

Athens e-IRG Workshop report

9-10 June 2014



Event Summary

Workshop location: Divani Caravel Hotel, Makedonia Room, Mezzanine Level

A total of **71 participants** attended the workshop. The objective of the workshop was to address important topics related to the **e-Infrastructure Commons, Data Management, User Aspects**, and the **Future of e-Infrastructures**. The keynote in the opening session tackled the issue of the e-Infrastructure policy of the European Commission. Four plenary tracks aimed to discuss the following:

1. “Coordination and Collaboration” addressing the implementation of the European e-Infrastructure Commons and the collaboration between ESFRI and e-IRG.
2. “Data aspects” focusing on international cooperation, data management and challenges.
3. “User aspects” addressing a broad range of topics including legal issues, core services, impact of e-Infrastructures, skills and training.
4. “Horizon 2020 Work Programme 2016-2017”, a highly interactive track about the future directions of European e-Infrastructures.

Each track concluded with an expert panel discussion providing ample opportunity for interaction with the audience.

Key Conclusions

The following are the key points raised during the sessions and associated panel discussions (focusing on measures to achieve the following).

DAY 1

- Allow end-users to participate in the governance of RI’s and in particular e-Infrastructures.
- Integrate various sources of reflection: GEANT Expert Group, ASPIRE Study, e-IRG Papers.
- Involve (and then manage) private entities in e-Infrastructure service delivery and provision
- Sustain RI’s to ensure cooperation and synergies
- Governments to create legal frameworks needed to increase public trust

- e-Infrastructure providers to create higher assurance regimes to secure data, developments of AAls and applicable standards
- e-IRG may want to investigate health care data as a case study
- Sustain dialogue among all RI stakeholders
- **Synchronise ESFRI and e-IRG at all levels to identify needs related to the e-Infrastructure ‘Commons’**
- **Sustainability, networking and energy are common problems for the vast majority of e-Infrastructures. All major e-Infrastructures should be making a common case for sustainability. Sustainable results need sustainable organisations.**

DAY 2

- Agree on processes for 1/ disclosing data and 2/licencing - The rights of data owners and producers – they need to be comfortable with data ownership rights.
- Organise trainings on data preservation at the doctoral level and for all data users/producers
- Related stakeholders to work together on some common issues – business/sustainability models, organisational structures and users’ involvement
- Develop business models for sustainability
- Prepare for Big Data technologies

Contents

Event Summary	1
Key Conclusions.....	1
Opening	4
Sverker Holmgren (e-IRG) - Welcome address; goals of the workshop.....	4
Christos Vasilakos, General Secretary Research and Technology - Welcome address by Greek ministry.....	4
Panos Argyrakis, AUTH and GRNET and Yannis Ioannidis ATHENA Research Center - Welcome address by local host.....	4
Anni Hellman (EC/DG CNECT) - e-Infrastructure policy EC	4
Rosette Vandenbroucke (VUB) - Charter of Access	5
Track 1 Coordination and collaboration.....	6
Bob Day (JANET) - e-Infrastructure Commons.....	6
Yannis Ioannidis (ESFRI) – Collaboration ESFRI – e-IRG	7
Panel discussion on e-Infrastructure coordination-collaboration	7
Sergio Andreozzi (EGI).....	7
Giuseppe Fiameni (CINECA/EUDAT).....	8

Françoise Genova (CDS)	8
Sergi Girona (PRACE)	8
Kees Neggers (SURF)	9
Track 2 Data aspects.....	10
Juan Bicarregui (STFC and RDA) – International data collaboration	10
Kevin Ashley/UK Digital Curation Centre - Data: institutional & data centre roles - The DCC experience	10
Dimitar Misev (Jacobs University Bremen) - Big data management. challenges.....	11
Andrew Smith (ELIXIR) - ELIXIR data management challenges	11
Track 3 User aspects.....	12
Andrew Cormack (JANET) – Legal issues.....	12
Christos Kanellopoulos (GRNET) – eidentity services	12
Susan Reilly, LIBER - Data Libraries needs/experiences	13
Colin Wright, CSIR (South Africa) - Thoughts on training / educating of e-Infrastructure users	13
Panel discussion for tracks 2 and 3.	14
Horizon 2020 Work Programme 2016-17	15
Anni Hellman (EC) - Horizon 2020 Work Programme 2016-2017	15
Panel discussion & discussion with audience	16
Closing session	17

