



# e-ScienceTalk

## 9<sup>TH</sup> E-INFRASTRUCTURE CONCERTATION REPORT

## TABLE OF CONTENTS

<b>1 INTRODUCTION</b> .....	<b>3</b>
<b>2 SCIENTIFIC DATA</b> .....	<b>4</b>
2.1 Overview.....	4
2.2 Discussion/Recommendations.....	4
<b>3 E-SCIENCE ENVIRONMENTS</b> .....	<b>7</b>
3.1 Overview.....	7
3.2 Discussion .....	7
<b>4 E-HPC STATE OF PLAY</b> .....	<b>10</b>
4.1 Overview.....	10
4.2 Discussion .....	10
<b>5 INFRASTRUCTURES FUTURE PROSPECTS UNDER HORIZON 2020</b> .....	<b>12</b>
5.1 Overview.....	12
5.2 Discussion .....	12
<b>6 EXHIBITION OF NEW PROJECTS</b> .....	<b>14</b>
6.1 Scientific Data .....	14
6.2 e-Science Infrastructures .....	14
6.3 Support Actions and NCP .....	14
<b>7 CONFERENCE DISSEMINATION</b> .....	<b>16</b>
<b>8 STATISTICS</b> .....	<b>17</b>
<b>9 CONFERENCE FEEDBACK</b> .....	<b>18</b>
<b>10 CONCLUSIONS</b> .....	<b>19</b>

## 1 INTRODUCTION

The EC's e-Infrastructure Concertation Meetings are a series of events that bring together key players in the community who are working towards a long term sustainable e-Infrastructure for scientific research in Europe. The event described in this document was targeted at the e-Infrastructures' community in the new Distributed Computing Infrastructures era.

The 9th e-Infrastructure Concertation Meeting, organised in collaboration with the e-ScienceTalk project, took place at the Lyon Conference Centre in Lyon, France on the 22<sup>nd</sup> and 23<sup>rd</sup> of September 2011. The programme for the event aimed to further the EC's objective to develop a world-class resource as part of a strategy to achieve European leadership in e-Science. The meeting provided an overview of the EGI community and a review of the community's plans and progress towards the adoption of a federated virtualised infrastructure for the European user community.

The two day meeting brought together 150 representatives from the e-Infrastructure landscape, policy makers and funding agencies, and shone a spotlight on the new e-Infrastructures projects starting in 2011. Subjects included taking stock of progress and discussing future perspectives on data, cloud technologies, HPC, software and e-Science environments, as well as future prospects under the Horizon 2020 funding programme.

This year's meeting attracted an exceptional number of delegates (a third more than last year) and also proved to be a draw for a wider audience through the live web cast. Over the two days, 454 individual visitors viewed the webcast of the proceedings. Delegates commented that the format, content and organisation of the meeting worked particularly well, and it provided an ideal opportunity for networking for the e-Infrastructure community.

## 2 SCIENTIFIC DATA

### 2.1 Overview

The session dedicated to Scientific Data was opened by a presentation on the current status of activity in the field of scientific data one year after the publication of the high level group report "Riding the Wave". This was followed by a presentation on the results of a study on 'Digital Object Identifiers'. A panel discussion followed where the debate focussed on establishing a European Data Infrastructure and a data access interoperability task force. Panellists also considered how to ensure that the areas of open publications, open data, open software and open education are more productively integrated.

### 2.2 Discussion/Recommendations

The first presentation was given by Wouter Los from LifeWatch, who began by highlighting the problems with building a Data Infrastructure e.g. interdisciplinary and national border challenges. The importance of trust and curation was emphasised. With a proper scientific e-Infrastructure, researchers in different domains can collaborate on the same data set, finding new insights and solving grand challenges. Introducing some of the project components of ENVRI, Los described some of the requirements to study complexities (e.g. access to interoperable data and workflows, options for fast computation of the effect of changes in parameter data).

