

e-IRG Workshop Summary

10-11 Nov 2014 Rome



Glossary

AAI - Authentication and Authorisation Infrastructure

AEG – Assessment Expert Group

AISBL - Association Internationale Sans But Lucratif

BOD – Board of Directors

CoE – Centre of Excellence for Computing Applications

EC – The European Commission

e-IRG – e-Infrastructure Reflection Group

ESA – European Space Agency

ESFRI - European Strategy Forum on Research Infrastructures

FPA - Framework Partnership Agreement

IAC – Industrial Advisory Committee

KB - Knowledge-Base

KPI – Key Performance Indicator

PPP - Public-Private Partnership

RI – Research Infrastructure

ROI – Return on Investment

SHAPE – SME HPC Adoption Programme in Europe

SSC – Scientific Steering Committee

SWG - Strategic Working Group

SWG - Strategy Work Group

VP – Vice-President

WG – Working Group

WP – White Paper

WPG – Work Programme

WW – World-wide

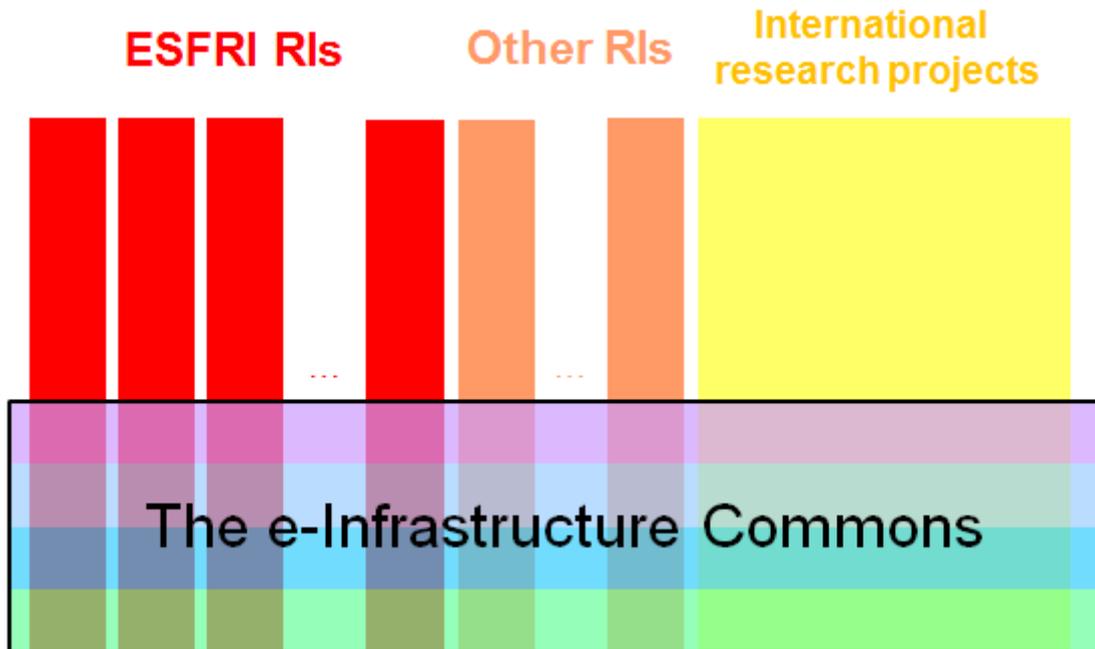
Contents

Glossary	1
Main conclusions.....	4
Day 1.....	5
11:00 Sverker Holmgren (e-IRG Chair) Welcome address; goals of the workshop.....	5
Workshop Notes.....	5
11:30 Anni Hellman (EC DG CNECT) Presentation from the EC	6
Workshop Notes.....	6
12:00 Giorgio Rossi (University of Milan, Italy - vice-chair ESFRI) Towards the new ESFRI Roadmap 2016	7
Workshop Notes.....	7
Chair: Enzo Valente Track 1: Outlook towards eScience and International Research Infrastructures. 8	
14:00 Patrick Aerts (NLeSC, The Netherlands) PLAN-E: European co-operation of eScience centres 8	
Abstract	8
Workshop Notes.....	9
14:25 Federico Ruggieri (GARR/INFN, Italy) CHAIN REDS: Coordination of intercontinental infrastructures.....	10
Abstract	10
Workshop Notes.....	10
14:50 Davide Calonico (INRIM, Italy - Director of Research) Metrology infrastructure for research 11	
Workshop Notes.....	11
15:15 Ingrid Mann (EISCAT, Sweden - Head of Projects, EISCAT) EISCAT-3D	11
Workshop Notes.....	11
Chair: Françoise Genova Track 2: ESFRI projects	12
16:20 Massimo Cocco, INGV, Italy EPOS: European Plate Observing System.....	12
Abstract	12
Workshop Notes:.....	13
16:45 Rafael Jimenez (EMBL-EBI, UK - ELIXIR Chief Technical Officer) ELIXIR: The European Life-Science Infrastructure for Biological Information.....	14
Workshop Notes.....	14
17:10 Mark Hagen (ESS - Head of the ESS Data management and Software Centre) ESS: The European Spallation Source	15
Workshop Notes:.....	15