Acronyms:

RI – Research Infrastructures

e-RI - e-Infrastructure

KPI – Key Performance Indicator

PPP – Public-Private Partnership

cPPP – Contractual PPP

CoE – Centre of Excellence

PCP – Pre-Commercial Procurement

Opening

Sverker Holmgren (e-IRG) - Welcome address; goals of the workshop

The Chair emphasises the need for a **One-Stop Shop** for all RI/e-Infrastructure users. The current growing population of RI's is mature to undergo some level of integration, identifying commonalities and synergies.

Users are key in the process of RI governance and there is need to determine models for their involvement and levels of responsibility. The RIs should satisfy their needs and letting users participate in the running of an RI will ensure the achievement of this objective.

There are a lot of challenges in Data Management and the data track of this workshop will continue to investigate these issues.

The main objective of the workshop is to send input from this workshop to the EC.

Christos Vasilakos, General Secretary Research and Technology - Welcome address by Greek ministry

The State Secretary for Research welcomes all participants and states that everybody should be very pleased to see e-IRG coming back to its birth place, 11 years after its first workshop in June 2003. In its 11 years of history e-IRG has made significant contributions to the development and advancement of e-Infrastructures through policy making and harmonization activities, starting from the research networking and computing areas, both high-throughput and high-performance computing, and in the last years expanding also to the cloud computing, software and data areas. e-IRG, through its renewed strategy, is also working towards coordination and harmonisation of ICT-based e-Infrastructures, bringing together all these stakeholders and facilitating dialogue towards a single e-Infrastructure "Commons" and a single interface towards researchers. He mentions that he is also very glad to see that ESFRI and e-IRG have been also cooperating strongly together, in particular on the data management aspects, which are vital for both RIs and e-Infrastructures. He also reminds of the first joint ESFRI and e-IRG meeting at Megaron in April, during the week of the very well received International Conference on Research Infrastructures (ICRI 2014). He closes by wishing a fruitful and successful workshop and meeting.

Panos Argyrakis, AUTH and GRNET and Yannis Ioannidis ATHENA Research Center - Welcome address by local host

Panos Argyrakis and Yannis Ioannidis also warmly welcome the participants and briefly introduce the role of their respective institutes, GRNET and the ATHENA Research Center, in the development and coordination of e-Infrastructures (but also for Research Infrastructures for ATHENA RC) nationally (in Greece), regionally (the South Eastern Europe and the Mediterranean region) and also at the European level.

Anni Hellman (EC/DG CNECT) - e-Infrastructure policy EC

The EC's objective is to enable world-class research and innovation through the provision of research infrastructures, for every researcher in Europe – independently from what is possible in a given country. The EC supports "making every researcher in Europe digital", connected and in reach of the best virtual tools, repositories, software and HPC available.

Open access and open publications – including data – is now a policy. The RIs need to adopt a model to tackle the Big Data Challenge, ensuring open access.

e-Infrastructure Commons is an EC priority now as well. The question is what structure (national, regional, thematic, etc.) should be used to enable this concept.

Questions/Comments:

1. **Q:** What is the role of industry? How to integrate industrial solutions and research; is open source/open access a way to involve industry? **A:** This is still open. Industry should be involved in defining the approach. PPP is a model that has raised interest although complex and open access is the way to go.
2. **Q:** Is there a KPI/benchmark to measure the use of the e-Infrastructure Commons, e.g. 50% of researchers to use the Commons? **A:** This is interesting and needs to be discussed further.
3. **Q:** What business model is needed for industry involvement? **A:** Sustainability is now expected off each project. Business models as an issue has been identified. We need to work more together in this area.
4. **Q:** What does it mean to get the private side involved in a PPP? What is the business model? **A:** An example is the cPPP for HPC – it is taking the lead in the provision of policy. This all needs to be tested but this is the way forward.
5. **Q:** Are the e-IRG reports useful for the EC? **A:** Yes – and we should work more together.