The second session focussed on the findings from a project called DIGOIDUNA which has collected quality information to provide recommendations on the growth of digital identifiers. Digital Identifiers (DIs) are an essential building block for enabling effective and efficient technical solutions for the creation of value-added services. These include data access, search and navigation as well reputation assessment and citation indexes. The benefits are widely recognised by various stakeholders and DIs are key to cost effective data management.

Managing identifiers is recognised as an essential component of present and future scientific data infrastructures. Difficulties arise, however, when dealing with data and information created and managed across national, organisational, disciplinary, cultural and technological boundaries. Unfortunately, there is a different level of maturity between more mature solutions for digital objects and the gradually emerging solutions for authors. Trusted institutions should support the definition of agreements between the relevant stakeholders and users, especially when there are potentially conflicting interests.

Recommendations were made by Paolo Bouquet, who works on DIGOIDUNA. Measures included creating a critical mass of coordinated DI systems, securing sustainability and long term of viability as well as wider demand for usage/exploitation of e-science results. He emphasised it is important to renew commitments from various stakeholders.

Some of the main recommendations are listed below:

- The flexibility of funding sources should be enhanced, allowing the reallocation of funds in the portfolio to enable the rapid scaling of promising solutions that embed or promote the value (usage) of identifiers
- Funding bodies must support the development of collaborative models and actions to create synergies and exchange opportunities between the private/commercial sector and scientific sector (“DI-related PPPs”)
- Provide innovative incentives for data producers in publicly funded projects to either adopt one of the commonly recognised methods for identifying digital objects and authors where they exist, or directly feed into solutions that overcome the obstacles
- Stakeholders should require systems compatible with the open web standards and practices for identifying entities on the web.

The Data Access Interoperability Task Force panel, moderated by Herbert Van de Sompel of the Los Alamos National Laboratory, offered a very interesting occasion to focus on interoperability challenges. The scope of the discussion was around how to achieve interoperability in data access across disciplines, projects, regions and continents. Both technical and organizational aspects were considered.

Two perspectives were heard before the panel discussion from Daan Broeder (DAITF/EUDAT) and Donatella Castelli (CNR-ISTI, Pisa). Daan Broeder’s suggestion was to collect a core group of experts from a few domains, and continue working on preparatory docs and forum discussions. Workshops will start in April/May 2012 as planned by EUDAT and OpenAIREplus. Castelli felt the Taskforce should consist of community practitioners and stakeholders as well as IT researchers and engineers.

Below are some of the suggestions and comments from the panel:

**Andrew Treloar**, Director of Technology for the [Australian National Data Service](#) (ANDS), Australia

- Researchers are all international and this should be recognised in our infrastructure.
- As the public is funding the data, we should build public access into interoperability.
- Let’s be pragmatic and start small. We should meet the needs of some people and build from there.

**David Giaretta**, STFC and Alliance for Permanent Access (APA), UK

- Persistent Identifiers are critical but traditional business models are broken. Pay once and they are out of business but keep paying and they’re not being persistent.
- We need the support from both scientists and funders. We also need to involve the commercial world and government, and examine privacy and access issues.
- The unfamiliar and unknown must be made usable which means researchers need to be able to use the infrastructure without necessarily having knowledge of the software behind it.

**Eloy Rodriguez**, Universidade Do Minho, Portugal

- The technical challenge must be easily addressed.

- Solutions should also offer different levels of interoperability.
- We should reuse what we have and there is no need to 'reinvent the wheel'.

**Matthew Dovey**, Programme Director, Digital Infrastructure, JISC, UK

- Questions that need to be addressed: *How do we motivate researchers to manage their research? What can be delivered through institutions and what at a national level? Will researchers recognise data at citation level? What does an individual researcher need? What training do they need? How will it be administered?*
- We need a longer term solution and should think pragmatically by understanding the context as you collect the data. We need annotation of data in context e.g. column marked T – is it time or temp?

**Tor Bloch**, Consultant at Rimac, Denmark

- There are lots of lessons to be learnt from the IPSOS health project. There is different data access in different countries.

