SUMMARY e-IRG POZNAN WORKSHOP 12-13 October 2011

Around 120 participants attended the e-IRG workshop organized in Poznan on 12-13 October 2011, under the auspices of the Polish EU Presidency. Policy issues related to e-Infrastructures was the main focus of the first day of the workshop, with topics such as sustainable e-Infrastructure, possible synergies between EU funding and national funding, as well as presentations of Polish examples of sustainable and reliable local infrastructure. The second day was devoted to technology aspects, with focus on data infrastructures and data management.

Jan Weglarz, Director of PSNC, Poland, warmly welcomed the participants and gave a short overview of the e-Infrastructure development in Poland. **Maria E. Orlowska**, Secretary of State, Ministry of Science and higher education, Poland, further elaborated on the e-Infrastructure development in Poland over the past 20 years and highlighted the importance of involving the policy level for sustainable development. Orlowska also highlighted the substantial funding available in Poland for ICT research infrastructure on the national level as well as the use of EU structural funds.

Gudmund Høst, Chair of e-IRG, spoke about the challenges for the European e-Infrastructure, in particular the need for Pan-European e-Infrastructures crossing multiple administrative and national domains, which is an important part of the e-IRG mission. He summarised the e-IRG White Paper 2011 and highlighted that data management policies are extremely important for RIs. Høst also underlined the opportunities offered in the EU, with examples of FP7 clusters of ESFRI infrastructures and pan-European FP7 data e-infrastructure initiatives such as EUDAT.

Kostas Glinos, Head of Unit GÉANT & e-Infrastructure in DG INFSO, EC, underlined the transition towards an e-Infrastructure of services, where relevant governance and cost models and a proper regulatory framework have growing importance as well the connection to the industry. Glinos also presented the draft plan of Horizon2020, to be agreed on in mid-2013, hopefully with a first call by the end of 2013. The draft plan shows a re-distribution in key policy areas, with more allocation of funds to Research and Innovation and Infrastructure. The outcome of the consultation of Horizon2020 showed strong support for the concept to bring research and innovation closer together and the need for simplification of access of funding. The EC Workshop on RIs (4th July 2011) showed that the role of e-IRG and ESFRI should be further strengthened and also supported the need for training and education as well as partnership with the industry. To assess and evaluate research infrastructures EuRoRIsNET[1] was being asked to make a catalogue of all research infrastructure and e-infrastructures currently in operation or being planned. This system for assessment and evaluation is needed for future orientation. The resulting discussions will trigger strong interaction with e-IRG. Glinos was asked by the audience to comment on the development of ERIC (legal framework for a European Research Infrastructure Consortium). Glinos stated that the first ERICs have been created but so far there is no rush from the infrastructure to become ERICs and it is still a learning phase. This may change once the industry becomes further involved, according to Glinos.

Matthiew Dovey, JISC, UK, gave an overview of the two ERANET projects e-Infranet and SIM4RDM. e-Infranet is a policy project with the aim to develop and strengthen policy cooperation and coordination between national e-infrastructure programmes, ensuring their smooth and efficient integration in the European Research Area. Main focus is put on the themes Green ICT, Cloud computing and Openness (Open Access, Open Source, Open Borders and Open Opportunity). Support Infrastructure Models for Research Data Management (SIM4RDM) is a new project,

[1] EuroRIsNET is the network of contact points at European level for research infrastructures

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

which supports infrastructure models for research data management and will include publishers and different funding models in various Member States.

Sverker Holmgren, NordForsk, Sweden, presented the Nordic eScience Globalisation initiative, as an example of regional coordination, support and funding of large scale research projects. The eScience initiative has a combined top-down, bottom-up approach and is a coordinated effort of eScience research, education and e-Infrastructure. Funding has been given in a common pot from the Nordic countries (apart from Denmark). Focus is put on supporting leading eScience research and education, mainly in the area of health and climate/environment.

