

## e-IRG Workshop 4-5 November 2013, Vilnius



64 participants attended the e-IRG workshop on 4-5 November 2013 in Vilnius during the Lithuanian EU Presidency. 'Infrastructures in the Digital Age: the Data Challenge' and 'e-Infrastructure commons' were the two main themes that have been discussed during the workshop.

### Track 1: Implementation of the e-Infrastructure commons

#### e-Infrastructure Commons 2020: Integrated services via interoperable infrastructures

In the first track 'Infrastructures in the Digital Age: the Data Challenge' **Sverker Holmgren** and **Arjen van Rijn** introduced the main concern of e-IRG's White Paper 2013; a single e-Infrastructure for knowledge, science and innovation. A commons is understood as resources including management system and governance that are accessible to all members of a community. The properties of the proposed e-Infrastructure Commons are openness, accessibility and the adaptability to change. The key elements of such e-Infrastructure Commons are high-quality services that are well managed and seamlessly integrated from the users' perspective. These services should be flexible, efficient and able to change dynamically to meet future requirements.

The users' inability to differentiate and thus to decide what type of resources, and which type of service to use for what purpose, feeds the need for integrated service. Furthermore the uncertainty about the long-term availability of services and resources are recognised as obstacle to use e-Infrastructures. *e-IRG has identified the insufficient collaboration, coordination and integration of these services as the main issues, which are reason for the difficulties for researchers to navigate in the field of e-Infrastructures.* According to **Sverker Holmgren** the e-IRG Workshop in Vilnius should be used as first step towards a joint, integrated setting. e-IRG acknowledges that the integration of infrastructures across countries borders have to face many issues, legal concerns being one category of these issues. The *e-IRG Task Force on Legal Issues* has published a report, which presents small steps to get a solution in an efficient way.

To realize an e-Infrastructure Commons e-IRG sees the necessity to implement three distinct core functions via a common strategic effort between users, primary strategic actors and providers:

- High-level community building, as a strategic effort with a single organization proposed that is in a central role.
- Service provision layer, where production-level services are provided in a flexible, open and competitive approach.

[www.e-irg.eu](http://www.e-irg.eu)

**e-IRG**  
e-Infrastructure  
Reflection Group

e-IRG Secretariat  
P.O. Box 93575  
NL-2509 AN The Hague  
The Netherlands  
Phone: +31 70 344 0526  
secretariat [at] e-irg.eu

Visiting address  
c/o The Netherlands Organisation for Scientific Research (NWO)  
Java Building, Laan van Nieuw Oost-Indië 300  
NL-2593 CE The Hague  
The Netherlands

- Innovation, which is needed to keep the system alive. There is a major role for funding agencies to support innovation.

To enable the users to fulfil their duties, they need to be strengthened strategically and learn to drive their long-term strategy for their infrastructure requirements.

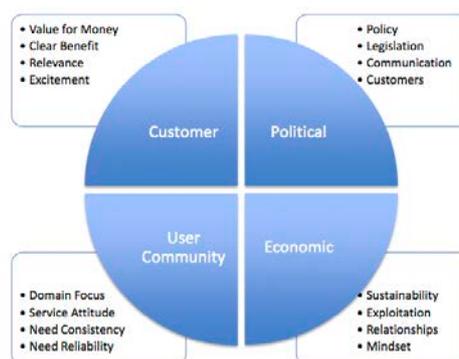
The national governments should provide a basic funding-level for national e-Infrastructures, particularly to get these continuously innovated and should remove regulatory constraints especially to allow private research parties to make use of public funded e-infrastructures.

The European Commission should empower and fund the European user communities as the ESFRI projects to influence the development of the infrastructures, enable and promote the use of structural funds for e-infrastructure development, provide input for strategy setting, harmonization and provide clear guidelines for regulations.