The main summary points from the Q&A are included below:

- We should not forget the human aspect and we need to emphasise this.
- A top down approach meeting a bottom up approach may cause problems and we therefore need to be careful.
- Should we try to make this happen so rapid natural selection can happen?

## 3 E-SCIENCE ENVIRONMENTS

### 3.1 Overview

e-Science environments can allow scientists to connect to their peers, share scientific resources and collaborate on research across organisational boundaries. The EC has supported the development of these through a number of calls, in both virtual organisations and e-science environments. This session looked at the developments thus far as well as best practices for e-science environments through presentations and a panel discussion on the topic.

### 3.2 Discussion

Through Call 1 in 2007 and Call 7 in 2010, a number of projects were funded under the Virtual communities banner. The aims of the Virtual Research Communities calls were to:

- address the specific needs of new scientific communities regardless of the location of their research facilities.
- offer advanced applications and capabilities to more researchers, capturing commonalities, fostering interoperability, promoting open standards and federating approaches across disciplines.
- assist user communities who may require the adaptation of methods and scientific practices to exploit the extended capabilities of the e-Infrastructure.
- incorporate users from academia and industry from one or more scientific or engineering communities, computational scientists and e-infrastructure providers
- provide end-to-end e-infrastructure services and tools
- create user-configured virtual research facilities/test-beds by coalition of existing resources
- create sustainable virtual research communities
- integrate regional e-Infrastructures

The e-science environment call in 2011 featured the following evolved objectives:

- support the development and deployment of e-Science environments based on a seamless and integrated e-Infrastructure (networking, computing and data infrastructures and services)
- foster user oriented services

The main obstacles and questions which arise when developing these e-science environments include:

- Standardisation of gateways between the technical systems and networks underpinning integrated e-Science services.

- Does standardising the scientific process stifle innovation?
- Common models of scientific workflows
- Is there a need for new tools for working with content in collaboration mode?
- How and by whom will the rules/models be established?
- Should scientists use available wikis and social networking or do we need to tailor tools to their specific needs?
- Will scientists be able to openly share their ideas, data and resources?
- How will the notions of research excellence evolve in the e-Science environment?

One vision for the e-Science environment is to create a European Cloud Computing Infrastructure to store and access data, as well as for data processing and analysis. The project would be led by the European Space Agency together with CERN, EMBL, EGI.eu, national space agencies and European big commercial ICT players. According to a Forrester Research report, the global cloud computing market will shift from \$40.7 billion in 2011 to more than \$241 billion by 2020.

This infrastructure aims to be considered as a natural infrastructure for the global science community similar to the road or telecommunication infrastructure for the general public today. By acting as a place in which to store vast quantities of data, along with an unrivalled array of open source tools, and an infinite amount of computing power which is accessible and usable from any kind of computer, smart phone or tablet device, it will enable scientists to mine data in one place and address major challenges.

In contrast to China and the US, industrial cloud policy in Europe is quite heterogeneous. The development and exploitation of a Cloud Computing Infrastructure will initially be based on the needs of European IT-intensive scientific research organisations, while also allowing the inclusion of other stakeholders' needs (governments, businesses and citizens). However due to the scale and complexity of services it will require the collaboration of a variety of service providers. To join this infrastructure, just like any other public infrastructure, a service provider must adhere to internationally recognised policies and quality standards that will be adopted by the governance structure involving all stakeholders.

Initial actions for implementing a European Cloud Computing Infrastructure are to:

- Identify and adopt suitable policies for trust, security and privacy on a European-level.
- Create a light-weight governance structure that involves all stakeholders and can evolve over time as the infrastructure, services and user-base grow.
- Define a short- and mid-term funding scheme base involving the three stakeholder groups (service suppliers, users, EC and national funding agencies) into a Public-Private-Partnership model to implement a Cloud Computing Infrastructure that delivers a sustainable business environment adhering to European-level policies.