17:35	Miles Deegan (SKA, UK - Engineering Project Manager SKA) The Square Kilometre Array	15
	Abstract	15
	Workshop Notes.....	16
18:00	Tibor Kalman (GWDC - DARIAH-ERIC - Co-Head VCC1) and Eveline Wandl-Vogt (Austrian Academy of Sciences - DARIAH-ERIC - Co-Head VCC1) DARIAH-ERIC: Towards a sustainable social and technical European eResearch Infrastructure for the Arts and Humanities	16
	Abstract	16
	Workshop Notes:.....	17
18:25	Panel.....	17
	Day 2.....	21
	Chair: Ilmars Slaidins Track 3: European e-Infrastructures for Data Management and Computing. 21	
09:00	Sanzio Bassini (CINECA, Italy - Chair of the PRACE Council) The PRACE Supercomputing Research Infrastructure	21
	Workshop Notes.....	21
09:25	Damien Lecarpentier (CSC - IT Center for Science, Finland - EUDAT Project Director) EUDAT Collaborative Data Infrastructure: future perspectives	21
	Abstract	21
	Workshop Notes.....	22
09:50	Tiziana Ferrari (EGI.eu - Technical Director) Open Science Commons for the European Research Area	23
	Abstract	23
	Workshop Notes.....	23
10:45	Christian Grimm (DFN, Germany - General Manager DFN) and Bob Day (Janet, UK) Geant perspectives in H2020 era	25
	Workshop Notes.....	25
11:10	Panel discussion & discussion with audience	25
12:00	Sverker Holmgren Wrap up and conclusion.....	27
12:10	Sverker Holmgren, Enzo Valente, Ilmars Slaidins Final words, Remarks, Actions	27

Main conclusions

The main conclusion: The intersection between vertical and horizontal research infrastructure needs further discussions. The adopted model could look like this:



Main Workshop messages: please refer to Page 1

Day 1

Opening (11 05 am)

Enzo Valente opens the event. Refers to the past of the location. Emphasises Italy's focus on cultural areas of science and research.

11:00 Sverker Holmgren (e-IRG Chair) Welcome address; goals of the workshop

Workshop Notes

Sverker Holmgren welcomes the guests. Thanks should go to Italian presidency and the organisers. *This is an open workshop and discussion is encouraged.*

The task of e-IRG: It provides advice and guidelines on European e-Infrastructures in the area of science and research. Delegates are appointed by ministries and the EC. Its task is to promote open, constantly innovating e-Infrastructures for science and research. The e-Infrastructures in question are in the area of connectivity, computing, data, tools provision and services.

The current challenges and questions:

- Data - Bringing all components together because data is produced and shared by most RI's. In the current landscape all RI's deal with data.
- Connection between discipline-oriented and horizontal RI's – Discipline oriented RI's are the ESFRI projects on national or European level. Horizontal RI's support all types of research for many projects, with tools and services provided in various ways.
- Point of contact – One of the key questions is '**Whom to approach to fulfil my RI needs?**'

Next White Paper Topics (*Best e-Infrastructure practices for large-scale research* (due the end of 2014)):

- e-Infrastructure Commons - There should be a **single e-Infrastructure Commons** for all users - a single interface, for users in order to easily choose the services they need (and let scientists and the RI's do what they do best). The e-Infrastructure Commons should span across all e-Infrastructures, ESFRI Projects, International Projects, Data infrastructures and Projects, etc.

e-IRG/ESFRI Coordination - A lot of discussions have taken place on this relationship including horizontal RI's. e-IRG submitted input to the ESFRI roadmap (e.g. landscape analysis, it also takes part in scientific evaluation). There is an **Overarching WG** consisting of all e-IRG representatives in ESFRI working groups. The next e-IRG White Paper will tackle ESFRI and related guidelines.

Enzo Valente – Sets priorities for the event and describes the contents (discussions on supercomputing, data, open science and the RI's needed). Describes the context of WP 2014.