**Arjen van Rijn** further analysed the situation in the Netherlands and explained how the aforementioned criteria are met in the Dutch context by SURF. SURF is the Dutch e-Infrastructure umbrella organisation that coordinates all components of the e-Infrastructure, which consists of capability computing, capacity computing, storage, networking and an eScience centre. In his analysis Arjen van Rijn came to the conclusion that not all services are fully integrated in the e-Infrastructures context, but that some examples exist where researchers use several connected services to accomplish their tasks. But what is still missing is the clear implementation of the user community involvement, business models and pricing of services.

### e-Infrastructure: A view from the edge

**Steven Robertshaw** introduced the PUCE model to reflect upon the given technology Distributed Computing Infrastructures (DCI) in the four different contexts of **P**olitic, **U**ser community, **C**ustomer and **E**conomy, whereas social aspects are considered in the economic part of the model as social capital and human capital. All four contexts have an impact on DCI and the DCI itself has also an impact on the four contexts, so that implies a two-way conversation or two-way flows and exchange of ideas at least between the DCI communities and the people present in those different contexts. At the moment there is no one DCI community; there are several small communities, but speaking with one voice is necessary to gain influence and to get impact with the goal to harmonize the legislation and influence funding streams at the European level.



In the economic context a DCI can be seen as an entity in an ecosystem, whose sustainability can be guaranteed if the DCI engages with the rest of the ecosystem and provides value. To “survive” in this ecosystem a DCI should gain relevance by services and particularly those kinds of services

that the people who are going to use the infrastructure actually want. DCI providers have to provide services that help the communities to solve their problems. A DCI European Research Infrastructure Consortium (ERIC legal entity) will be a good structure to wrap up all activities and provide a unified interface to the user communities.

Business models and the development of a coherent community strategy require additional and complementary skills, which need to be brought into the community in order for the DCI, whether ERIC or not, to continue to survive in the long-term.

**Steven Robertshaw** separated user communities and customers into two different contexts. The users of a DCI are the research communities, which use the infrastructures to do their research, but actually do not pay for the usage. The customers on the other hand are the taxpayers, which are really quite interested in research and research results. To reach the customers through the media, significant, exciting events and narratives are necessary. Engaging with the customers provides benefits in the political context for the DCI community.

**Steven Robertshaw** summarized that the DCI communities should work together to develop a critical mass (an ERIC seems to provide a reasonable structure). It will be necessary to develop strategic plans and then communicate them. The goal must be a single focal point for end-users and also suppliers.

A research infrastructure is an ideal test-bed for new, advanced technologies, so research results should be integrated at a prototype-level before they move on into more market-orientated commercial activities. This will have direct relevance to the other objectives in the area of Horizon 2020, which is to improve competitiveness of Europe.

## **Digital Cultural Heritage moving towards an e-Infrastructure-based approach to digital preservation**

**Antonella Fresa** explained that in the Digital Cultural Heritage and Humanities (DCH) preservation is at very high priority, because the investments in the production in the digital culture and heritage are very high. In order to produce data for research in digital culture and heritage, comprehensive human intervention is necessary to produce the metadata that are associated with the digital content. In the DCH it is estimated that the total costs of digitizing the collections of the European museums, archives and libraries including the audio-visual material is in the range of €100bn, which is €10bn per year if we consider investments over the next 10 years. And the costs of preserving and providing access to this material over a 10-year period after digitization would be in the order of 10bn to €25bn. These huge costs for the DCH sector explains the needs for an implementation of a permanent identification of digital cultural objects and the issue of governance and management of these repositories. Further requirements of the cultural heritage sector are:

- distributed, safe storage and preservation ranging from short to medium and longer-term
- seamless powerful access to and a very high level of trust in the infrastructure

[www.e-irg.eu](http://www.e-irg.eu)

e-IRG secretariat  
c/o Netherlands National Computer Facilities Foundation (NCF)  
P.O. Box 93575  
NL-2509 AN The Hague, The Netherlands  
Phone: +31 (0)70 344 0526  
secretariat@e-irg.eu

Visiting address  
Anna van Saksenlaan 51  
NL-2593 HW The Hague  
The Netherlands

**e-IRG**  
e-Infrastructure  
Reflection Group

- semantic search and the search within the complex data
- aggregation and interoperability of existing features and services

**Antonella Fresa** expressed the needs for a European-wide harmonisation of legislations in the DCH community. An example for the existing diversities is the different authorities in the German *Länder*, the Belgium communities and the centralized organisation of a cultural heritage in Italy and France.

