.egi.eu/keynotes*

Lightning talks

One of the new features of the EGI Community Forum 2014 will be a series of Lightning Talks by researchers who can demonstrate the scientific impact of grid computing.

After a successful call for participation, the organisers selected the following talks for the Opening Plenary (19 May):

> *Towards the big data strategies for EISCAT-3D*, presented by Ingemar Haggstrom

> *Implementing an air quality forecasting system on the grid*, presented by Anastasia Poupkou

> *Large-scale search for periodic solutions to the Newtonian three-body problem*, presented by Milovan Suvakov

The range of talks aims to cover different areas benefitting from the support of EGI as an e-Infrastructure for science. Ingemar's talk highlights the importance of EGI for large-scale, collaborative, pan-European projects in the ESFRI roadmap. Anastasia's contribution demonstrates the importance of distributed computing to tackle societal challenges. And Milovan's work is an example of excellent science with high scientific impact in its field.

And more...

We will also have a dedicated session for networking, posters and demos on Wednesday 21 May 11:00-12:30. The Networking Session will be the ideal occasion to interact with colleagues, see what is new and brainstorm about ideas and future collaborations.

http://cf2014.egi.eu

The VAPOR portal: toolkit for VO management

Franck Michel introduces a new set of tools to make VOs easier and cheaper to run

Virtual Organisation (VO) managers and their support teams have a large variety of tools and portals available to assist them in managing their community and operating resources. A few examples are GStat to browse the information system, SAM Nagios to monitor availability, the CERN Experiment Dashboard or the VO Operations Portal. Although generic, some of these tools are often designed to meet the needs of somewhat specific contexts (e.g. a large community with well-established resource management policies), and their operation requires solid IT support.

For small or newly emerging communities, starting up the operations of a VO, having an outlook on the set of storage and computing resources that support the VO, monitoring activity, dealing with the files spread out by users, can be difficult tasks.

VAPOR

The **V**o **A**dministration and operations **POR**tal, was developed to help small to medium-sized grid user communities to perform common VO administration and operation tasks, at a reduced human cost. Such communities typically have little or no dedicated IT support and may use resources either in reserved or opportunistic manner. VAPOR is generic, experiment-independent, and complements existing tools such as SAM Nagios and the VO Operations Portal.

VAPOR's features were designed to (1) provide statistical reports and status indicators about the resources supporting a VO, and (2) facilitate the management of VO users' data.

With VAPOR, VO managers can:

(1) generate on-demand reports to (i) visualise status information of resources supporting a VO, (ii) monitor the running and waiting jobs to help detect peaks of activity and bottlenecks, (iii) monitor computing elements success rate and time response, and (iv) compute a so-called white list of computing elements to feed a job submission system.

(2) use data management features to (i) monitor storage elements to prevent them from feeling up, (ii) check and fix inconsistencies between the storage elements and the file catalogue (dark data, lost files), and (iii) clean-up files left behind by former users.

VAPOR has been progressively deployed during the first quarter of 2014 and is already supporting the VOs biomed, compchem, enmr.eu, vlemed, shiwa-workflow.eu and vo.france-grilles.fr.

We believe that VAPOR will help those existing communities in daily administrative and operational tasks, and will possibly enable the mutualisation of tasks between several communities.

How to use VAPOR

VOs can start using VAPOR on demand by sending a request to vt-vapor@mailman.egi.eu.

Or they can easily deploy their own instance of VAPOR using detailed documentation: http://go.egi.eu/vapor-install

Project wiki: https://wiki.egi.eu/wiki/VT_VAPOR

Bringing grid computing to the masses: the SURFsara & EGI Massive Open Online Course

Jan Bot on the experience of creating a MOOC from scratch

© Marie-Lan Nguyen / Wikimedia Commons / CC-BY 2.5

Massive Open Online Courses (MOOCs) are complete courses taught to many students using an online platform, with the content being freely available to anyone interested. This new form of education has recently become popular and many universities have experimented teaching their courses online, using educational platforms such as Coursera, EdX or Udacity.

Grid computing is perceived by many as being hard to learn. The theoretical background in distributed computing and the need for hands-on experience is hard to obtain for many scientists working in fields with little or no experience in large-scale computing.

Seeing this opportunity, we developed a MOOC to help these scientists taking their first few steps on grid computing.

Grid computing is well suited to be taught as a MOOC, due to its distributed and international nature. With the backing of the EGI-InSPIRE project, we set out to create a course which would explain some of the fundamental concepts behind grid computing, while also providing the hands-on experience needed to start working with it directly.