Manfred Horvat, Vienna University of Technology, Austria, rapporteur of Synergies Expert Group (SEG) started by reflecting on the massive changes since his first involvement in EU cooperation in the 1980s, when the fax machine was the most indispensable technical tool for smooth cooperation. SEG aims to develop synergies between EU programmes for the current and the next programming period, for instance by aligning the cooperation between policy frameworks, programmes and actions. Some of the findings from SEG are fragmentation of innovation policies at EU level, lack of coherent governance and of instruments for pooling European and national funds. One recommendation for the current programming period is to test possibilities for interoperability of programmes. For the next programming period Horvat concluded that close coordination and cooperation of HORIZON 2020 and the Cohesion Policy Funds oriented towards the objectives of the Europe 2020 and Innovation Union will strengthen R&I in the EU. It will become an important European competitive advantage, with the EC as "facilitator" and the Member States and regions as the main actors.

Possible synergies was also the topic for **Christos Vasilakos** DG Research and Innovation, EC, who presented "RIs in Horizon2020 and possible EU actions to increase synergies". Vasilakos described the RIs as the core of efficient research and innovation strategy and the backbone in construction of ERA and therefore they require world-level quality in all aspects of their activities. Vasilakos concluded that in order to increase synergies the role for EU in the coming programming period could be a more leveraging one to help developing a framework for ERA, with better industrial access to facilities. The role of national/regional authorities could be to exploit the potential for synergies when implementing the different EU policies and to foster coordinated access to the EU funds.

Daniela Zaharie, West University of Timisoara, Romania, described the impact on the RTD activities of the support from National programmes, EC Structural Funds and EC FP6/FP7 programmes in the case of a Romanian team from 2002 onwards. There was an increase in success from FP6 to FP7; the money attracted had increased significantly. Her conclusion was that there are both national and international elements which have a critical impact in reaching competitiveness and even if there is no such thing as a "success recipe", there is no success without a synergetic exploitation of European and national opportunities.

The workshop session on "Polish feedback for sustainable and reliable local infrastructure" started with a presentation by **Artur Binczewski** PSNC, Poland, who spoke about PIONIER – the National Research and Education Network. Binczewski gave a brief history on the Polish e-Infrastructure for science and explained the transmission topology of PIONIER. Examples of international connectivity were GÉANT and cross-border Dark Fibre connectivity to neighbouring countries, especially to Germany. Some examples of projects using the e-Infrastructure were Radio astronomy and Future Internet Engineering.

Marek Niezgodka and **Maciej Filocha**, University of Warsaw, Poland, presented the POWIEW project, Polish High Performance Computing Infrastructure. The project is focusing on high-end computing for grand challenges of Science and Engineering and is funded by the EU structural funds. The aim is also to facilitate and support new scale research in Poland. Niezgodka and Filocha underlined the importance of keeping a balance between "easy to use" and interoperability, to keep access simple. Examples of research areas given were new scales of resolution in numerical weather prediction and modelling of the structure of the universe. Some of the future outlooks are synchronised key infrastructure investments and petascale computing for Polish academic community.

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

Lukasz Dutka, ACK Cyfronet AGH, Poland presented the project "PL-Grid", a Polish grid initiative. Its main objectives are the creation of a common base infrastructure, plans for HPC and scalability computing and to have the capacity to construct specialized domain grid systems. The grid infrastructure has so far been used in many different fields such as biology, quantum chemistry and astronomy and it will be operable for at least 5 years. Dutka summarised that heterogeneity of grids is the key, but in Europe we often try to be homogenous to facilitate administration of work.

The PLATON project – a service platform for eScience was the topic of the presentation given by **Robert Pekal**, PSNC, Poland. The project (2008-2012), is largely financed by EU and the Polish Ministry of Science and Higher Education and consists of five services: Eduroam, Campus services, Archiving, Science HD TV services and Video conference services. The campus services works as a "cloud" for scientific computing and education and the user chooses the time when he/she needs the resources by a calendar-like interface. Pekal summarised that there are five basic yet important and modern services that constitute an excellent base to build in future more sophisticated services/platforms or upper layer complex services and that this approach is one of the directions of NRENs development. Pekal got the question from the audience if the cloud service he described was directed to end-users or to IT-departments and he clarified that the end-users are the target group.

Marek Niezgodka, University of Warsaw, Poland, talked about the development in Poland regarding Open Knowledge infrastructures. He mentioned among other things the Virtual Library of Science which gives free unlimited access for research and higher education (since 1996) and the Federation of Digital Libraries as well as SyNaT - national collaboration of 16 partners in research and system development. Future perspectives for Poland are open data act, open mandates for scientific publications and enhanced publications - an integrated infrastructure and National research publications.