To date, initiatives that have used ad-hoc infrastructures of this sort have seen success. For example the GAIA mission has made a 3D map of the Milky Way by using cloud computing. However no telecoms communications company can yet provide the amount of storage and power needed to





support all scientific disciplines that have the potential to use clouds. The 'Science Clouds' initial flagship use cases, hopefully followed by the addition of further user communities, will help to create an incentive for industry to provide the necessary computing power for future e-science.

## 4 E-HPC STATE OF PLAY

### 4.1 Overview

Over the past years the European Commission has invested in HPC infrastructures across Europe particularly through the DEISA and PRACE projects. This session looked at HPC infrastructure as well as software in the HPC community.

### 4.2 Discussion

In the HPC arena PRACE is going well – its resources are six times oversubscribed. Calls to date have addressed issues such as more effective solar cells, biochemistry, fluid dynamics, particle and plasma physics, weather and climate models, material science, and astrophysics.

ICT is acknowledged as a fundamental enabler for research and innovation but, while the HPC market is healthy, the EU is still behind. In the last two years Europe has lost 10% of its HPC capabilities while Asia and the US have increased their capabilities. For example only one machine in the current Top500 is European. The average yearly spend on HPC is around half of the US GDP spending; to stay competitive Europe needs to double its current investment.

There is currently fragmentation of European HPC efforts across many countries. PRACE can help to combine and reinforce the efforts of national and EU funding authorities. There is also a need to develop a new HPC strategy for industry involvement.

Investments in HPC infrastructures require a long-term perspective. For example the transition to petascale and exascale computing can offer new opportunities for both science and computing. The EC also recommends we exploit the innovative potential of HPC services beyond science.

The EU has a lot of talent in HPC; it is also home to a number of successful software firms. The EESI (European Exascale Software Initiative) is a very good example of engagement with industry; a significant amount of important HPC software has also been developed within the EU. However currently only a few codes scale beyond 100,000 cores - this is something that will need to be developed for the future. In Europe it is more likely that efforts in HPC will pay off in the software not the hardware side.

Key policy actions from the European Commission in HPC are:

- Develop EU-level governance
- Spend more (Member States, EU, industry)
- Development of EU native capability through
- Pre-commercial procurement



- Level-playing field for EU supply industry
- Increase HPC use in industry, especially by SMEs
- Share application and software development with global partners

## 5 INFRASTRUCTURES FUTURE PROSPECTS UNDER HORIZON 2020

### 5.1 Overview

This session looked at the future plans of e-science infrastructures and projects under the next Framework Programme (Horizon 2020) which will run from 2014 to 2020. The outcomes of the discussions and presentations are summarised below.

### 5.2 Discussion

e-Infrastructures need greater attention. This can be fostered through for example, better engagement with industry in the hope that this will lead to wider availability and use of sustained standard products that are necessary to reach a critical mass. GÉANT for example aims to improve engagement with users as well as initiating more engagement with industry. The European Grid Infrastructure (EGI) will also ensure that future efforts are highly user driven.

There is a need to close the gap between what the e-infrastructure community is doing and what the computing services organisations in HEIs and research institutes are doing for general research teaching and learning support. On the whole HEIs and other research institutes will not be open to looking after a system developed by an EU-funded project once the project is finished. However if the institute has been involved from the start they will be more likely to take on and use tools and platforms developed within a project, as it has been developed with their needs in mind.

Communities can drive the innovation of development and e-infrastructures can try to surf on these waves. A good example of this is neuGRID and the associated outGRID and N4U projects. Neuroscience is an advanced user community, with state of the art needs. However what started as a vertical impetus within neuroscience is now spreading horizontally into other medical fields. Developers can draw out commonalities from bleeding edge research to support other research areas looking to use e-science tools. By driving common cross technology elements in this way we can achieve economies of scale; this is how we have achieved ubiquitous web access. However while finding commonalities is important, both horizontal and vertical development is needed as communities will sometimes have specific needs unique to them.

Funding remains a problem both when developing e-Science environments and ensuring their sustainability. Over the past years the e-infrastructure community has learnt many lessons; their aim now is for longer-term and more sustainable infrastructures. However in order for e-infrastructures to develop long-term plans, they need access to funds upfront. A move to more individual funding sources could help to increase their sustainability. New tools are also needed to ensure national commitment of resources, financial and others.