**Antonella Fresa** presented the *Digital Cultural Heritage Roadmap for Preservation project* (DCH-RP), which is a small coordination action. The first outcome of this project is a registry of services that has been implemented and published online, with the cooperation and under the coordination of the Poznan Supercomputing and Networking Center. A first proof of concept was run under the coordination of EGI. Insights from this project are:

- Cultural data are curated by many different persons and therefore data management, data administration, user access control are fundamental issues.
- Security of data is very important for cultural institutions. Thus trust building becomes a key factor in particular when it is not determined where data is stored.

The DCH-RP project is just one step that belongs to a longer process. In the last decades cultural institutions started to move intensively their content on the digital world and this has implied to create knowledge about new instruments, new rules of access, new standards to exhibit digital heritage, new ways of communication. This is particularly important, because the concept of open science has not yet arrived fully in the cultural heritage sector. **Antonella Fresa** concluded that there are obstacles in the DCH sector that must be taken into account in order to speed up processes and the use of the e-Infrastructures is a pillar in the direction of speeding up these processes. And the cultural heritage sector should progress towards its full integration in the concept of open science.

### The e-Infrastructure needs of ESFRI projects

**Yannis Ioannidis** reported about ESFRI, which supports a strategy for building and operating research infrastructures across Europe and aims to capture the needs of European research for the long-term (10-20 years and possibly beyond) and covers all fields of science and technology. The projects in the ESFRI roadmap 2010 are from seven categories:

- social sciences and humanities
- life sciences
- environmental sciences
- energy
- materials and analytical sciences
- physics and astronomy

[www.e-irg.eu](http://www.e-irg.eu)

- e-infrastructures

Yannis Ioannidis explained that many of the research infrastructures in these categories are e-Infrastructures (e.g. CLARIN, DARIAH or ELIXIR) and underlined the importance of the data driven developments like persistent storage, preservation of data and data management services for ESFRI and the need for standards and persistent IDs to enable interoperability. Other issues to ESFRI are the human factor, training and the global scope of research.

Then Yannis Ioannidis described how data-driven research transformed the way how research is done; in the past scientists form a hypothesis, then they set up an experiment and collect the data to analyse it and see if the theory holds or not. Ioannidis states that the advent of research infrastructures allows the collection of data without having a hypothesis and that the analysis of the data determines what kind of hypothesis could be true or not.

According to Ioannidis the ESFRI research infrastructures are not only data producers but also consumers of Big Data and he concluded that it is not possible to move the data to the computation, because it does not scale and therefore the computation and the analysis should take place in the research infrastructures.

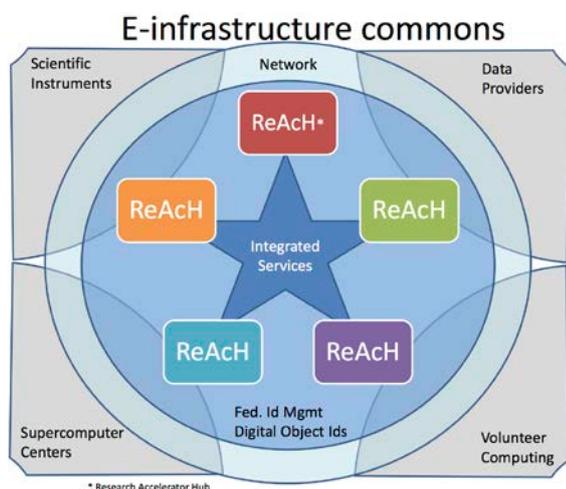
## A Vision for European e-Infrastructure for the 21st Century

**Bob Jones** presented EIROforum's vision of how e-Infrastructures could be used and evolve in the future. The EIROforum is the grouping of eight intergovernmental scientific research organisations in Europe (CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, XFEL and ILL <http://www.eiroforum.org/about/organisations/index.html>).