Giulio Notarstefano, OGS, Italy, started the first workshop session of the second day; "Grand Challenges", by presenting the Euro-Argo project, the European contribution to worldwide ocean monitoring. Euro-Argo is an ESFRI project in its preparatory phase. Euro-Argo is the European part of a global ocean observing system (Argo), which is based on numerous floats distributed worldwide on all seas. The goal is to deploy, maintain and operate 800 floats. Euro-Argo aims to be organised as an ERIC from 1 January 2012, which provides the overall coordination of the project. There was a question from the audience regarding the challenges on the data management of all these data collected and how they are used by researchers and can be made useable/accessible by the researchers. Notarstefano answered that the structure of data management is well organized in the sense that the data are transmitted to the national data centres, and then transmitted to the central data centre. These data are freely available and any application that can allow the user to get the data quickly can be useful.

Pär Strand, Chalmers, Sweden, presented "ITER: a driver for science and e-infrastructures?" The aim of ITER is to demonstrate that it is possible to produce commercial energy from fusion. It is an experimental facility and an international collaboration between the EU, China, South Korea, Japan, Russia and the USA. This requires data sharing, deployment and replication in several offsite repositories and also (close to) real time data streaming to all areas of the world. The fusion reactor itself is still under construction. The hope is to have the first plasma in 2019, and to be in full operation in 2026. The ITER Modelling Framework (IMAS) shall be operational well before ITER Operation and the infrastructure and modelling have to live over 30 years. The IPR is already a challenging issue, middleware and authentication are complex questions and also data integration for multiple resources. Most serious issue is the topic of governance models. Is ITER a driver for e-Science and e-Infrastructures? Yes, according to Strand, but now there is time to influence from the infrastructure point of view. For one or two years the ITERIS project will provide recommendations to ITER on these issues, which will be discussed with all the partners.

Rosette Vandenbroucke, University of Brussels, Belgium, gave a presentation on the OSIRIS project, which deals with challenges to found a sustainable e-infrastructure. Vandenbroucke suggested a mixed approach of top-down and bottom-up, to include both user communities and governments. She mentioned in particular four categories of parameters to consider: Governance, Access Policy, Operational Principles and Sustainability. Vandenbroucke also suggested to reflect on the various funding types (project based, use-based/service-based?) and sources (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

national industry, institutes, paying users?) to be able to find the best possible solution. To consider the expenses is also of importance – will it be mostly capital or operational? Another issue is the "evolution" with questions such as number of users, stakeholder requirements and the time horizon. The OSIRIS project will produce a document with recommendations, which will available by the end of the year.

The next session was focusing on Data Infrastructures and Data Management. **Frank Schluenzen**, DESY, Germany, spoke about photon and neutron facilities. PANdata brings together eleven major world class European research infrastructures to create a fully integrated, pan-European, information infrastructure supporting the scientific process. PANdata shared data infrastructure focus on Neutron and Photon areas and provide these user communities with data repositories and data management tools. One example of a scientific experiment is the investigation of a new species of an early human ancestor by x-ray microtomography. Many PANdata users come without experiment or compute experience and the data infrastructure needs to comply with these inexperienced users. Schluenzen concluded that encouragement (rather than recommendations) is needed from funding agencies, policy makers, publishers, since the users are very reluctant to share data. Also collaboration regarding interoperability is very important for the project.

Leif Laaksonen, CSC, Finland, told the audience about the newly started EU project EUDAT - Towards a European Collaborative Data Infrastructure. Laaksonen explained that he wanted to open up for a discussion for future collaboration. The main objective of EUDAT is to build the generic data infrastructure layer of the e-Infra vision. Some of the key activities are capturing communities' requirements and providing fundamental core services such as long-term preservation and data access, upload and workspaces. The main expected benefits are to enable the multi-disciplinary data intensive research and collaboration, ensuring wide access to and preservation of data in a sustainable way. EUDAT has 25 partners from national data centers, technology providers, research communities and funding agencies and Laaksonen underlined that additional partners can still be relevant in the future.

