e science TALK

GridBriefings

Talking about e-science

Mapping the e-Infrastructure Landscape

Today the World Wide Web provides information for people across the globe but, as yet, no single networked system provides a similar service for researchers to help them access, share, store and process large amounts of data. With this in mind three reports – a "Blue Paper" from the e-Infrastructure Reflection group (e-IRG), a High Level Expert Group report on Scientific Data and the Distributed Computing Infrastructure Collaborative Roadmap – have recently detailed ways in which Europe's e-Infrastructures can work together to ensure a more integrated service. This briefing details the findings of these reports and the actions we can take to provide a harmonised landscape of services for our researchers.

A global vision

Building a coordinated e-Infrastructure landscape is a little like building a railway before matching track sizes were agreed – or even standard time zones! e-Infrastructure providers are currently laying the 'tracks' that will allow researchers to access whatever data or computing power they need, simply and quickly. However, in order for this to work, each provider needs to ensure that their tracks are built to a set of standards, so that the trains that run over them have unimpeded access to the entire network (different e-Infrastructure providers) and don't stop short.

In practice, developing such an integrated e-Infrastructure is much more complicated than making sure that one size fits all – no scientific discipline makes the same demands on the infrastructure as the next. But providers expect that by working together, and pinpointing the areas of common need, they can present researchers in every field with an integrated e-Infrastructure service that can grow and evolve over time to suit their needs.

Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda : *"Science has always been based on exchange of information and intense interactions between researchers. Today, thanks to the availability of global communication networks, we profit from truly global and massive scientific collaborations. To this end, the EU's Digital Agenda for Europe has called for the development of research infrastructures and e-Infrastructures, including for scientific data."*

What the experts say

In Autumn 2010 three different reports were released which set out ways in which Europe can achieve its vision of a united e-Infrastructure vision:

The e-IRG Blue Paper: The e-IRG is an inter-governmental policy body comprised of national delegates from more than 30 European countries; it works to define and recommend best practices for pan-European e-Infrastructures.

Sharing pictures from space

SMOS satellite
Credits: ESA - P. Carril

European satellites collect a wealth of information about Earth from space. The data produced can be used by researchers to gain a better picture of our planet and increase our understanding of key issues such as climate change. But in order for this to happen satellite operators, space agencies and data providers need to work together so that data is accessible and exploitable to those who need it.

Traditionally in Europe there has been poor cooperation in this field - in the past there was no common approach for long-term preservation and access to space data. Now, things are changing.

The European Space Agency, recognising the need for cooperation and sharing, has launched a long-term data preservation programme that merges all earth observation data from across Europe and Canada. This will give researchers access to the data they need, when they need it. A set of guidelines will help to ensure data can be used across the board while close collaboration with researchers will benefit both providers and users by making sure their ideas work together

Working with ESFRI

The ESFRI (European Strategy Forum on Research Infrastructures) projects are large research infrastructures that span social and biomedical sciences, earth and physical sciences, energy, infrastructures and analytical facilities. For example ELIXIR, an ESFRI project, will build a distributed but interlinked collection of biological data resources and literature for life science researchers.

The ESFRI projects are likely to be key users of Europe's e-Infrastructure. As such it is vital that providers work together with these projects to understand their needs. Only through close collaboration can we provide them with a useable and useful service. These issues are assessed in the e-IRG Blue Paper. Another report by the EEF (European e-Infrastructure Forum) examined the requirements of the ESFRI-projects and outlined the services and resources that the e-Infrastructure community can offer (see GridBriefing no 13, Future needs of the ESFRI projects).

In October 2009, ESFRI asked the e-IRG to examine ways in which ESFRI projects and their users can engage and exploit common e-Infrastructure services. The resulting Blue Paper, released in September 2010, stressed the importance of bringing communities together in order to improve their mutual understanding and collaboration.

The DCI Collaborative Roadmap: In an effort to explore the collaborative opportunities available within the community, six Distributed Computing Infrastructure projects funded by the EU have set out ways in which they can work together over the next two years. The projects – EGI-InSPIRE, European Middleware Initiative (EMI), Initiative for Globus in Europe (IGE), European Desktop Grid Initiative (EDGI), StratusLab and VENUS-C – cover areas from managing a European Grid Infrastructure through to middleware engineering and cloud computing. This DCI Collaborative Roadmap is a starting point to strengthen potential collaborative opportunities between the funded projects.