The key idea is to provide a sustainable research infrastructure by bringing together publicly funded infrastructures and commercial partners to work together into a hybrid-model. The focus of the participating centres called Research Accelerator Hubs (REACH) will be to provide IT-services to the user communities, but also other things like consultancy and training. The first of these REACHs are going to production in 2014 as prototypes. Bob Jones presented examples for these REACHs from EMBL, a number of tier 3 data centres from the life sciences domain at

European Bioinformatics Institute (EBI) in the UK, and at CERN, the 3.5 Megawatt data centre in Budapest.

Bob Jones said that every REACH would offer different services; tailored to the specific communities they are working with. So that each REACH has its own portfolio of services and its own funding model. He also stated that REACHs will need a long-term funding commitment, and they will provide their services as a pay-per-usage model. By using a



[www.e-irg.eu](http://www.e-irg.eu)

**e-IRG secretariat**  
 c/o Netherlands National Computer Facilities Foundation (NCF)  
 P.O. Box 93575  
 NL-2509 AN The Hague, The Netherlands  
 Phone: +31 (0)70 344 0526  
 secretariat@e-irg.eu

**Visiting address**  
 Anna van Saksenlaan 51  
 NL-2593 HW The Hague  
 The Netherlands

**e-IRG**  
 e-Infrastructure  
 Reflection Group

pay-per-usage model the funding is linked to the usage, so that the funding agencies can immediately see the impact of a service and hence they have a justification for their investment. Bob Jones further explained that a user forum, bringing together the organizations and the projects that operated on international level would provide a platform to present the common needs from user communities and research infrastructures to policy makers and e-infrastructure providers. The proposed user forum will be the first user forum, which will really “have teeth”, and will be different to existing models where the users are consulted “as advisory boards in the e-Infrastructure model but are not necessarily being part of the governance of it”.

### Panel Discussion on Implementation of the e-Infrastructure Commons

**Kees Neggers** chaired the panel discussion, the panellists were **Antonella Fresa (Promoter Srl)**, **Steven Robertshaw (xkavate)**, **Arjen van Rijn (SURF)** and **Bob Jones (CERN)**. **Arjen van Rijn** opened the discussion by summarising the commonalities of the presentations. As a first commonality Arjen van Rijn identified the issue about the sustainability of the infrastructures. This is addressed through the different funding models, where the use of the infrastructure in order to understand what to fund more and what is more demanded by the users is considered. The second issue is the involvement of the users, which are individual researchers and communities of the researchers but also citizen scientists. So the question is how the e-infrastructures, which have grown with the limited and under control communities of users, can prepare for users that are not so much under control. And the third issue are policies for data, which means also trust building in the infrastructure because the data are no more under the control. **Stephen Robertshaw** agreed about the commonalities but stated that an agreement is necessary to become a unified entity and then to move forward to implement this agreement. He also stated that this implementation is more about services that address end-users’ needs than technologies. **Bob Jones** stated that the users will vote “with their feet” if a pay-per-usage model is taken into account, but he supported that a user forum must be part of the decision making process.

## Track 2: Infrastructures in the Digital Age: the Data Challenge

### Result of ESFRI/e-IRG working group on Data Policy

The e-IRG Chair **Sverker Holmgren** opened the second track “Infrastructures in the Digital Age: the Data Challenge” by recalling the two e-IRG Blue paper related to data. The first Blue Paper in 2010 discussed how large-scale projects and their users could engage in a common e-Infrastructure. A more focused discussion on data management results in the Blue Paper from 2012, which addresses crosscutting themes for research infrastructures in general. To provide a more policy- or recommendation-oriented approach for ESFRI, a joint informal working group from e-IRG and ESFRI extracted policy recommendations like the need for sustainable e-Infrastructures to enable, storage of, curation, preservation and access to data. The joint working group identified also a number of actions needed i.e. the clear definition of roles for the different actors like users, data

[www.e-irg.eu](http://www.e-irg.eu)

e-IRG secretariat  
c/o Netherlands National Computer Facilities Foundation (NCF)  
P.O. Box 93575  
NL-2509 AN The Hague, The Netherlands  
Phone: +31 (0)70 344 0526  
secretariat@e-irg.eu