Paolo Manghi, CNR, Italy presented the OpenAIREPlus project, which aims to deliver technology, developing an open access data infrastructure for scientific information that includes publications, datasets, funding and to interlink all of them. It capitalises on the OpenAIRE infrastructure that was built primarily to include FP7 and ERC publications. In parallel, there are also research activities dealing with the design of a data model, content management services and an orphan repository for those who have no repository to store their research data and publications. Guidelines for publication providers and repository managers will be provided, especially on metadata and protocols to be used. The project will be using the D-NET software toolkit and INVENIO repository from CERN that are widely used in European projects.

A presentation of the plans for a data infrastructure for Digital Cultural Heritage (DCH) was made by **Antonella Fresa**, Ministry of Culture, Italy. The amount of digitized material in the European Cultural sector is growing rapidly. This generation of data is accelerated by Europeana that is fostering the European cultural institutions. The needs of DCH include high quality information technology management, access facilities and interoperation. Fresa underlined that it is not a new infrastructure but a new approach based on national and regional interoperable systems using existing resources. The main actions required in the DCH area are to improve awareness, promote trust, establish priorities, consult stakeholders, and to promote international cooperation. The three main projects to take care of these actions are DC-NET: joint programming for DCH e-Infra implementation, INDICATE: international cooperation user case studies, pilots and LINKED HERITAGE: a best practice network on metadata and standards, linked data and persistent identifiers, multilingual vocabularies, aggregation of content to Europeana.

Andrew Lyall, ELIXIR Project Manager, European Bioinformatics Institute, Hinxton, Cambridge, UK, presented the BioMedBridges, a cluster project which will start in January 2012, with the aim to build bridges between the ESFRI BMS RIs, clustering them together and linking basic biological research data to data in the other domains. Its scope is to provide secure, robust and ethical access to biodata. BioMedBridges will construct the e-infrastructure to allow interoperability between data and services in the biological, medical, translational and clinical domains. It will also include major e-Infrastructure stakeholders such as GEANT/DANTE, EGI.eu, PRACE and CERN.

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

Birger Jerlehag, SND, Sweden, introduced the audience to the DASISH project- the harmonization of five ESFRI Infrastructures within Social Sciences and Humanities. DASISH is getting together CESSDA, CLARIN, ESS (European Social Survey), DARIAH and SHARE (survey of Health, Ageing and Retirement in Europe). These projects are dealing with same kind of data such as personal data and registries and have common challenges, such as how to achieve integration and interoperability beyond the borders of the individual projects given different data organizations? How can we improve the quality of our data to enable advanced and cross-disciplinary access and enrichment operations? How to establish trust of SSH researchers in the infrastructure services? Some of the solutions found are to work on all aspects that will foster shared data access and enrichment and to take care of legal and ethical issues that are of relevance for all SSH domains in a cross-disciplinary activity and work on simplified solutions.

The last part of the e-IRG workshop in Poznan was devoted to panel discussions. The panel discussions focused on the questions on "how to integrate the data infrastructure with the existing grid and HPC infrastructures" and "Users and infrastructure providers – demands vs. offers". The panel consisted of **Peter Wittenburg**, MPI, NL (User community representative), **Richard Kenway**, Chair of the PRACE Scientific Board, University of Edinburgh, UK (HPC e-Infra provider), **Steven Newhouse**, EGI.eu Director, NL (Grid e-Infra provider), **Leif Laaksonen**, CSC, FI (Data e-Infra provider) and **Lars Börjesson**, Chalmers University of technology, SE (ESFRI vice chair).

Peter Wittenburg stated that researchers need a framework to create virtual collections integrating data and access them (distributed and multidisciplinary) and they want to deploy services close to the data but they don't care about the technology. They are interested in a cheap technology that is transparent to the users or user communities. The user communities need an explicit cost model and services, which are separate from innovation. Wittenburg claimed that there will be no single technical solution to all issues the user communities cope with in this heterogeneous world regarding data.