The first step was to work on the support materials: slide decks, quiz questions and assignments. Then, we asked different people from the EGI community to contribute use cases to show how grid computing is currently being used. The use cases were, just as the course's lectures, recorded as screencasts with voice-overs to explain the material and covered a variety of scientific topics, including astronomy, life sciences and climate modelling.

The course ran from November 2013 until the end of January 2014. Over 350 people joined, of which 30 completed the final assignment. A 10% graduation rate is actually to be expected when working on a MOOC, as most students are only interested in exploring some parts of the content.

Our students included PhD students from different fields, programmers based at Dutch medical centres and astronomers from LOFAR.

The students had the opportunity to get hands-on experience on grid computing by using a pre-configured Virtual Machine (VM) with all the necessary software tools installed. The VM provided the grid environment so that the students could test their understanding on quizzes, weekly exercises and the final test. The quizzes were graded and students with an average score above 60% advanced to the final assignment.

The feedback we received is encouraging. Of the students who completed the final survey, 92% (23) gave the MOOC a 4 or 5 out of 5 grade, and about half of these said they will use grid computing for their research.

The MOOC course materials have been added to the SURFsara website where they are now part of the grid documentation. The material created for the course is still available and can be used by anyone interested.

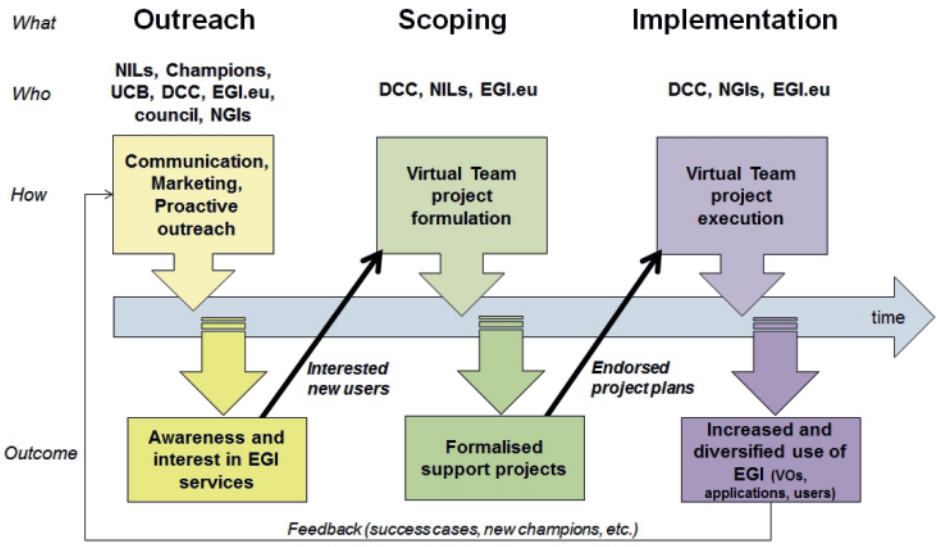
The MOOC team

Anatoli Danezi,
Jeroen Schot,
Tijs de Kler,
Nico Kruithof,
and Jan Bot

url: http://mooc.uva.nl/portal

The EGI Engagement Strategy

Gergely Sipos summarises plans for future engagement with user communities



Sustainability is essential for e-Infrastructures and the scientific communities that they support. Many of these scientific communities have research agendas measured in decades and need to be assured of the continued operational presence of the e-Infrastructures that they adopt.

EGI's sustainability plans have become increasingly coupled with its long-term strategy to *connect researchers from all fields of science across the whole European Research Area (ERA) with the reliable and innovative ICT services from EGI that they need to undertake their collaborative world-class and world-inclusive research.*

Engagement is a key element of this EGI strategy with the following goals:

1. Identify scientific communities from the ERA that can benefit from the use of solutions provided by the EGI community.

2. Reach out and engage with communities about ICT technologies to capture their requirements.

3. Help communities to address scientific challenges with existing EGI solutions, or to develop new solutions as needed.

4. Support scientific communities to become self-sufficient users of the EGI e-Infrastructure services.

The **Engagement Strategy** describes the goals and targets of EGI engagement activities and provides information about the human networks, online resources and tools that will help us to implement these activities.

In the next two years a growing number of Research Infrastructures (RIs) from the ESFRI and national roadmaps are expected to reach implementation or operational stage. These RIs are already exploring the current and future needs of their user communities and they bring together a wide diversity of stakeholders looking for solutions to many of the problems science is facing today.