Rosette Vandembroucke (VUB) - Charter of Access

The European Charter for Access to Research Infrastructures document is a result of an EC initiative. It defines principles and guidelines to be used as a reference when defining access rules and conditions for access to RIs. The document has no legally binding nature. The work is done in the spirit that no distinction will be made between Research Infrastructures and e-Infrastructures. It defines different access regimes: quality-, quota-, market-, and ubiquitous-based or a combination of the above. Data handling is an issue. The document will be worked on further in the next semester.

Questions/Comments:

1. **C:** e-Infrastructures should be mentioned in the title of the document.
2. **Q:** Is this a public draft? **A:** No, but it can be sent to the workshop participants.

Track 1 Coordination and collaboration

Chair: Vasilis Maglaris (National Technical University of Athens)

The main objective of the session is to explore synergies among major e-Infrastructures providers towards an “e-Infrastructure Commons” to better serve user aspects and in particular the ESFRI project communities. Current developments and ideas/policies for further cooperation will be presented from various key stakeholders.

Vasilis Maglaris stresses the importance of global e-Infrastructures and of innovation and how the e-Infrastructure Commons may be related to industry and PPPs. In addition, sustainability of the Commons and its constituent e-Infrastructures is key and sustainable business models need to be worked out (e.g. charging customers). He invites the speakers to address such issues.

Bob Day (JANET) - e-Infrastructure Commons

The term “Commons” stems from the 18th century when certain pieces of land used to be shared and be available to everyone. The term e-Infrastructure and its components are explained. An Infrastructure is a process-led activity. It comprises many elements. We do not know yet how to develop all of them (although the bottom/basic part is in place). **There is a need to integrate various sources of reflection: GEANT Expert Group, ASPIRE Study, e-IRG White Papers, etc.** And to integrate European-level e-infrastructure services, along with evolving the current governance to enable this. NRENs collective response towards this direction is to integrate DANTE and TERENA into a GEANT association, making a single organisation, more inclusive and with wider remit. Main issues:

1. Users in governance

What role beyond influence and ‘a seat at the table’? Some RIs have complex user involvement mechanisms already.

2. Private providers in service delivery

There is a shift from system integration to brokerage. The current role is more of a market manager rather than provider – managing the market on behalf of the users. Clarity is essential on the **role** of publicly-funded players. How to implement this correctly within the current legal framework?

3. Governments in ensuring persistence-sustainability

There needs to be planning – investment cycle of **10 years** to ensure sustainability. Issues: cost-inefficiency, fragmentation, lack of global reach

4. Information governance – societal impact

Case Study: UK National Health Service. Medical data were made available for research and issues emerged through that. The system was nearly put into place and publication was up to the patient. The whole project is put aside for the time being due to privacy issues and this may be considered as a setback for research. The guiding principles in any case should be **secure data management** and **information used to benefit society.**

Questions/Comments:

1. Application software should not be part of the infrastructure: This should be written by users.
2. Governance of e-Infrastructure: There should be a 'Users' Committee' but not running the show..

Data at hospitals: –There are many questions and open issues in this area such as: is that for health care or research? Anonymity may not be sufficient for health care. Example: MRS scans – how can these be collected by a number of researchers? Who receives the credit for inventions? – **e-IRG should look at health care data as a separate case study.**

Yannis Ioannidis (ESFRI) – Collaboration ESFRI – e-IRG

Yannis Ioannidis gives an overview of ESFRI explaining that it was set up by the EU Council of Research Ministers in 2002, bringing together representatives of ministers of the 28 Member States, 9 Associated States, and of the EC supporting a coherent and strategy-led approach to policy making on Research Infrastructures. Its mandate was to develop a Roadmap with pan-European RIs or major up-upgrades to existing RIs covering the needs of European research communities in the next 10-20 years in all fields of sciences and technologies regardless of possible location. The first Roadmap was published in 2006 and its updates in 2008 and 2010. It included 48 projects in 7 thematic categories with a required investment of 20 Billion Euros and 2 Billion per annum to operate them. A first evaluation of current projects was performed while a new update of the Roadmap is ready to be launched with a plan to be finalised in 2016. The roadmap update process is explained. There are several ESFRI projects which are e-Infrastructures, while there are diverse but in many cases common needs for several ESFRI projects. Then ELIXIR is presented as a case study being a data infrastructure for Europe's life science research sector, explaining its ICT needs.