11:30 Anni Hellman (EC DG CNECT) Presentation from the EC

Workshop Notes

Explains the structure of the new European Commission. Main features are: 7 VPs', they have assembled their respective portfolios to achieve convergence and synergies, each VP looks after a specific area. DG Connect – under Günther Oettinger– Digital Economy and Society, report to VP for Digital Single Market Andrus Ansip. The Commissioner for Research is Carlos Moedas; reporting lines are more complicated now; there is a strong ICT agenda in place.

The new **Commissioner for Digital Economy and Society** is Günther Oettinger; a letter from Commission President Jean-Claude Juncker has the following message: *Digitalise Europe*; this involves revolution, legal steps, high quality digital network/infrastructure, innovation always related to industry, creative industries/media; the task is to turn digital research into a success story by supporting entrepreneurship, start-ups, new business, and innovation.

Commissioner for Research – Carlos Moedas. RI a priority, a more focused portfolio is in place now; there is a clear message: research and its results should lead to growth and jobs.

Work Programme 14/15 in progress (next **WPG 16/17** is being prepared). Investment is in the range of € 200-230 million/ year. **WPG 14/15** has two e-Infrastructure related sub-calls, one in 2015. Support Actions – one in progress one in 2015. The deadline for *e-Infrastructure for Open Access* was in 2014 (has passed - one proposal for € 13 million).

Timetable: evaluation - 5 months after call closure, results one day after announcement, 3 months to sign (these are tight deadlines, no negotiation part now).

2nd September 2014 – a lot of calls were closed, e.g. data RI, HPC, HPC for SMEs, core services, research and education networking, e-Infrastructure policy development (42 proposals, most for the e-Infrastructure big data/29/), Support action – € 4, 82million in total, 14 proposals for international cooperation. Timetable: October -evaluation, ethical assessment – in progress now, January – reports and results, April signing and grant agreements.

Two more calls, Infra and Support Action with a **14th January 2015** deadline (e.g. CoEs – Centres of Excellence for Computing Applications, VRE- e-Infrastructure for virtual research environments, Support Action – new professions and skills). The process is to close in June 2015 (5months after the call), ethical screening – March/April, signatures – August.

WPG 16/17: Consultation in progress. WPG structure –to be approved by EC in 2nd half of 2014/, draft 1st in the 1st half of 2015, final - 2nd half 2015. Focus on integration, interoperability, common approaches. Scoping papers to be drafted and advisory boards to be set up.

Other initiatives: Framework Partnership Agreement (FPA) for Geant. cPPP's (contractual PPPs) – HPC and Big Data Value.

Other issues:

- e-IRG to suggest structure for WPG 16/17 by 4th December. New financial instruments to be discussed, quicker reactions from e-Infrastructures are expected, advisory group to meet next week (17th November 2014).
- e-IRG met DG Connect on 4th November (such meetings should continue). The Executive Board of e-IRG will also meet with Directorate Excellence in Science/*DG CONNECT*.
- Key issues: Commons, Collaboration with ESFRI, e-IRG contribution to WPG 16/17 (this role is recognised).
- Innovation Workshop took place on 3rd October. Topics: European e-Infrastructures and Innovation Clusters, SME innovation, WPG 16/17, (should there be a common action? summary report available, discussion group set up).
- Mathematics and Digital Science Workshop took place on 6th November. Topic: Mathematics for HPC/Big Data and vice versa in the context of WPG 16/17 (Only mathematicians, workshop summary will be available – there was a wealth of ideas).
- **Digital Area Forum** – it will meet 9th December.

Questions and Answers:

- Summary of Innovation workshop – it should be on the web.
- Maths workshop summary – should be available end of 2014

12:00 **Giorgio Rossi (University of Milan, Italy - vice-chair ESFRI) Towards the new ESFRI Roadmap 2016**

Workshop Notes

There is a common understanding that RI's are important as European competitiveness depends on their availability and quality. Europe's competitiveness is built on investment in education and access to first class RI's. Coordination is required to optimise the limited funds available and maximise the ROI. Data sharing between data producing RIs must be optimised.

ESFRI is working off its 2010 Roadmap. It contains 42 RI's, distributed across fields, some implemented; some others will have to go. ESFRI was mandated by the Competitiveness Council in July 2014. There will be a new roadmap for 2016 which will include two parts: a landscape analysis of RI's in the EU and international context.

There are gaps in the EU RI eco-system. There is a process for developing the new roadmap. Proposals are to be submitted through national members. There are two assessment processes: scientific and governance assessments. There will be only 25 projects in the new landscape, including 8-10 new projects for all disciplines. The landscape analysis will be completed first (to close in winter). There is one member of e-IRG in each SWG. Examples of RI's required: European Astronomy and Astrophysics/Particles, Exa –Analytical Facilities, EPOS, Health and Food Research, Customised Healthcare. The connections between the RI's will be part of the landscape work (e.g. Energy research RI's /production and efficient use of energy/ and Social and Cultural Innovation RI's will cross-connect with other domains).