Involving industry could be an alternative way to develop sustainable models, however in this case difficulties could arise due to the tension between retaining commercial knowledge and ensuring open access for all.

There are already e-infrastructures that are making impacts on the global level but continued development is difficult. Money is not necessarily the problem here – they also need nurturing over long periods of time.

Connections between researchers and providers are of vital importance. Those in the e-Science community need to build trust between them and their current and potential user communities. For example e-infrastructure providers need to be clearer about the services they provide for researchers. In particular acronyms can act as barriers towards researchers' understanding – providers should strive to be more transparent.

Very large networks for remote sensing (the 'internet of things') were not specifically discussed but should also be recognised in this landscape. In the last call, the EC received no proposals in this sub topic, it seems that is difficult to address this community. Microelectronics often gets forgotten about - this is ICT rather than an e-Infrastructure. However the field of microelectronics is an enabler for HPC, grid and so on. OSIRIS includes this sort of Research Infrastructure in their studies which are often different in approach from other infrastructures.

Software should be seen as an infrastructure alongside computing and data. It is a tool to be used, sustained and supported in its own right. Scalable application development will be important for exascale computing – the focus is moving from an HPC data approach to a software development approach.

Computers do not solve the problems, human beings do. We need to invest in training, a global collaboration in funding of human development. Training Work Packages within projects are a starting point but wider horizontal funding would be beneficial. It would be useful to explore links between ERC and Marie Curie programmes in this respect.

Europe needs to democratise the use of HPC. There are not enough scientists and engineers and the money to fund training is lacking. Open data is tougher to solve on a European basis compared to a national level – and there is a question of who meets the costs.

National agencies are often seen as obstructive because researchers have to get past them to get funded. OSIRIS is collecting information for public authorities on players in the field e.g. governance, policy, operations so that wheels do not have to be reinvented. This information is useful for new e-Infrastructures and for comparison of existing ones. National agencies also get asked for money for the Future & Emerging Technology ICT flagships – calls for funding need to be integrated across EU initiatives. Research Infrastructures should become a key component when making ERC proposals – the best researchers should use the best infrastructures. Smaller countries should also be included but the question remains whether new member states afford to use these expensive infrastructures.

## 6 EXHIBITION OF NEW PROJECTS

Following the 8th e-concertation meeting, organised by the e-ScienceTalk project and held at the Globe in CERN, Geneva in November 2010 it was decided that new e-Infrastructures projects funded under Call-9 would have the opportunity to present themselves via an exhibition rather than in a presentation session. These poster sessions were held during the course of the first day of the event. Projects who were invited to present their work during the exhibition poster session included the following.

### 6.1 *Scientific Data*

- 1 agINFRA
- 2 diXa
- 3 ENGAGE
- 4 ESPAS
- 5 EUDAT
- 6 iMarine
- 7 OpenAIREplus
- 8 PanDataODI
- 9 SCIDIP-ES
- 10 transPLANT
- 11 BioMedBridges
- 12 DASISH
- 13 ENVRI
- 14 SIM4RDM

### 6.2 *e-Science Infrastructures*

- 1 BioVeL
- 2 DRIHM
- 3 EarthServer
- 4 GLORIA
- 5 N4U
- 6 SCI-BUS
- 7 VERCE

### 6.3 *Support Actions and NCP*

- 1 Discover the COSMOS
- 2 ELLA



- 3 FISCAL
- 4 GLOBAL excursion
- 5 ORIENTplus
- 6 Virtual Campus Hub
- 7 EuroRIs-Net+

## 7 CONFERENCE DISSEMINATION

Along with the logistical organisation of these meetings e-ScienceTalk put together a communications strategy to highlight the projects and discussions coming out of the meeting. This included the following:

- The event was publicised on the e-ScienceTalk homepage <http://www.e-sciencetalk.org>.
- The event was webcast online for those who were not invited or could not attend in person by the GRDI2020 project, at <http://www.grdi2020.eu>
- The slides from the meeting were available at <https://www.egi.eu/indico/conferenceTimeTable.py?confId=452#20110919>
- The e-ScienceTalk team, plus colleagues covered the event on our blog, GridCast, through written posts and web videos, [www.gridcast.org](http://www.gridcast.org)
- The tag #concertation was chosen for use on Twitter, so delegates and followers could track news from the conference online, [www.twitter.com/e\\_scitalk](http://www.twitter.com/e_scitalk)



## 8 STATISTICS

The 9<sup>th</sup> e-Infrastructure Concertation meeting was attended in person by 150 delegates. This is an increase of more than a third on last year's attendance of 110 delegates. Attendees included representatives from the EC, as well as those involved in EC funded projects. In addition, a further 305 visitors made use of the live streaming via the web on the first day of the conference. On the second day 149 visitors accessed the live webcast.

This year, there was a great deal of interest from those who could not make the meeting. The first day of live streaming attracted 825 total page views from 36 countries. The meeting had a wide international appeal and the top 10 viewing countries were Italy, Germany, Belgium, Slovenia, Portugal, US, UK, Columbia and the Netherlands. On Friday 23<sup>rd</sup> of September, there were a total of 403 page views from 21 different countries. A number of non-European countries such as Columbia, which was the sixth top connection, were tuning in on the second day of the meeting.

From the 22<sup>nd</sup> to 23<sup>rd</sup> September, there were 42 visits from 42 unique visitors to the e-ScienceTalk website, which contained information on the e-Infrastructure Concertation meeting. Forty nine percent of these visits were from new visitors to the site.

At the co-hosted, EGI Technical Forum, there were 27 blog posts, 9 videos and 10 bloggers. Seven authors blogged at least once. From the 22<sup>nd</sup> to the 23<sup>rd</sup> of September there were a total of 249 visits to the GridCast blog, from 176 unique visitors. On average visitors spent 3.48 minutes on the site when visiting and 51% of these visits were from new visitors. During the e-Concertation meeting, there were 63 tweets on the first day and 34 on the second day. Tweets came from a number of sources including EUIndiaGrid2, Venus-C, SIENA, EUBrazilOpenBio, EUMEDGRID and FP7 Training. During the EGI-TF there were 251 tweets in total.

## 9 CONFERENCE FEEDBACK

In addition to the statistics gathered above the following anecdotal feedback on the 9<sup>th</sup> e-Concertation Meeting was also gathered in an effort to improve upon the meeting next year. This was as follows:

- Participants stated that this event gave them very good networking opportunities allowing them to exchange views and best practice with projects they had never interacted with before.
- Delegates enjoyed the meeting overall, and commented that the venue, catering and access were good in general.
- One recommendation for next year would be for the exhibition for future projects to be in adjoining room or more clearly laid out, perhaps with presentations next to posters.
- Other formats suggested for dissemination included iCalendar (ics) format, poster, screen, phone app, email alert
- Another recommendation was for the e-concertation website to link to all funded projects. A wrap-up session could also have been a useful for summarising key points.
- The dinner at Brasserie Georges was enjoyed by delegates.
- The webcasting of the event also went well. A number of delegates referred their colleagues towards it if they were not present at the meeting.

## 10 CONCLUSIONS

The 9<sup>th</sup> e-Infrastructure Concertation meeting was well-received and attended both in person and online via the live webcast. The event provided an ideal opportunity for e-Infrastructure projects to learn more about each other, to network and build relationships that could lead to future collaborations. The webcast attracted a large degree of interest amongst those who were unable to attend, and ensured that the meetings outcomes were disseminated to a wider audience.

Feedback from the meeting indicates it was useful to have a conference exhibition instead of project presentations to lighten the load of the sessions during the event. However, next year it might be more effective to have projects provide short presentations during the exhibition. The organisation of the event went smoothly and delegates were happy with the venue. e-ScienceTalk will aim to ensure future events have similar opportunities.