Visiting address  
Anna van Saksenlaan 51  
NL-2593 HW The Hague  
The Netherlands

**e-IRG**  
e-Infrastructure  
Reflection Group

owners, infrastructure providers, etc. and the need for governance structures and mandates for the different kinds of actions. Furthermore Sverker Holmgren explained that procedures, which allows and encourage innovation, are needed to enable science and research in ways that we cannot imagine today. The connection of data and resources in a transparent, coherent and easy way are an example for this thesis, especially under the viewpoint of the 48 research infrastructure of ESFRI's roadmap. The ESFRI cluster projects, which structure these RIs topic-wise enable the discussion of common needs and solutions within each cluster, but due to the very different disciplines a connection between these four clusters is difficult. So coordinated activities between those clusters are recommended to enable cooperation, accessibility, interoperability, harmonisation, utilization and standardisation. ESFRI should approach national funders and the Commission to explain the further need for coordination of the data infrastructure facilities and ESFRI should ask e-IRG to become more active in the organisation of these activities and to collaborate more closely together.

### **ELIXIR: European Bioinformatics Research Infrastructure**

**Rolf Apweiler** presented the ESFRI research infrastructure ELIXIR, which is a research infrastructure for European Bioinformatics Institute (EBI). EBI's service mission is to enable life sciences research and its translation to medicine, agriculture, bio-industry and society by providing biological data, information and knowledge. So the day-to-day activities of EBI are the collection of biological data from laboratories from around the world, their processing, classification and analysis. The data range from genes, genomes to protein and metabolite expression, protein sequences, molecular and cellular structures and that alike. The advancements in the DNA sequencing lead to a yearly duplication of the amount of data. To create a robust infrastructure for the resulting biological and clinic information is a bigger task than individual organisation or nations can take alone and life sciences has huge data needs and by far the largest research community. ELIXIR's mission is to build a European infrastructure for biological information, supporting life science and translation to society, to bio-industries, to environment and to medicine. In this infrastructure the central hub is hosted at EBI, which coordinates the activities of all the nodes and national centres. The services offered to the community by ELIXIR are manifold, but have one common point; they are all open access and free of charge, except data with privacy restrictions, which are under controlled access. ELIXIR provides tools and tries to integrate existing tools, enabling data access and mining by developing an interoperable tools infrastructure. There is a huge demand for training and a need to work on the development of standards for biological and medical nomenclature and controlled vocabularies and ontologies of reshaping among our databases. The future challenge for life science data services is to scale and to sustain the funding for a distributed infrastructure with really more than a million users, for managing and interoperating the big and heterogeneous data.

### **The growing challenges of Big Data in the Agricultural and Ecological Sciences**

[www.e-irg.eu](http://www.e-irg.eu)

**e-IRG secretariat**

c/o Netherlands National Computer Facilities Foundation (NCF)  
P.O. Box 93575  
NL-2509 AN The Hague, The Netherlands  
Phone: +31 (0)70 344 0526  
secretariat@e-irg.eu

**Visiting address**

Anna van Saksenlaan 51  
NL-2593 HW The Hague  
The Netherlands

**e-IRG**  
e-Infrastructure  
Reflection Group

**Chris Rawling** reported from the agriculture and ecological science, an area, which recently moved into some of the issues of big data. The particular focus of his research is to underpin the solutions to sustainable agriculture. While Elixir was largely concerned with dealing with the molecular data describing the genotype, Chris Rawling reported about the needs of the two ESFRI projects ANAEE and ISBE, which focus on the physical behaviour, structure and function of organism, the phenotypes and trades. The infrastructures needs are driven by the approach of the systems biology and importance of modelling and simulation, so that the needs are not only on data storage and distribution, but also on computational requirements. The advances in genome sequencing and related technologies create a data deluge, which requires distributed centres around the globe for the data analysis and modelling. And the next data deluge is in sight due to image-based technologies used in high-throughput phenotyping.