Questions/Comments:

1. **Q:** Why were e-Infrastructures not included in the previous ESFRI Roadmaps? **A:** PRACE is the only one included in the 2010 Roadmap update, but now ESFRI and e-IRG will be working together and others may be included.
2. **Important to synchronise ESFRI and e-IRG on all levels.** We need to break away from the silos mentality.

Panel discussion on e-Infrastructure coordination-collaboration

Sergio Andreozzi (EGI)

- Sergio Andreozzi proposes three main actions required: 1/ a single entry point for all e-Infrastructures including a common service catalogue, 2/ then awareness raising and 3/ joint promotion to users (such as the ESFRI projects) of this single entry point is required.
- Resource allocation per type of customer and coordination of support mechanisms. Business models for the long-term sustainability of the e-Infrastructure services. Funding is expected from both member states and the EC; member states at the national level, EC funds at the trans-national and EU levels providing added value. In both cases long-term thinking is needed.

Giuseppe Fiameni (CINECA/EUDAT)

Giuseppe Fiameni underlines the following points:

- Raising awareness on e-Infrastructure services is one of the objectives. Communities do not always know what e-Infrastructures really provide and how to access their services.
- e-Infrastructures integration is complex but effort is needed to make services interoperable. EGI, PRACE and EUDAT among others have already started collaborating to ensure the interoperability of their services. The existence of different resources allocation models do not facilitate cross-utilisation scenarios. Cross-utilisation pilots of e-Infrastructures need to be launched for the communities to shape and harmonise the services.
- The involvement of user communities on such pilot services and the overall deployment of services by communities should be encouraged.

A single entity to access them all is indeed required.

Françoise Genova (CDS)

Françoise Genova explains that the user point of view should be taken into account. She initially explains that there can be a gap to bridge between the actual RI facility users and the wider community of data users. Data users want to discover, access, use and compare/integrate data and this requires that data is in the system, it is available, discoverable, comparable to other data. In users may want to perform computations on the data. These points are then further analysed and the following points are made:

- ESFRI projects have to setup data archives and funding is required (that may compete with other RI-specific funding).
- An open data policy is required to promote reuse and competition possibly with an embargo period as some users do not want to share data. Funding agencies have a role to play in the above issues.
- Agreed and as much as possible common data formats and descriptions are needed.
- A registry of resources is required along with mechanisms to discover and compare data with other sources.

Overall, ESFRI projects have a role to play as major data providers, to initiate discussions with the communities or to be major actors in the discussion organised at community level on formats, semantics, registries. Most likely it won't be a single system in the end, but the use of generic building blocks when possible should be encouraged (such as EUDAT or RDA). There is need for dialogue to create the e-Infrastructure 'Commons' set of integrated services.

Sergi Girona (PRACE)

Sergi Girona explains that there is indeed a need to identify common needs among the various required services but also in other areas such as training, although there are not huge commonalities. **Sustainability, networking and energy are all common problems.** What is the reason for e-Infrastructure coordination and cooperation when RIs tend to create their own e-Infrastructures? **All e-Infrastructures should be making a common case for sustainability. Sustainable results need sustainable organisations.** HPC cPPP to support technology and

users. So we need to include together the ESFRI projects and the other users together in the discussion for e-Infrastructure Commons.

Kees Neggers (SURF)

Kees Neggers stresses the fact that e-Infrastructures face more organisational rather than technical problems. To make progress and create the common layers available to all users, we need to work on better organisational structures. The time factor should not be neglected, as otherwise the big users may create their own e-Infrastructures.

Questions/Comments:

- **The issue of Trust** – people trust what has been built to be trusted and a recognised name (e.g. Google, Microsoft and Verisign) always helps. Besides, their services are available and usable.
- **Commons** – by using generic blocks we may be abandoning what we had in the past? Or should we allow free evolution? We should use the building blocks which are evident along with common sense without imposing any rigid structures. However, solutions should be adopted only if they are provided through sustainable organisations.
- **Scalability** is difficult to ensure – there are common mechanisms that can be used.
- **Sustainability** – are we going to charge for RI services?
- **Science management** should be a profession – it is underestimated now.