The ESFRI RIs, their preparatory projects, and other multinational scientific collaborations are considered as the primary potential beneficiaries of EGI services and therefore the prime targets of EGI Engagement activities.

A second target group are small research collaborations and research networks. Unlike RIs, these groups may not be aware of the benefits of e-Infrastructures to science, so discussions have to start at a more basic level.

The EGI Engagement workflow aligns elements into a process that helps EGI to reach new users, and support them through the use of EGI services. This workflow consists of three phases:

1. Outreach: Using marketing and communication, this phase aims to identify ERA members whose work could be lifted to the next level by EGI's e-Infrastructure services, and inform them about the EGI solutions and possible engagement options.

2. Scoping: In this phase, engagement with new users is deepened; detailed requirements from their e-Infrastructure use cases are captured and translated into focussed support project plans.

3. Implementation: This phase executes Virtual Team projects that, after successful completion, will enable the users to reach new frontiers in science. The projects will also indirectly result in a diversified use of EGI's solutions.

Designing a toxic chemical-eating bacteria, between the lab and the computer

Anna Slater explains why grid computing was important

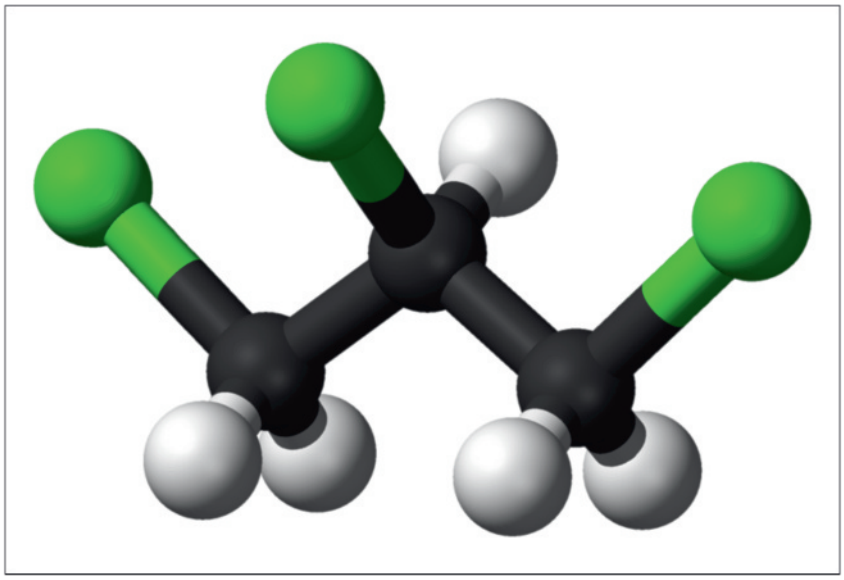


Diagram of the TCP molecule, with Cl atoms in green. (Source: wikicommons)

For centuries man has harnessed various chemicals in industry and research. They are not all benign though and contamination of land and water is a potential risk. Millions of years of evolution have created an astonishing diversity of bacteria capable of processing countless chemicals into more harmless substances. These too have been taken advantage of, but some chemical compounds used by man do not exist naturally.

TCP (1,2,3-trichloropropane) is one such compound. First created during the industrial revolution it is now used as a pesticide and in chemical manufacturing. It is also toxic and persistent, once released in to groundwater, it can stay there for over 100 years. A potential solution is clearly a bacteria that could convert TCP into a harmless compound, quickly and cost-effectively. Yet there hasn't been sufficient time for one to evolve. Now a group of scientists at Loschmidt Laboratories of Masaryk University in the Czech Republic have used the grid to help evolution, by engineering synthetic bacteria that can metabolise TCP into harmless glycerol.

The team, led by Jan Brezovsky and Zbynek Prokop, used a combination of grid-powered mathematical modelling and laboratory experiments to design and test a five-step metabolic pathway for a new bacterium. The key is balancing the activity of multiple enzymes to keep the bacterium alive, while sustaining a high processing rate. The team were also able to engineer new enzyme variants, which adds further complications of what variants to create, and their most effective combinations and concentrations.

The scientists used computer-assisted directed evolution to predict and plan the creation of the enzyme variants. These were then created and tested for their performance. Using these values the team then created a model to simulate the vast number of possible enzyme ratios, in order to find the optimum combination, which would balance toxicity and glycerol production.