Chris Rawling underlined the importance not just of the data and their delivery to the researchers, but in the use of these data to build mathematical models in order to be predictive and to answer the "what if"-questions. These mathematical models are very important and they contribute into a broader approach to biology, which is regularly called systems biology.

Concluding it is to say that biology is a big data discipline and ecology is becoming one, drivers of this are not just next-generation sequencing technologies but imaging at all biological and geographical scales. This results in driving demand in compute, storage and networking capacities across the globe, not just Europe. The major challenges are in the development of standards, the data integration and interoperability of data and computing systems, and visualization.

### **Future steps of European data e-Infrastructures**

**Damien Lecarpentier** presented an overview on EUDAT, a cross-disciplinary collaborative infrastructure. The mandate is to build a layer of common data services for the different research communities. Currently 26 partners, from 13 countries form the consortium of EUDAT, among these partners are 7 research communities from various areas of science: EPOS from the area of earth sciences, CLARIN from the linguistics, ENES for climate sciences, LIFEWATCH for biodiversity, VPH from the bioinformatics, INCF from neurosciences and DRIHM for the hydro-meteorology. To provide the needed core services to the communities EUDAT is capturing the different needs and requirements of the different research communities. Damien Lecarpentier provided three key messages: the first is that the capturing of the community requirements takes time; a lot of expertise and skilful people are needed who act as bridges between the different user communities in order to translate between the different disciplines. To gather valuable information from the research communities a strong interaction, in-depth discussions and meetings are necessary. The second key message is the EUDAT is not developing new technologies, but that they use, reuse and adapt technologies that are already available. The third point is that EUDAT is an infrastructure with site, hard- and software. Although EUDAT is engaged in standardisation and recommendations, it is more than a policy group, because it is also providing services on a real infrastructure.

## Global Data Infrastructure

**Herman Stehouver** reported on the efforts of the Research Data Alliance (RDA), which is besides EUDAT a direct result of the “Riding the wave” report from the high-level expert group on data. The RDA Forum is a high-level forum of several representatives from LIBER, Science Europe, ERF, ALLEA and others. According to Herman Stehouver RDA receives strong support from the G8+5 Group of Senior Officials. RDA Europe provides support to RDA in the secretariat, the council, the technical advisory board, the organizational advisory board and different working and interest groups. RDA focus is on data sharing and interoperability, thus RDA Europe approaches scientists, data practitioners and research communities through its analysis program to investigate on gaps and overlaps in their data practises. Other activities of RDA Europe are the prototype program, which helps people to develop prototypes for data exchange and the travel support program.

## Introduction to Infrastructures in Horizon 2020

**Costas Glinos** reviewed shortly the activities during the seventh framework program, which led to the RDA and projects like EUDAT and OpenAIRE and subsequently he presented the planned activities during the Horizon 2020 framework related to e-Infrastructures. Three vertical pillars will be the main part of Horizon 2020 Research Infrastructure activities, which are the development of new world-class RIs, the integration and opening of existing national RIs of pan-European interest and the development, deployment and operation of ICT-based e-Infrastructures. These three main pillars are horizontally crossed by two activities, which are concerned the innovation potential of RIs and their human capital, and the reinforcement of European RI policies and international cooperation.

Costas Glinos said that the EC is aiming to overcome the distinction between generic and service-driven infrastructure on the one hand, and the community driven infrastructures on the other hand. Furthermore the EC wants to remove the silos between research infrastructures and e-Infrastructures as well as e-IRG and ESFRI.

Another challenge is the integration between data and computing; Costas Glinos reminded that e.g. EGI started as a data grid but that we have now a separation between data and computing, which will be challenging to overcome. He said that the EC wants to have a rearrangement of computation and data infrastructures in this part. He explained that a continuation with some cooperation between the infrastructures would not work.

A third challenge arises from the policies on openness, e.g. the Open Access recommendations. In the work program the infrastructures are asked to support the Open Data policy and provide support to the people who have the demand to open their data. The Open Data policy in the Horizon 2020 framework will create a strong demand for support from the e-Infrastructures, since a significant amount of projects have no experience regarding the preservation and curation of data, metadata, data management, etc. The same questions arise regarding Open Software.