Conclusions

Vasilis Maglaris summarised the main outcomes of the session:

1. **Human networking and dialogue among communities** is needed.
2. Emergence of cloud and virtualisation technologies and services can help create a **common horizontal layer** acting as a substrate (the bottom-base layer)
3. **Peculiarities need to be respected**, not enforcing solutions or creating state monopolies. Open standardised environments to maximise synergies should be encouraged.

Track 2 Data aspects

Chair: Yannis Ioannidis (University of Athens and ATHENA Research Center)

The main objective of the session is to present data management challenges along with possible best practices to confront such challenges, always keeping in mind the user perspective and needs. Activities from RDA, the UK Digital Curation Centre, the ELIXIR ESFRI project and input from Computer Science Researchers that have been dealing with such challenges for many years will be presented. Finally cost aspects of data management will also be touched.

Juan Bicarregui (STFC and RDA) – International data collaboration

Recent policy statements from the G8, EC, US and others, recognise the growing importance of a strategic approach to the open research data management. These policies have articulated a clear and consistent vision of research data managed as an asset for the public good, accelerating innovation and enabling a new paradigm of data-driven science. The policies highlight the importance of ensuring that research data is discoverable, accessible, assessable, understandable, useable, and wherever possible interoperable to specific quality standards. He reviews some of these policy initiatives and describes how the Research Data Alliance (RDA) is aiming to build some of the social and technical bridges enabling their implementation connecting data and people. The RDA vision is that researchers around the world are sharing and using research data without barriers. The RDA mission is thus to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonization, and discoverability. This will be made possible through the development and adoption of infrastructure, policy, practice, standards, and other deliverables.

Questions/Comments:

- **Q:** Would you please give us some examples of Working Groups (WGs)? Is there any Working Group results? **A:** WG on Metadata, Practical Policies, Persistent Identifiers, Foundations, etc. There are no concrete outputs yet – RDA is a very young organisation. Some first ones are expected later this year.

Kevin Ashley/UK Digital Curation Centre - Data: institutional & data centre roles - The DCC experience

The respective roles of research data centres and universities (not just university libraries) in enabling data reuse is explained and what other pieces of infrastructure are necessary to support other parts of the research data lifecycle. The goal is to make data reusable. Some key messages included:

- Often data can tell stories that publications do not.
- Discipline-bounded data discovery doesn't give us all we need or want.
- Not all data comes from other researchers.
- One person's noise may be another person's signal!

But why care to make data reusable? First of all data is expensive and with investments economies of scale need to be made. With reuse more research, teaching & learning, and

planning can be performed. That is why now funders require data preservation and appropriate data management plans. Some institutional roles include leadership (coordination), audit (who has what), preservation, citability, data/publication linking, education. In addition, according to several studies the Return on Investment of Data Centres is huge (between 400-1200%). Decisions on what data to keep however are also needed, as cost of data centres is always an issue. Costs aspects have been dealt with by the 4C Project. Fundamental question is 'who pays' but answering to 'what does it cost' is also important. Data preservation should be managed internally as it might be more expensive to manage it externally. **Data created through public money belongs to all of us.**

Questions/Comments:

- **Q:** Should every country have something like a national DCC? **A:** Some high level of coordination is needed in every state, although in some smaller states the collaborative way may work without a national structure.

Dimitar Misev (Jacobs University Bremen) - Big data management challenges

Volume, Variety, Velocity and Veracity, are all important properties of Big Data, and lately even more are appearing, like Validity, Volatility, and yet others. Typically the focus is on the Volume of data, while more focus in this presentation is given on the Variety facet, as perhaps an even bigger challenge. So data comes in a various forms and types, and a big problem is that we are not aware of the importance and need for equally extensive support of each of them; multidimensional arrays and coverage data in general, are one of those data types lacking a much needed attention. Clean slate approaches, therefore, are needed for storage, access, and analytics on such multi-dimensional array data.

Andrew Smith (ELIXIR) - ELIXIR data management challenges

New technologies are generating vast amounts of data at unprecedented rate. These data are heterogeneous; they come from multiple sites, relate to different biological functions and often have specific legal and ethical concerns. Integration between these resources is critical and a Pan-European solution is required as no one institute or country can solve the challenge alone. Close collaboration with e-infrastructures is necessary in order to run these services. An example of data use is given on cancer and food shortages. ELIXIR receives data from large European projects. Data is generated all across Europe – users are global. Open Access is an opportunity. Data volume, integration and privacy are the challenges. Integration of data tools is key. Training component is also important. Finally, various countries can still adopt various approaches.