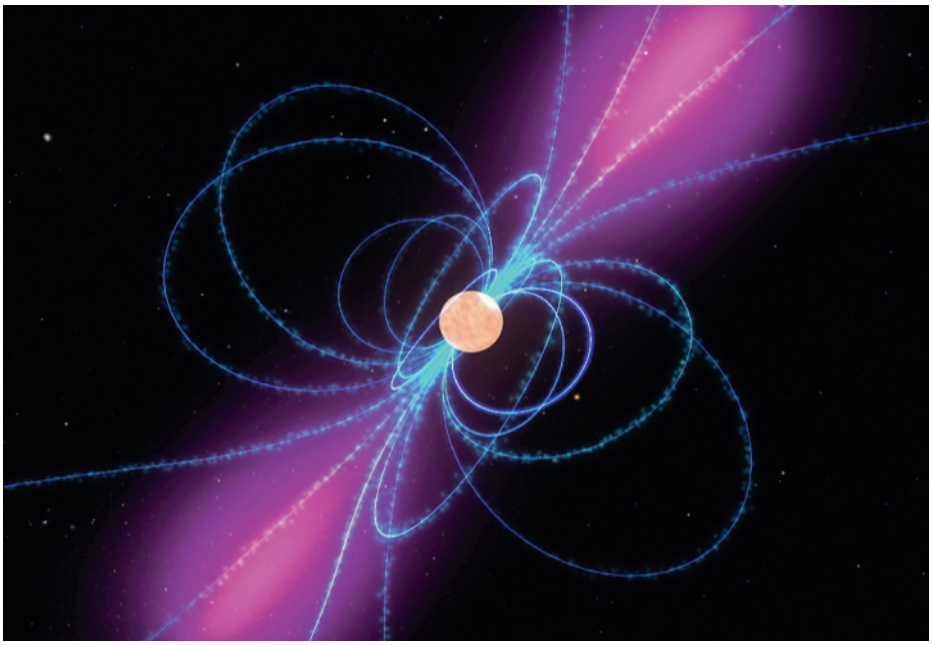
The team consumed around 100 CPU days using the grid computing resources provided by Metacentrum in the Czech Republic. The grid-assisted approach offered a twofold advantage. "We could explore combinations in much higher detail since we could afford to run variants with minimal mutual differences," said Jiri Damborsky, one of the scientists in the group. More importantly, grid resources enabled tight interaction between rounds of experimental work and the ongoing refinements and predictions of the model.

"Taking into account the requirements for high accuracy and the time demand of computationally intensive calculations of differential equations, grid calculation was the only reasonable choice."

The predictions of the model were used to create bacterial strains containing the optimal enzyme ratios. These real-life bacteria behaved very closely to what was predicted. The model also revealed limitations of specific enzyme variants, and predicted ways to improve pathway function. "Our study highlights the potential of forward engineering of micro-organisms for the degradation of toxic anthropogenic compounds, and the essential role of grid computing to design complex biological systems," Jiri concludes.

Searching high and LOFAR pulsars

Neasan O'Neill interviews Thijs Coenen about his research and the added value of grid computing



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 wikicommons)

What is your research?

I am an astronomy researcher with the LOFAR Pulsar Working Group based at the University of Amsterdam. My work focuses on discovering nearby weak pulsars using data from the LOFAR Pilot Pulsar Survey (LPPS) and the LOFAR Tied-Array Survey (LOTAS).

What are pulsars?

Sometimes near the end of a star's life their core gets super dense and collapses under its own weight. This results in a massive explosion, or super-nova. These in turn create neutron stars that rotate at high speeds and emit radio signals, light and radiation, like a cosmic lighthouse. These are known as pulsating radio stars or pulsars.

So what is LOFAR?

The Low Frequency Array (LOFAR) is a new radio telescope that consists of receivers spread out over Western Europe, with most being based in the north of the Netherlands. It uses a large number of antennas that are combined through software to form one large telescope to detect weak, low energy radio signals. LOFAR has a unique view of the sky, it is much more sensitive than other observatories operating at similar frequencies and is one of the largest in the world.

Why study pulsars?

Pulsars are valuable natural laboratories for astronomers because the conditions they are host to cannot be created on Earth. Since their discovery in 1967 they have given us insight into gravitational fields, planets outside our solar system and how matter behaves at extremely high density.

How has the grid been useful?