Richard Kenway described the transition of PRACE from a TIER-0 resource provider to an integrated infrastructure provider spanning the range TIER-0 and TIER-1 resources. He said that the approach of buying large petaflop systems and bringing them into the PRACE organisation is a very top-down approach. For some users these machines are not ideally configured to meet their needs, e.g. the climate community. PRACE is beginning to establish a user forum to address these issues. From now on the competition is open worldwide, since PRACE can only be a success if it is supporting the very best science in the world. Right now industry can come in, in collaboration with academia but from the call in May next year PRACE will be open to industry-only projects as well.

Steven Newhouse mentioned three challenges of the integration of data infrastructures into existing HPC and grid infrastructures: governance, technical services and the operation infrastructure of those technical services. Newhouse emphasised EGI's national network of resource providers, which provides an unprecedented organisational network of over 350 sites and centers. Regarding the topic user and infrastructure – demand versus offers, Newhouse commented that the users quite rarely have demands on how they expect the infrastructure to be and they need to provide collaboration, recognition and endorsement of the infrastructure that they are using.

Leif Laaksonen stated that there is no solution that fits everybody, that it is necessary to build up something which is a patch of different things with a very pragmatic approach. He underlined that the infrastructure providers should work closely together to build complementing solutions and not fall into a competition. Laaksonen stated that the data integration requires a lot of inside knowledge and collaboration and it is important for the key players to share knowledge and to work together.

Lars Börjesson gave an overview of the current ESFRI roadmap, which comprises a strategy plan and an action plan, and includes 48 projects from which 60 % should be implemented until 2014. Börjesson pointed out that from the ESFRI side, the e-infrastructures issues are important for all of the RI projects, which are in the ESFRI roadmap. They are also important for the synergies between the RI projects and for addressing the grand challenges. There is a demand from ESFRI on a larger degree for cross-disciplinary data integration. These issues should be addressed by the research infrastructures joint program initiative. ESFRI is now developing large pan-European research infrastructures which faces the grand challenges, but also the large societal challenges projects which makes data

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

sharing and data interoperability absolutely essential. Börjesson concluded that EUDAT can be the answer to all of these different challenges, since it tries to unify everything.

Starting the discussion **Norbert Meyer** (PSNC) asked if a data infrastructure besides the existing HPC and grid infrastructure is needed. **Wittenberg** answered that in some areas HPC and data are very close together, other communities are organised differently and an interesting outcome of EUDAT will be to see how to abstract from concrete cases to something more generic. **Kenway** stated that the issue of data is going to change as we head towards exascale when the dominating cost is going to be moving data. **Newhouse** explained that data exists as a sort of separate, first class member of the e-infrastructure ecosystem. He urged the needs of the data orientated services and the service providers to get integrated into the existing governance structures that we have. **Laaksonen** stated that data has not been properly understood but he expects the value of data to be growing over time. **Börjesson** pointed out that data is necessary to look upon the integration of all the different layers. **Françoise Genova** (member of the High Level Expert Group on scientific data) was worried about that everything should be integrated into the existing structure, since the diversity answers some of the needs of the different communities. She suggested taking a bottom-up approach before integrating into an existing single structure.

Meyer also asked the panelists about the major problems in terms of supporting the user requirements regarding data, which are on users' site. **Wittenberg** explained that the existing data infrastructures do not serve all of the heterogeneous wishes of the different user communities and it is necessary to have a system that has clear identity providers and agreements between the countries. **Wittenberg** also called for a Persistent Identifier (PID) and a persistent identifier generator authority and there will be a workshop in April 2012, which will address these things. **Newhouse** stated that all communities seem to have agreed that identity, identity access and the identification of data are important and we should be careful to enforce to use that system over other systems.

Morris Riedel (EMI) asked the panelists about the opinion on software regarding the maintenance, on the provisioning and the development. **Wittenberg** suggested making the costs explicit and asking the research organizations finally – not even the countries. **Laaksonen** recommended a rewarding system for software and data. **Newhouse** missed "responsibility" in the discussion - people who are making the most benefit of various services in each layer need to take a responsibility for the maintenance and development. **Kenway** referred to a recent IDC report, where the idea of virtual centers of excellence for software is proposed.

The workshop ended with some concluding remarks by **Gudmund Høst**, who saw an emerging theme on the role of e-IRG: Action instead of Reflection! He thanked the audience and the Polish organizers, especially Norbert Meyer and reminded about the next e-IRG workshop which will be under the Danish presidency in June 2012.

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