Questions/Comments:

- Data is open access and free for science and industry. Partnerships are created to leverage resources.
- **Q:** Is there any work being done on explaining the value added/KPIs to the citizen? **A:** This will be part of the project (through Working Groups). ELIXIR brings together resources that are already there.
- ELIXIR nodes are in 1/3 of the cases integrated in EGI or PRACE nodes

- Users and nodes are represented in Governance
- **Q:** Open Access to publications and data – will there be a 1-year embargo on data reuse? **A:** We need to convince people that it is against their own interest. Sometimes this does not depend from the researchers. Data should be published on paper publication.

Track 3 User aspects

Chair: Andrew Cormack (JANET)

The main objective of the session is to present a series of e-Infrastructure users aspects, both by the users' and by the providers' side. Topics include legal issues, authentication (e.g. eduGAIN), PIDs, training and education, and data aspects/needs by research libraries. There is a need for user requirements to be taken into account by e-Infrastructure providers swiftly and that there is more e-Infrastructures coordination in favour of the users (i.e. that users easily find their way to the different services offered by e-Infrastructure providers).

Andrew Cormack (JANET) – Legal issues

Updates on legal aspects after the publication of the e-IRG Task Force on Legal Issues report are presented. Changes to relevant legislation in the past year include among others the Parliament's version of the Data Protection Regulation (changes to research provisions), their version of the NIS Directive (which make it less likely that e-Infrastructures are considered as 'market operators'), and the Commission's new Research, Development & Innovation (R&D&I) State Aid proposals (which are intended to help research collaborations).

Procurement presents issues with Public Private Partnership (PPP) definition and the use of PreCommercial Procurement. Network Regulation is required, while Access Policies and Software licences should be also addressed.

Questions/Comments:

- **Q:** What is the difference between RI and e-Infrastructure in this framework?
A: The e-Infrastructure has the networking bit

Christos Kanellopoulos (GRNET) – eIdentity services

One of the visions of the European Commission for the Horizon 2020 funding period is that all researchers should be able to access the necessary compute and data services via the use of a single electronic identity. GÉANT and the eduGAIN service have been in the forefront interconnecting federations to link services and users worldwide. Recent developments in eduGAIN are explained, along with their relevance to the Scientific Communities and the future that lies ahead. Identity federations should thus be interconnected (inter-federated), providing legal and technical frameworks to make them interoperable. As of this May all production Federations worldwide are members of eduGAIN and GÉANT has established active collaborations with the research communities. The time is right in order to proceed together towards the future.

(DAY 2)

Susan Reilly, LIBER - Data Libraries needs/experiences

LIBER is the Association of European Research Libraries, while LERU is the League of European Research Universities. The LERU roadmap for research data and its implications for institutions and libraries is introduced. The roadmap promotes open by default and identifies 6 key area for action. The area that was highlighted as crucial is advocacy and engagement with researchers. Institutions and libraries are adapting to funder mandates on data management, but also recognising that research data is a valuable institutional asset, which can be realised through good data management. Coherent policies and defined roles and responsibilities are essential going forward. More work needs to be done to define what data should be made available, especially from a disciplinary perspective. No one size fits all and this needs to be reflected in policy, infrastructure and support services offered. Finally organisations need to know the cost of storing data, while a Working Group on research data management has produced 10 recommendations.

Colin Wright, CSIR (South Africa) - Thoughts on training / educating of e-Infrastructure users

A brief overview of own experiences in teaching Computational Mathematics and Computational Sciences is given, including institutional level challenges. As a consequence of a recent review of South African e-Infrastructures, a project is underway to establish a national Tier 1 e-Infrastructure organisation as a coherent entity. This will include HPC, NREN, Data and also a Training and Skills development initiative. Insights arising from a survey of some international initiatives re. training e-Infrastructure users across the whole e-Infrastructure ecosystem include:

- Develop skills at multiple levels: e.g. training should be present in doctoral and student training programmes.
- Training module scope to include computational science, numerical algorithms, parallel programming, cloud computing, data-centric computing, computer animation / graphics. Data management needs particular attention in view of data deluge including analysis and visualisation, curation and long term preservation.
- Commerce & industry training modules on “commercial basis”.
- Domain specific workshops in partnership with research community.
- Ad hoc and programmatic interventions.
- The Silos mentality should be removed
- The South African response to the challenge of training the spectrum of e-Infrastructure participants by establishing a cross-cutting Skills and Training is given.