Comments:

- e-IRG and ESFRI should support integrated solutions for **data producers and data management**.
- DG RTG and Connect should issue **integrated calls** for IT solutions for RIs.
- We need to participate in the definition of **common data formats**.

The ESFRI Process:

Milestones: landscape completed, Strategy Work Group analysis, scientific analysis, pan-European relevance assessment matrix, e-IRG White Paper, End of 2015 – candidate projects and emerging projects, rejected projects, Assessment of maturity – using the method of AEG-2012, Final decision by Executive Board -- 2016 Roadmap (10 years rule, 10 projects phasing out, ESFRI update 2 due in 2018).

Questions:

1/What is the exact timing? –Project presentation by the end of March 2015, then peer reviewers will be identified (neutral fully independent referees; KPIs, criteria, assessment metrics available on the website), autumn – recommendations to be presented, end of year – short list available.

2/ What should RI's and ESFRI do jointly? - Projects capable of cooperating, exchanging data and good standards for metadata, etc. New RI's should make an effort to open their science. ESFRI only produces a landscape analysis trying to indicate what features projects should have - it does not organise calls.

3/ Who is in ESFRI? - National delegations: 28 delegates and 10 associated states, 2 delegates per country. There is some financial support from the states. There is a request from the Competitiveness Council to implement the priorities in the roadmap. All projects are deemed important and for some there is no implementation method specified.

Lunch break

Chair: Enzo Valente Track 1: Outlook towards eScience and International Research Infrastructures.

14:00 Patrick Aerts (NLeSC, The Netherlands) PLAN-E: European co-operation of eScience centres

Abstract

Dr. Patrick J.C. Aerts

Director Strategic Alliances, NLeSC

The advancement of science has been strongly stimulated by the very existence of advanced e-infrastructures, such as provided by European and national back bone networks and the resource infrastructures on top of that (PRACE, EGI, EUDAT, ESFRI-facilities). Gradually, however, the focus on provisioning e-infrastructures and ICT is shifting towards their innovative deployment in science. Enhancing science and so facilitating new discoveries by (optimal) use and re-use of techniques, software,

tools and methodologies across disciplines is what the future of science, research and development will depend on. eScience is the discipline that addresses this shift in focus that goes well beyond the Big Data wave, however much this wave floods our present thoughts about conducting modern science.

In order to strengthen the position of eScience/Big Data Research as a domain per se, the aspirations of the field, the skills level of the scientists working in that field and the broader educational aspects a European platform of centers has been implemented in a constituting meeting September 29-30 2014 in Amsterdam, to bundle present knowledge and expertise across Europe and to define a practical work program towards close cooperation between centers involved in conducting eScience.

The presentation will focus on the goals and actions lines of PLAN-E.

Workshop Notes

PLAN-E is a European cooperation of e-Science Centres. Science and society are interconnected.

Example: E-Science centre in Holland with a mission to enable digitally enhanced research through use of e-science and similar tools. The application domains are those of e-Infrastructures. NLeSC in the coordinator and it is a platform of e-science and data research centres in Holland with the leading role of NWO (e-Science Centre) and SURF (HPC).

e-Science experts can apply to become members. The Platform is formed by e-Science and Data Research Centres. It will define common objectives. It is a continuation of the work of ARCADE (Advanced Research Computing Academic Discussion Group in Europe) consisting of funding and policy agencies and HPC centres. Its main output was a Knowledge Base: An Overview of HPC in Europe. Its work has been transferred to e-IRG.

Why do we need PLAN-E?

Its task is to facilitate transferring knowledge to science on how to use RI's. It will share knowledge of tools, e.g. data management. It will attract attention to and obtain support for important issues. It will form a community of e-Science users. It will foster the development of academic PhD skills that are needed at this level. It will focus on layer between the ICT and e-Science

It is a voluntary organisation with a kernel of active members. Terms of Reference will be defined as well the organisational aspects of PLAN-E will be dealt with.

Questions:

1/ Was this discussed with the participants before? We might have enough bodies in Europe. - People agree there is something missing between ICT and e-Science.

2/ Who from the countries listed joins? Organisations and not countries join – we are trying to approach the most representative organisations in each country.

Comment: Why are you using the term e-Science? It was first used in the UK, could be e-nhanced.