High Level Expert Group report on Scientific Data: The High Level Expert Group on Scientific Data was asked by the European Commission (EC) to develop a vision of scientific data e-Infrastructures in 2030. The group's resulting report was presented to the Commission in October 2010.

Together these reports encompass every feature of the e-Infrastructure landscape – from fast networks providing the foundations of connectivity through to shared resources and the researchers using them. More detailed findings on the issues these three reports cover are given in the following section.

Connecting people together

Research networks provide researchers with high-quality internet services, allowing them to connect easily and quickly to collaborators. For example, the GÉANT network, facilitated by the EC through the GN3 project, works with Europe's national research networks to connect 40 million users in over 8,000 institutions across 40 countries.

High quality networks provide the keys to cutting-edge Research Infrastuctures (RIs) and as such must stay accessible and easy-to-use in the face of changing research needs. To aid this, the e-IRG recommends that new RIs participate in networking coordination bodies to define, test and use new networking services.

As research resources become more interconnected, aligning authentication and authorisation across infrastructures is also vital. Future pan-European e-Infrastructure and ESFRI projects are encouraged to define their access control policies and mechanisms from the start, in accordance with the standards and best practices adopted by the research community.

www.sxc.hu

The development and spread of remote instrumentation techniques and technologies will also open new opportunities for scientific communities. Sharing expensive scientific equipment such as radio telescopes and synchrotrons through remote use can cut costs and open up the facilities to more researchers. An increasing reliance on remote instrumentation has been identified as a particular issue for the environmental science ESFRI projects. To ensure that these tools can be used and accessed across all of Europe's e-Infrastructure, the e-IRG recommends developing standard interfaces for these technologies.

Steven Newhouse, EGI - *"The EGI-InSPIRE project is establishing mechanisms to bring new innovative technologies into the European Grid Infrastructure in order to support the innovative research taking place within the ESFRI and other research communities. The results from the other DCI projects and activities taking place within the National Grid Initiatives will provide a platform for the development of innovative software services and tools to help support distributed data analysis."*

Riding the data wave

How to tackle the growing amounts of data being produced worldwide is one of our biggest challenges. Data initiatives must work out how to store, access and preserve data for researchers in the decades to come. In response to this, the EC asked a High-Level Group on Scientific Data to look at the challenges and benefits resulting from the rise in data. Their findings are detailed in the recent report 'Riding the Wave: How Europe can gain from the rising tide of scientific data'.

www.sxc.hu

In the report the Group detailed the potential benefits of developing an e-Infrastructure for data. It can allow different domains to collaborate and enable the use, re-use and combination of data while still maintaining the data's integrity and ownership. But to enable this to happen they recommend the following:

- The EU to should develop a framework for a collaborative data infrastructure.
- Additional funds should be found to develop such an infrastructure.
- Incentives and rewards for data sharing should be introduced, as well as ways to measure data value.
- Researchers should be trained so they recognise the importance of sharing data.
- Create incentives for green technologies within data infrastructures.
- Establish a high-level international group to plan for data infrastructure.

Monica Marinucci, Oracle - *"Technology advancements, networking capabilities and tighter integration of ICT in the scientific process are opening up new frontiers for science and enabling scientists to explore new ways of doing research. It is crucial for the future of science and economy to set up shared research infrastructures and to consolidate basic research services that will allow scientists and their institutions to do better and innovative research in an efficient, scalable and collaborative manner. Industry should support this process by providing cost effective, reliable and open solutions for the whole research lifecycle. Innovation and technology produced then translates into technology progress, jobs and wealth for the society at large."*

Up in the clouds

The VENUS-C project is set to explore how cloud computing can be used in European scientific settings. It will initially look at applications ranging from biomedicine, civil protection and emergencies, civil engineering and data for science, expanding its scope over time.

credit: D4Science

Aquamaps, from the D4Science project, is one such application which hopes to benefit from VENUS-C. Aquamaps uses environmental information, such as sea water salinity and temperature, along with data on fish environments to map the biodiversity of our oceans. This gives the fishery community a way to predict the existence of fish anywhere in the world at any given time. They can also gauge the impact of changing climatic factors as well as of pollution, natural and man-made disasters. By moving these calculations onto the cloud, VENUS-C hopes to provide a quicker way to map species, and ultimately provide scientists and decision makers with more information, faster.

VENUS-C will work with other DCIs in a number of activities. For instance, security and accounting have been identified as two potential fields of cooperation with EMI.