Questions/Comments:

- **Q:** What does the NICIS new organisation aim to provide? Does it include training?
A: First of all it aims to provide integrated e-Infrastructure services. We target universities mainly – to build coherence and get buy-in from the community. And then creating new skills and training.

Panel discussion for tracks 2 and 3.

Andrew Cormack (JANET-chair), Panellists: Juan Bicarregui (STFC-RDA), Kevin Ashley (UK DCC), Christos Kanellopoulos (GRNET), Susan Reilly (LIBER -LERU), Hans Pfeiffenberger (ScienceEurope)

The Chair introduces the panel stating that a lot has been done in several areas and asks what is still missing and what the biggest barrier in each of the areas of the panellists would be.

Several points are made:

- **Open should be the default on access to data** and should be eventually be dropped! Simple and universal solutions are required to resolve the main challenges such as the commercial use of data.
- Services are not always visible to the users. **A virtual marketplace for services for researchers needs to be developed**, which will make life for users much easier. Furthermore, harmonised identity inter-federation is key to make the services accessible and EduGAIN brings new potential.
- **Training for early-career researchers is needed.** A new generation of researchers with new expectations needs to be created.
- **Compliance mechanisms and standards are needed, but also the right attitude.**
- The aim should not be openness, rather the means. Current practices (such as privacy, embargo periods, etc.) reinforce closing data. The important is to **give recognition for the ones who share data**. We need a flexible model for that and training and time are needed.
- The barrier to get extra value out of data is the Intellectual Property/copyrights. The copyright law was not designed for the digital ERA. The G8 principles are positive but the database directive does exactly the opposite. **Until we have suitable legal directives, bottom up won't work.** Policy makers need to help to align everything. As an answer it is stated that Creative Commons license can be used to circumvent the database rights.
- **Creative commons licence – as long as you give me credit, I am happy.** We need to be comfortable with our rights.
- **Embargo periods should be considered as a transition technology.** New generation of users need systems easy to use with new tools, performing without a flaw, and they don't need to know much about it. Another panellist says that embargo period works for some disciplines.
- **Scientific groups do not know what services are available and they write their own code.** If they knew they would have reused tools and services available, as well as whole applications.
- Data description – should librarians do that? Should they differentiate into various domains? **Training is required, evolving librarians towards disciplinary specialisation.**
- **Libraries may also be specialising in domains** but also collaborating among them.
- It will take time to go open and digital. **The more disciplines use this approach (open and digital), the more we can disseminate and persuade others to adopt it.**
- **Champions pushing agendas are also necessary.**

- In USA, there is a new NSF rule that in order to get funding, one needs to get products (patents and the like) not papers. Also publications, data and software need to be citable to get funding for the next project. Evaluators need to know what an important product is. EU is also working on data management policies plans but resources are needed to enforce it.
- One challenge that has not been tackled yet is an **IT environment for accessing protected data**. It requires rights information but also organisational provisions. In other words managing complex rights and coding the IP rights workflow.

Horizon 2020 Work Programme 2016-17

Chair: Sverker Holmgren, e-IRG Chair

The main objective of this session is to present the current plans for the next e-Infrastructure Work Programme under Horizon 2020 (WP 2016-2017) and to get feedback from the audience (in an open consultation mode). Input from the previous sessions will be used.