Processing the data from pulsar surveys is 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 and connected to the European grid I could efficiently search the data. It made my life a lot easier both in terms of the overall time taken by the computational task and the time spent interacting with and babysitting the data processing.

What kind of results have you had?

Back in November as part of the defence of my PhD I was able to announce the first new pulsars discovered by LOFAR. It not only was a big part of my PhD but also validates the approach we have taken for the LOFAR pulsar surveys and bodes well for the full-strength pulsar survey currently underway

And what does the future hold?

I continue to work with LOFAR on pulsars and I expect we will have more announcements about new pulsars as time moves on. However it is a large project and LOFAR will have an impact on many other areas of astronomy from our planet's own ionosphere to massive, distant, galaxies formed at the very beginning of the universe.

More information

url: http://www.lofar.org/

Can Open Science boost impact in a 'Publish or Perish' reality and make EGI's research outputs more visible?

Sergio Andreozzi and Ivo Grigorov on the advantages of Open Science

© Marie-Lan Nguyen / Wikimedia Commons / CC-BY 2.5

Excellence in science, bright ideas and innovative thinking will always be a top skill when embarking on a research career. But in this highly-competitive age, 'how' we do research can contribute as much to impact as the ingenuity of the research ideas themselves.

The definition of impact is changing in the eyes of the funders from simple citation counts to how much societal engagement is achieved by the research results. Funders are no longer just satisfied with clever ideas, but want to see them accessible to citizen scientists, the private sector, and society at large. The Horizon 2020 criteria are just one last example of how open research results receive equal importance to the innovative character of that research.

Open Science provides concrete ways to improve the day-to-day research routine. Instead of being just another burden necessary to pass the review process, Open Science in the Internet age offers the potential to make individual research more visible, talked about, followed and understood by the wider scientific community, inviting serendipity and potential new collaborations.

How to optimise the chances for the evaluation process, while boosting the personal profile of the individual researcher, is the topic of session 'Open Access to EGI Research Outputs' during the upcoming EGI Community Forum in Helsinki (21 May 2014). The session will also cover the services EGI is developing to support the implementation of the Open Access policy and to allow researchers to publish and share the research outputs that are possible thanks to EGI. The aim is to increase discoverability and reuse in order to accelerate innovation and research.

Making EGI's value visible

Open Access also allows EGI to demonstrate its scientific impact by linking research outputs to its services and resources, or to the services and resources provided by the National Grid Initiatives and virtual organisations.

The collaboration with the OpenAIRE project is key to this goal and the new functionalities that this project has been developing for EGI will also be presented at the Community Forum. EGI.eu will present its plans for developing a federation of digital repositories for open data to be offered to the communities.

With this, EGI aims to stimulate a change in the mind set of researchers and highlight that research and e-Infrastructures should become a primary concept that should be citable and linked to publications, data, applications and so on to recognise their value, but also to increase the visibility as enablers of new scientific discovery for today's researchers and tomorrow's citizen scientists.

More information

Sergio Andreozzi is Strategy & Policy Manager at EGI.eu
sergio.andreozzi@egi.eu

Ivo Grigorov is a project manager for Technical University of Denmark and member of the FP7 FOSTER Steering Committee
ivgr@aqua.dtu.dk

Open Science workshop @CF2014
http://go.egi.eu/egicf14-oa

The EGI Solutions Portfolio

Javier Jiménez writes about the new version of the EGI solutions

The European Grid Infrastructure, as a community, serves the needs of multiple resource providers, research communities and the National Grid Initiatives (NGIs). All these consumers have different needs, problems and expectations, which are evolving constantly and quickly.

The EGI solution portfolio was developed during 2013 to present dedicated answers to specific user needs. But as the users' requirements evolve quickly, so do the EGI Solutions.

The EGI solutions

The **Community-Driven Innovation & Support solution** addresses the way EGI responds to the researchers' support queries. Whenever researchers encounter a challenge accessing EGI resources, they can, as before, knock on many doors. But if the problem requires a new technology, it is now possible to summon a group of experts to put their brains together and create an innovative answer. This will, then, become part of the pool of previously existing applications, workflows or any other already existing approach.

The **Federated Cloud solution** is the long-awaited response to the demand for a European federation of academic clouds. With this solution, researchers obtain a single cloud system for their research activities, which they are able to scale to their requirements, which is fully resilient and free from vendor lock-in. The user-researchers can focus on their core work and obtain new approaches to their work.