[www.e-irg.eu](http://www.e-irg.eu)

e-IRG secretariat  
c/o Netherlands National Computer Facilities Foundation (NCF)  
P.O. Box 93575  
NL-2509 AN The Hague, The Netherlands  
Phone: +31 (0)70 344 0526  
secretariat@e-irg.eu

Visiting address  
Anna van Saksenlaan 51  
NL-2593 HW The Hague  
The Netherlands

**e-IRG**  
e-Infrastructure  
Reflection Group

Further Costas Glinos identified the sharing and provision of basic operational service throughout the community as another challenge in Horizon 2020. As an example he took the registration authority for PIDs. Questions like ‘who will be the provider of the authentication and authorization services’ and ‘how this could be completely federated, so that the service is interoperable across Europe’ have to be solved.

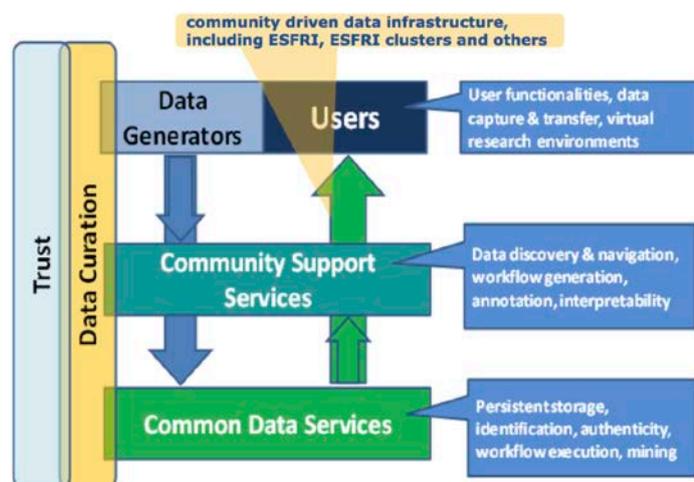
Finally the notion of business planning is seen as a challenge in Horizon 2020 since those infrastructures, which are providing basic services will be serious about their financial sustainability, which is in both interest, users, as well as providers. The EC does not privilege any particular solution but the projects have to argue why they prefer the selected business plan; people should think carefully upfront how their business planning will be and present this within their proposals.

### Panel Discussion on Infrastructures in the Digital Age: the Data Challenge

**Alf Game (ESFRI)** chaired the panel discussion on Infrastructures in the Digital Age. He invited the panellists **Sverker Holmgren (e-IRG)**, **Costas Glinos (EC)**, **Elena Righi-Steele (EC)** and **Yannis Ioannidis (ESFRI)** to draw together some themes and conclusions with the view to come up with some ideas about what e-IRG and ESFRI might be able to do to improve the general effectiveness for research infrastructures and have an integrated vision for them, including e-infrastructure requirements in particular data.

**Sverker Holmgren** identified the connections between the community support services and the common data services as an issue that need to be tackled. He stated that it must be made clear to the researchers and the communities that they benefit from this model. The first step to bring these users to the e-infrastructures and the data services is to encourage them to use the national infrastructures. To improve the relation between horizontal services and domain-specific services one has to start at the national initiatives.

#### data infrastructure architecture



**Costas Glinos** said that the majority of EU funded projects are not from the infrastructures programme or from ESFRI, so it would be a major step ahead if these other projects are aware of the generic services and use them.

**Yannis Ioannidis** remarked that techniques and services come from specific areas and become more generally applicable if the community recognize this general applicability. Research

infrastructures and e-infrastructures need to collaborate so that the framework exists, which allows these specific techniques or services to become more generic. Furthermore he stated that computing infrastructure and data infrastructures depend on each other, so that one can serve the other. **Norbert Meyer** said that data infrastructures in the different communities are currently isolated islands, which need to be connected. He agrees that new initiatives and new architectures are necessary, but existing databases and infrastructures need to be taken into account.