14:25 Federico Ruggieri (GARR/INFN, Italy) CHAIN REDS: Coordination of intercontinental infrastructures

Abstract

Prof. Federico Ruggieri

Head of Distributed Computing and Storage Department GARR

Director of Research INFN Roma Tre

Research and Education need efficient communication and innovative services that can jointly be named e-Infrastructure. Several e-Infrastructures have been deployed in different regions of the world providing services ranging from the Network connectivity to Grid, Cloud and HPC Computing. Research and Education, however, is now globalised and Virtual Research Communities can address new scientific challenges thanks to the collaboration of groups distributed worldwide. European and non-EU e-Infrastructures have thus to interoperate to address the requirements of cross-continental research communities. Coordination and harmonisation of e-Infrastructures among different regions of the world is the aim of the CHAIN-REDS project. The presentation shows the current achievements of the CHAIN-REDS project and the technical and organisational challenges that Regional e-Infrastructures have to face today and in the near future.

Workshop Notes

This project has the following international partners:

- Africa Connect tasked with improving NREN capacity and cross-border connectivity.
- Arab States Research and Education Network – an evolution of a previous project (EU funded)
- China - CSTNet, CERNET and CNGI (university, academy of science, future internet), CNGrid – HPC Network
- India – National Knowledge Network, a state-of-art multi-gigabit network, ultra-speed new, GARDUA – grid, HPC grid
- Latin America – Scalac, (EU co-funded), ROC-LA (Grid), Network HPC in all countries

The Vision of CHAIN REDS: Scientific collaboration between RIs, Promoting interoperability, Defining a path towards global e-Infrastructure eco-system.

Projects elements:

- Knowledge Base – all e-Infrastructures on a single map, now extended onto repositories of documents and data.
- Promoting interoperations, MoU's between –e-Infrastructures and local centres
- Clouds for Research and Education – Federation demo, EG Federated Cloud Force
- Federated entities
- Use Cases from various domains. Example: APHRC – Societal health, African Population Health Research Centre/Kenya. Using data for research - bettering societies, challenges in Sub-Saharan Africa, eliminating data inaccuracy within global organisations.
- Consultancy, Access to EPIC PID
- Coordination and Harmonisation

Inter-regional coordination does not exist with a few exceptions although there is need to coordinate at this level as well.

Next events – 10-11 December, Oman; 15-17 December, India

Conclusions:

- The Mission is to promote coordinated activities among regional e-Infrastructures.
- The demos implemented have proved successful
- The project has five use cases
- There is a need for Intra-regional coordination
- Peer-to-peer MoU is not enough (between regions)
- e-IRG contribution needed

14:50 Davide Calonico (INRIM, Italy - Director of Research) Metrology infrastructure for research

Workshop Notes

Time and Frequency Methodology is used in telecoms, dating, defence, etc. The WW Market – Frequency Control Devices is around € 4.5 billion/year. Measurements are based on the Metre Convention in which each member country has its own meter institute. Atomic clocks (Atomic Cs Fountains are installed world-wide – A Nobel Prize has been awarded in this area) Nobel Prizes for the clocks.

The **INRIM** clock ensemble is tasked with transferring the clock time reference to users. Satellite and optical fibre (LIFT project) are used to distribute accurate time. Timing: 4 hours to spread commercial time value, 20 days for Cs Fountains.

In the **LIFT** project Torino linked to Florence through Bologna via optical fibre. More links are planned (including London-Paris and Torino-Munich).

The users are: Science – Radio-astronomical telescopes, metrology, space geodesy, atomic physics; Industry – e.g. atomic clocks in space.

Question:

How will that compare with ESA? - Tight cooperation, complementary activities are in place. They are one of the INRIM users and the project can offer capacity to any space experiment (e.g. 20-30 optical clocks).

15:15 Ingrid Mann (EISCAT, Sweden - Head of Projects, EISCAT) EISCAT-3D

Workshop Notes

This project is concerned with Earth Atmosphere coupled to Space. It is on the on ESFRI Roadmap (Environment), its preparatory project is ending now, the full project to be launched in 2019. The project has international sites, 3 partners and 6 associated partners. The task is observing earth and charged constituents in the earth atmosphere with high power radio waves. It uses global radars, rocker and satellites. It observes e.g. space transition at polar atmosphere. Measurements happen during an 11 year solar cycle and time series data is produced with Incoherent scatter. New system will be needed: huge antennas, multi-static phased arrays, with

10K antennas per site. The RI has sites in the Nordic countries. It has established data centres with existing national RIs, including common archives.

The RI's further needs are: on-site computing, network connections, operational needs, etc. developing links to e-science and e-Infrastructures. The project needs to remain science driven and Data driven. The next step will be EInSCAT_3D Implementation.