Anni Hellman (EC) - Horizon 2020 Work Programme 2016-2017

Anni Hellman explains the process for coming up with the next Work Programme 2016-2017 and the involved stakeholders. These include Advisory Groups of experts, Programme Committee of Member States national delegates, and other groups including e-IRG and ESFRI. A harmonised set of questions address all Advisory Groups and all stakeholder. Advisory Groups are expected to provide input by end of June 2014. The vision for e-Infrastructure in 2020 includes also specific steps that need to be included in WP2016-2017. Main topics include the focus areas, the main target priorities, what is missing in the previous WP, and what should continue and what not. Some of the key questions and the answers from the Advisory Group on RIs including e-Infrastructures are the following:

Biggest Challenges:

- Integration (e-Infrastructure Commons)
- Flexible long-term funding solutions
- Business models
- Big Data challenge and related data initiatives such as RDA
- Open Data and Open Access
- Skills

Key assumptions (driving future developments)

- Fierce competition
- Science more multidisciplinary and data intensive
- Prioritisation – not all can be publicly funded.

Output and Impact, Success metrics

- 1st class sustainable RIs and services open to industry and research, public, policy makers

- A less fragmented funding model
- VREs – interoperable, open and innovative infrastructures for better collaboration and intercultural communication in EU and worldwide

Bottlenecks, risks and uncertainties

- Lack of skills, culture of appreciation of data scientists
- Short term politically popular winning over long term perspective
- Unclear funding synergies, funding constraints
- Governance structures for RIs
- Legal barriers

Gaps and game changers

- Changes in political priorities
- EU-wide legislation harmonisation
- Global harmonisation and standards

Questions/Comments:

- Lack of long term funding by the EC is an issue
- Resolve the identity of the person accessing the data. On the other hand, it may or shall not be needed to know who access the data (anonymity).

Panel discussion & discussion with audience

Anni Hellman (EC), Sverker Holmgren (e-IRG Chair), Yannis Ioannidis (ESFRI), Arjen Van Rijn (e-IRG WG on e-Infrastructure Commons)

The e-IRG Chair introduces the e-Infrastructure Commons as described in the e-IRG papers, namely the e-IRG Roadmap and White Paper and then poses some questions.

- What approach is needed? National, thematic or technological?
 - All three are probably needed, but there is a priority order. And the challenges are mainly organizational.
 - A one-stop-shop for e-Infrastructure services is needed.
- And then how can Horizon 2020 support the establishment of the Commons?
 - Coordination needs – e.g. business models, involving users in governance, DANTE/GEANT are starting on that now with the GEANT association.
 - Is the network component supposed to govern all the e-Infrastructures?
 - We need to get the national e-Infrastructures ready for this fight

The objective of the Commons is to realise an ecosystem of public and commercial services, where users can easily find the service they want and providers can innovate. This will be achieved through a joint strategic effort between users and primary strategic actors and suppliers. The issues that require attention and related recommendations intended for the next Work Programme of Horizon 2020 (2016-2017) include:

- Ease of use between national and European institutions
 - Stimulate service portfolio harmonisation
- e-Infrastructure aims to be a general purpose infrastructure
 - Encourage innovation and development of generic reusable tools and services (common AAI, PIDs, service registry and discovery)
- e-Infrastructures need well organized user communities
 - Encourage communities such as ESFRI to organise and formulate their requirements, and then eventually position them to pay for their services
- Different levels of maturity
 - Encourage the development of strong national e-infra organisations as building blocks for European coordination
- European research communities require European e-Infra coordination
 - Encourage the development of a European coordinating body for all e-infra components in which these issues are solved coherently

The main challenges are then discussed.

Main challenges-issues:

- Big Data related.
- Skills needed - Making data science a profession. Cross-domain skills (mainly IT + thematic). Bioinformatics a best practice.
- Business models/long term sustainability
- Adopt some services from industry (such as dropbox) and not reinvent the wheel. Negotiate with scale with commercial providers.
- Membership – should that be as inclusive as possible (RDA)
- Centres of Excellence – we need to look at that adopting a thematic approach

Closing session

Wrap up and conclusions- Sverker Holmgren

The e-IRG chair closes the session expressing his satisfaction about the outcome, stating that there were very good presentations, very active discussions and concrete questions, focusing in the issues and ideas for their solutions.

Final words - Sverker Holmgren, Panos Argyrakis, Enzo Valente

Enzo Valente invites all speakers to the next workshop that will take place in the week of the 10th of November and Panos Argyrakis as the local host thanks everybody for participating and for their active involvement.

Marcin Ostasz-BSC, Fotis Karayannis-NWO, Final version, 3 November 2014