**Costas Glinos** pointed out that it would be difficult to recognize which of the service should be generic and which user-specific. He said that there will be many mistakes assessing services concerning their general applicability, so he suggested to develop guiding principles to overcome these issues.

**Alf Game** concluded that the underlying consensus is to broaden the thinking of the people so that they understand where they may sit in the diagram and have graceful awareness of the processes.

**Françoise Genova** commented that the astronomy community has long experience with re-use and sharing of scientific data and that the community is ready to share this experience. Although each community has its specificities because of its culture and history, to share the experience will help to build a generic system. To re-use data the input from the origin community is necessary to create the metadata, without which data is useless.

**Tiziana Ferrari** explained that the issue of engaging with the users means to use existing user communities as a bridge, but also introducing and collaborating with the user communities that support the other users. Tiziana Ferrari said that the diagram shows the technical structure of the infrastructure but is lacking support services where the users can get support from the infrastructure providers, from the technology providers, but also from the user communities that have already solved some of the issues. **Kees Neggers** stated that the diagram shows the data infrastructure architecture, but is lacking a control plane, which shows how the system is maintained, governed and funded. If this control plane is defined with all the actors that keep the system healthy, then the e-Infrastructure Commons will be there.

**Elena Righi-Steele** said that she saw the same problems with all communities irrespective of the field of science. She said that there are several barriers that need to be broken, horizontally and vertically. She underlined the need for common services that need to be applied to all the communities. The inter-community engagement will help to provide faster response to all sorts of issues. She pleads for a dialog between the user communities, the providers of the access and the funders of these providers, the national agencies, the research infrastructure providers, the ministries and the funding agencies at all levels. **Françoise Genova** agreed completely that people have to share their experiences and lessons learned, and said that it is important to bring practitioners to participate in building solutions, which can be used by many.

**Costas Glinos** stated that RDA is the implementation of the arrow in the diagram that goes from the community support to the common data service. Regarding the user involvement he added for consideration that this is not a panacea, because users tendency to build their own systems. The realization of the cost of things and the necessity to look for accordance of scale forces to aim for

a generic solution and will finally lead to the business model. The connection to possible funding models will build an important incentive for the evolution of the infrastructure. **Kees Neggers** agreed that users have the tendency to build their own system, but that they do so if they lack a service provider who provides the services in question. He explained that in the end, most users have interest in their discipline and not in tool development. According to Kees Neggers the challenge is in the generic tool development and service provision to be ahead of the users' needs and therefore the input from the users is needed in the early phases of the development, which creates the need for upfront investments and innovation. He explained that in the Netherlands the government supports the innovation and that the users pay for the resulting services, which provides a feedback loop that is essential. **Bob Jones** explained that in the Helix Nebula project commercial suppliers develop services on their own cost and that the users pay for the usage of these services.

**Yannis Ioannidis** said that the scientist often don't know about generic tools available to solve their problems. He agreed that RDA is a very good framework that attracts the right people, but that the end-users are still missing and that some effort is needed to bring these end-users into RDA to communicate their needs. Yannis Ioannidis concludes that the collaboration between computer scientists, data practitioners and end users should also permeate the funding structures. Further he called e-IRG and ESFRI to collaborate closer together on all levels. **Sverker Holmgren** agreed that some effort is needed, together with some incentives that have to be brought in from a higher level. **Alf Game** said that part of this discussion is to realize that the reason scientists are doing science and the reason funders pay for it are not the same. And he was very excited that RDA has put the scientific data on the agenda of G8. **Bob Jones** added that the objectives are to get these e-Infrastructure Commons recognised at the same level as the ESFRI RIs in some years.

[www.e-irg.eu](http://www.e-irg.eu)

**e-IRG secretariat**

c/o Netherlands National Computer Facilities Foundation (NCF)  
P.O. Box 93575  
NL-2509 AN The Hague, The Netherlands  
Phone: +31 (0)70 344 0526  
secretariat@e-irg.eu

**Visiting address**

Anna van Saksenlaan 51  
NL-2593 HW The Hague  
The Netherlands

**e-IRG**  
e-Infrastructure  
Reflection Group