Conclusions:

- Use common solutions
- Flexibility vs. sustainability – who pays for the time series?
- Link between research and data handling

Question:

Are observations possible without communication between sites? – The current set-up is not flexible enough, not responsive to change of mode (real-time and off-line, both modes are needed), communication helps respond to events in real-time.

Break

Chair: Françoise Genova Track 2: ESFRI projects

16:20 Massimo Cocco, INGV, Italy EPOS: European Plate Observing System

Abstract

Massimo Cocco

Istituto Nazionale di Geofisica e Vulcanologia, Seismology and Tectonophysics, Rome, Italy

The European Plate Observing System (EPOS) is a long-term plan to facilitate integrated use of data, data products and facilities from distributed research infrastructures for solid Earth science in Europe. EPOS aims to obtain a holistic, sustainable, multidisciplinary research platform that will provide coordinated access to harmonized and quality-controlled data from diverse Earth science disciplines, together with tools for their use in analysis and modelling. This integrated platform requires a significant coordination between, among others, disciplinary (thematic) communities, national research infrastructures policies and initiatives, and geo- and IT-scientists.

The EPOS mission is to integrate the existing research infrastructures (RIs) in solid Earth science warranting increased accessibility and usability of multidisciplinary data from monitoring networks, laboratory experiments and computational simulations. This is expected to enhance worldwide interoperability in the Earth Sciences and establish a leading, integrated European infrastructure offering services to researchers and other stakeholders. EPOS is promoting open access to geophysical and geological data as well as modelling/processing tools, enabling a step change in multidisciplinary scientific research for Earth Sciences. The EPOS Preparatory Phase (funded by the European Commission within the Capacities program) aimed at leveraging the project to the level of maturity required to implement the EPOS construction phase, with a defined legal structure, detailed technical planning and financial plan. The actual EPOS implementation phase will be built upon the successful achievements of its preparatory phase. EPOS will operate a full e-science environment including metadata and persistent identifiers.

In this presentation, we will describe the RIs to be integrated in EPOS and present the EPOS IT architecture in order to illustrate the integrated and thematic core services to be offered to the users. Some of the

thematic services at community level already exist and are operational. The presentation will also deal with the implications for the user community and funding agencies associated with the adoption of open data policies and access rules to facilities as well as the implications for the proper assessment of socio-economic impact of distributed, multidisciplinary RIs. We will also discuss the resources needed to tackle the challenge of fostering data driven research and big data applications. For Earth scientists, the prevalent problem is represented by the need of data, which must be promptly discovered, made accessible and downloadable, curated, minable and transferrable together with appropriate processing software and e-infrastructure resources. In general, there are a number of overlapping issues that regard data organization and their access, data transfer from (and to) supercomputing centres (HPC) and among the platforms of the federated communities. Finally, the presentation will also discuss the international cooperation initiatives and the global perspectives for solid Earth data infrastructures.

Workshop Notes:

EPOS is a single pan European distributed RI. EPOS deals with Solid Earth Sciences and it investigates the phenomena of earthquakes, volcanic eruptions, unrest episodes, tsunamis and tectonics, etc. It employs integrated use of data, models and facilities. There are various communities in EPOS, e.g. those involved in observing and monitoring volcanos. EPOS benefits the Citizen

EPOS Services: EPOS Integrated Core Services, Thematic Core Services and National RI and Facilities.

Current status: EPOS RI completed, each of the elements has a strategy. The current goal: Complete ICT Research Environment (Functional architecture). Preparation phase finished, EPOS is now entering the implementation phase; to be operational in 2020

EPOS Innovation Process: Creating Data, Processing Data, Analysing, Preserving Data, Giving Access to Data, Re-using Data. Data providers need to be involved in the process of building the RI.

Sustainability is an issue. E.g. the sustainability for various data and networks; huge amounts of data need to be integrated. EPOS is now approaching member states to present the impact of this RI and ask the governments to support.

Conclusions and Issues:

- Solid Earth Science in one RI – this is a choice that needs to be justified
- Co-design involving users to build a community
- Innovative and appropriate solutions needed to manage a pan-European RI
- Impact measurement requires tools (e.g. Technopolis approach)
- Sustainable architecture presents technical, governmental, legal and financial challenges
- EPOS has adopted a federated approach in its organisation and it is working on a federated approach to its IT solutions
- IT Procurement to start soon – Who should provide IT systems (PPP, European organisations, projects, national technology providers)?
- Convincing member states is the challenge now

Questions:

1/ Is there a US network collaborating with EPOS? – There is a global collaboration with US, although they have a different approach.

2/ What is the estimated cost? – € 90 million for national centres, they are local research centres and they already exist, € 8 million/year is needed for their operation.