Andrea Manieri, VENUS-C *"Cooperating with other European Distributed Computing Infrastructures offers an important opportunity to share knowledge on an evolving landscape. For VENUS-C, it is also an opportunity to evaluate how we can capitalise on European investments and expertise from a cloud computing perspective. By standing on the shoulders of giants, understanding the driving forces, costs and impact, VENUS-C will also play a part in defining a cloud computing strategy for European scientific communities."*

Neil Geddes, e-IRG delegate for the UK and chair of the Blue Paper editorial board - *"The Blue Paper represents one step on the road towards better exploitation of the opportunities presented by advanced ICT and computing for the European research communities. It acknowledges the importance attached to this area by both ESFRI and the e-IRG and illustrates the growing collaboration between these two groups. Both the e-IRG and ESFRI are prepared to transform the common findings stemming from the Blue Paper into actions supporting the ESFRI Roadmap projects in building the European Research Area."*

Supercomputers for super research

The Partnership for Advanced Computing in Europe (PRACE) provides Europe with world-class high performance computing (HPC) systems and service. Twenty European states are members of the PRACE association which is managed as a single European legal entity. It represents the top of the European HPC ecosystem and works closely with other European e-Infrastructure projects such as DEISA, EGI and HPC-Europa2.

credit: FZJ

PRACE's target is to offer state-of-the-art super-computing systems (Tier-0) to the European scientific communities. In the medium term, each system will provide researchers with several petaflops of computing power and by 2019 the aim is to reach exascale computing. PRACE is well integrated into the European HPC ecosystem and will include services presently offered by DEISA as of 2011.

At the moment PRACE has two Tier-0 machines (JUGENE in Germany and Curie in France).

David O'Callaghan, Trinity College Dublin - *"We need to work together to share requirements and results between projects and in particular to benefit from the strengths and specialities of national and international initiatives. StratusLab will work with EGI and the other distributed computing infrastructure projects to provide users with the mix of cloud, virtualisation and grid technologies they need to do research effectively."*

Powering research

Whether using a single PC, large-scale distributed data processing, or the world's fastest supercomputer, modern researchers invariably require access to computing power. In Europe, two international computing infrastructures – the European Grid Infrastructure (EGI) and the Partnership for Advanced Computing in Europe (PRACE) – are managing the development of grids of high throughput computing (HTC) and high performance computing (HPC) services, respectively.

For HPC to support researchers far into the future, the e-IRG recommends closer collaboration with users to better understand researchers' requirements of HPC and the opportunities it can deliver to them. But another challenge is on its way; in the next decade we're expected to reach exascale computing. This will provide computing a thousand times more powerful than now. But for these machines, simply scaling up existing software is not enough - it will have to be redesigned to work at such fast speeds.

In the case of distributed computing – whether delivered by grids, clouds, volunteer computing or shared data centres - stronger collaborations are needed between grid and cloud infrastructure users and resource providers. Six Distributed Computing Infrastructure (DCI) projects have detailed ways in which they expect to work together over the next five years in the DCI Collaborative Roadmap. Key areas for collaboration include providing each other with dissemination and training, and working together to define standards for interoperability.

Give the people what they want

Users and their needs are key when developing any e-Infrastructure. As a general, but important, message of the e-IRG Blue Paper, the e-IRG recommends that RI, e-Infrastructure and user requirements should evolve in tandem and collaboration between RI and e-Infrastructures at all levels should be actively supported. The annual EGI User Forum, for example, provides an opportunity for users to talk to e-Infrastructure providers.

Researchers are often grouped into Virtual Research Communities (VRCs), or Virtual Organisations (VOs) which rely on ICT to allow a group of geographically dispersed researchers to work together. The e-IRG recommends that VRC developments should proceed gradually, starting with domain-specific shared access to distributed resources, and expanding to integrate different research activities.

For more information:

DCI Collaborative Roadmap: https://documents.egi.eu/document/172

e-IRG Blue Paper: www.e-irg.eu/images/stories/eirg_bluepaper2010_final.pdf

High Level Expert Group report on Scientific Data: http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204

e-IRG: www.e-irg.eu

GÉANT: www.geant.net

PRACE: www.prace-ri.eu

VENUS-C: www.venus-c.eu

EGI: www.egi.eu

iSGTW: www.isgtw.org

e-ScienceTalk : www.e-sciencetalk.org