The **Federated Operations solution** is aimed at the resource infrastructure providers of EGI, to make their operations even more efficient and effective, or at those wishing to become members, to guarantee seamless integration. This solution relies on the lightweight standards family for service management in IT service provision, known as FitSM, which EGI has helped shape. This is a breakthrough in the implementation of service management routines in all federated IT service provisioning, not necessarily related with scientific production. With the focus on sustainability, the expertise achieved within EGI might also be valuable for other IT services organisations in their struggle for quality management in a federated environment.

The **High-Throughput Data Analysis solution** represents the core of the EGI activity, which is the provision of high quality data- and computation intensive resources in a distributed infrastructure.

New tools for make your work easier

The solutions are described in a series of white papers describing the problems they address, who are the beneficiaries from the solution and how the solutions are built to provide value.

These documents can be used and delivered to your contacts at all levels. They aim to help you in explaining what is what EGI does, and what EGI's value proposition is.

White papers

The EGI Solutions are described in the following white papers:

Federated Operations
http://go.egi.eu/2196

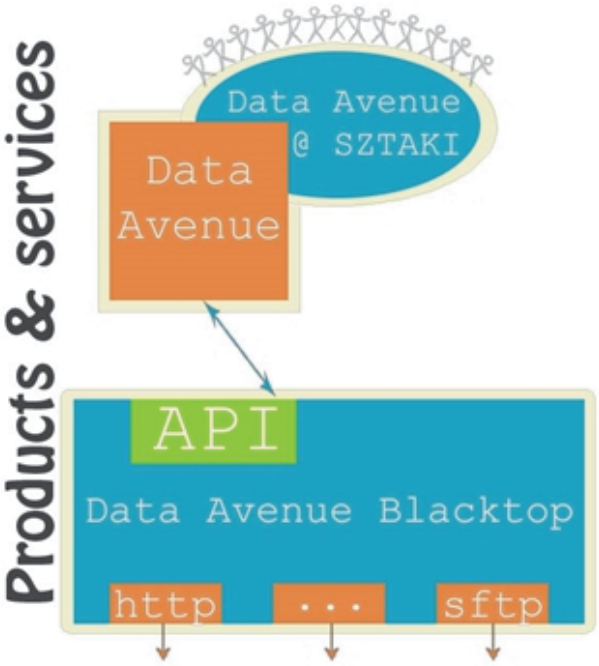
Federated Cloud
http://go.egi.eu/2197

HTC Data Analysis
http://go.egi.eu/2198

Community Support
http://go.egi.eu/2199

Data Avenue: a new tool for data transfer

Kitti Varga and Akos Hajnal on a new file commander tool for data transfer



Data Avenue is a file commander tool for data transfer, enabling easy data movement between storage services such as grids, clouds, clusters and supercomputers using various protocols, such as HTTP, HTTPS, SFTP, GSIFTP, SRM, and S3.

If you have ever faced problems transferring data to the environment where your experiment is running, then Data Avenue is made for you.

The Data Avenue interface allows you to browse, download and upload data to and from the supported data storages, and move data easily between them even if they are accessed by different protocols.

The storage solutions used today in distributed computing environments have been developed to serve specific goals by fulfilling specific criteria. Accessing these storage resources, however, requires dedicated protocols and tools to allow users to organise, manage, up- or download data via a graphical interface or command line. Selecting the most suitable storage resource for our requirements thus involves setting up the appropriate software environment, and learning how to use the related tools.

Data Avenue offers a great, easy-to-use solution to this problem.

Data Avenue has an easy-to-understand user interface and it is very similar to a file commander tool that everyone can use. It has two parallel windows to open the origin and the target of the file to be copied. The user has to authenticate both sides and after the successful log in, the files can be transferred between the sources without using the user's local computer storage.

Data Avenue can be accessed from everywhere, using a computer with an up-to-date browser and an internet connection.

The Data Avenue family consist of three members.

The Data Avenue@SZTAKI is a stand-alone installation of the Data Avenue portlet hosted at MTA SZTAKI that can be used easily and quickly to try out the features of Data Avenue without any installation.

The Integrated Data Avenue is a portlet integrated into WS-PGRADE/gUSE that offers a web-based interface for managing data storages with the help of Data Avenue Blacktop.

For developers, Data Avenue Blacktop is the core service that can be used to perform different operations related to storages through a web service interface.

More information

url: https://data-avenue.eu/

If you would like to learn more about the Data Avenue service, there will be a live demo of the Data Avenue service at the European Grid Infrastructure Community Forum 2014 in Helsinki, Finland.