3/ Are there any plans for achieving the sustainability of the entire RI among the partners? - It is a local task/national level task only; one needs to resolve the sustainability issue on national/other RI level; it depends on the maturity of other elements.

4/ What is the budget and the objective of the procurement process? – Not sure that if there will be a clear list of requirements; the EC should start this discussion.

5/ 10% investment for yearly operation is the usual level and EPOS seems to be more expensive. Why is that? – 10% needed to maintain the operations, but new RI hubs are also being built, hence the extra cost.

16:45 Rafael Jimenez (EMBL-EBI, UK - ELIXIR Chief Technical Officer) ELIXIR: The European Life-Science Infrastructure for Biological Information

Workshop Notes

ELIXIR deals with biological research data. It consists of sustainable, national centres connected together with EMBL-EBI as the main hub, on top of existing RI's. It is run by 17 EU states where there are ELIXIR nodes.

Data growth is a strategic driver. Challenges: the amount of resources (1800), non-centralised data, production data dispersed, data growth (doubling each 12 months). The technology is improving and the cost is getting lower. Network file transfer process: 1 day to produce, 4 days to transfer.

Sustaining data is a key challenge. There will be a budget shortage due to data growth.

Elixir technical activities: node activities, pilots (list available on the slide), task forces (TF): Cloud TF, Storage TF (discussions with EUDAT and EGI), AAI TF.

Collaborations with e-Infrastructures involve EUDAT, EGI and Geant.

Question

How is Biology different from other fields in that it produces huge amounts of data? Why do we need Elixir? – Our data is in many different places, stored in various ways, in different formats, etc.

17:10 Mark Hagen (ESS - Head of the ESS Data management and Software Centre) ESS: The European Spallation Source

Workshop Notes:

A key question is how e-Infrastructures help ESS? The plans of ESS: first accelerator – 2019, tools completed 2019 – 25. Application example: thermal neutron scattering (neutron and x-rays are used together, a technique rather than a field). A variety of users: medicine, batteries, chemistry). Another example: use of protons to produce neutrons.

Funding is in cash and in-kind by members. ESS will provide the equipment.

ESS organisation/Management structure includes DMSC – Data Management and Software Centre.

Data on disk is useless. Tools for analysis, computational resources, analysis are required and there are various projects to support that.

How can e-Infrastructure help?

- Providing reduced data sets
- Data analysis
- Deploying a semi-analytic model
- Using simulation

Questions:

1/ Can some parts of the processes be done remotely (e.g. sample preparation)? – Some things can have remote control, synchrotrons need 24h staffing and human beings need to be there.

2/ Is analysis done at home on downloaded data? Would it not be better to compute remotely – It needs to be shown that it works, e.g. some projects use time on the Berkeley HPC system.

17:35 Miles Deegan (SKA, UK - Engineering Project Manager SKA) The Square Kilometre Array

Abstract

Miles Deegan

Engineering Project Manager, Square Kilometre Array Organisation, Jodrell Bank Observatory, UK.

During the past 18 months the Square Kilometre Array (SKA) Project has made major advances. Since 2008, the global radio astronomy community has been engaged in the development of the SKA in a major effort - the 'Preparatory' phase of the project. The Preparatory phase ended in December 2011 and, following a number of major changes, the international SKA project has now progressed to the 'Pre-Construction' phase (2012-16). The Member Nations have set up the SKA Organisation, a not-for-profit company founded in the UK, to lead activities and the Pre-Construction work has been organised into a series of design work packages to be delivered by consortia from the Member Nations. This talk describes the organisation and scope of the work packages as the project begins the work of preparing for SKA1 construction. The talk will cover the e-Infrastructure requirements of the SKA, in particular the high performance computing (HPC) and big data aspects.

Workshop Notes

SKA is a huge user of data. SKA is a next generation radio interferometer. The current project phase (SKA Phase 1) is deploying 3 telescopes (Mid, Low, Survey) on 2 sites. They are 100x more sensitive than what is available now. Total cost is € 1.5 billion and the sites are in South Africa and Australia.

Examples of SKA Science: neutral hydrogen in the universe from cosmic dawn till now, evolution of galaxies, star formation, the cradle of life, fundamental forces.

SKA has 8 Science Working Groups. Global SKA Structure is as follows: Project office at Jodrell Bank, 11 technical Work Packages.

Transport requirements are: SKA 1 – 6 ExaBytes. There is a project called ‘Science Data Processor Consortium’ led by Cambridge University to support that.

SKA 2 will see huge changes in the number of dishes, e.g. South Africa will have 2500. The issues to resolve will be: signal transport, signal processing, software engineering, data storage.

Questions:

1/ How is data transport resolved? – This is still being worked on, perhaps there will be no direct link from one continent to another.

2/ HPC – why do you need that? Do you really need that? – Calculations done by GPU type technologies, there are commonalities there with HPC.

3/ How many atomic clocks are there per site? – No details on that, preliminary design as of now only.

18:00 Tibor Kalman (GWDG - DARIAH-ERIC - Co-Head VCC1) and Eveline Wandl-Vogt (Austrian Academy of Sciences - DARIAH-ERIC - Co-Head VCC1) DARIAH-ERIC: Towards a sustainable social and technical European eResearch Infrastructure for the Arts and Humanities

Abstract

Tibor Kalman

Gesellschaft für wissenschaftliche Datenverarbeitung (GWDG), Germany

DARIAH, Co-Chair Virtual Competency Centre 1 e-Infrastructures

Eveline Wandl-Vogt

Österreichische Akademie der Wissenschaften (ÖAW), Wien (AT)

DARIAH, Co-Chair Virtual Competency Centre 1 e-Infrastructures

DARIAH (Digital Research Infrastructure for the Arts and Humanities) aims to enhance and support digitally-enabled research across the Arts and Humanities by offering a portfolio of services centred around European research communities.

The DARIAH infrastructure is a social and technological infrastructure; it aims to be a connected network of tools, information, knowledge, people and methodologies for investigating, exploring and supporting research across the broad spectrum of the Digital Humanities and Arts.

The core strategy of DARIAH is to bring together national, regional, and local endeavours to form a cooperative infrastructure where complementarities and new challenges are clearly identified and acted upon. DARIAH is aiming to bridge the gap between traditional and digital Humanities and Arts, taking into account technical as well as social innovation.

DARIAH integrates national digital Arts and Humanities initiatives all over Europe and operates a platform to enable trans-national, interdisciplinary and trans-disciplinary research.

It offers a portfolio of services and activities centred around research communities. It develops a research infrastructure for sharing and sustaining digital Arts and Humanities knowledge.

By bringing together national activities from several countries, DARIAH will be able to offer a broad spectrum of services including training initiatives, such as summer schools and transnational curricula, a knowledge repository with standards and good practices for digital asset management, and guidance on repository certification and digitisation processes.

The DARIAH e-Infrastructure utilizes standards and best practices, allowing collaborations with several research infrastructures and offering opportunities for innovative research.

For the various affiliated projects, services for data sharing and digital publishing will be offered alongside technical systems for persistent identification, authentication and long-term preservation.

The DARIAH-ERIC was established on August, 15th 2014, by 15 Founding Members: Austria, Belgium, Croatia, Cyprus, Denmark, France (host), Germany, Greece, Ireland, Italy, Luxembourg, Malta, The Netherlands, Serbia and Slovenia.

Workshop Notes:

The Mission of DARIAH is to support digital research for humanities and arts. It is a social and technical RI. It provides access to tools, standards and it also answers questions by members and users. It supports social sciences and it cooperates with the European School on Social Innovation, creating common teams.

The challenges of DARIAH are:

- Big Data in Humanities
- Sustainability
- Citizen Science

Question:

Is the Call for participation in DARIAH open to everyone – Yes, it is open to everyone.

18:25 Panel

Prepared Questions and Panel answers:

1/ Are the ESFRI projects actively participating in RDA to develop standards for data access, data management etc.?

EPOS - Yes. We have IT people working in RDA and we are following RDA meetings.

ESS - They should do more in our community. We are only starting this work at this point.

EISCAT - Participating in related projects/proposals.

SKA - We attended a meeting between RDA and ESFRI projects in Brussels recently. But we should become more involved in these activities.

DARIAH - Yes, we are participating although the funding of this participation is an issue, people are taking part in different capacities. Funding of DARIAH is based on national activities, so they may participate under other hats.

ELIXIR - We are involved, there is an information flow between ELIXIR and RDA; RDA has a focus on generic solutions which do not always fit the needs of ELIXIR.

2/ Are the used standards and best practices in the ESFRI projects publicly accessible documented?

ESS - Partially yes. There are not fully publicly documented.

ESCAIT - We have identified some issues

SKA -None

DARIAH - Yes, partially. Standards to be developed for their provision, not every community would benefit from the publication of them.

ELIXIR - We have repository of formats and standards, which are diverse and hard to control and coordinate in order to help these communities

3/ What are your recommendations to new ESFRI projects for a successful application?

ESS - Distributed RI's are more complex. Co-design approach in which IT and data RI's work together.

EISCAT – In some cases impossible. Science comes first. The main effort should like with policy makers. Close links with Community need to be ensured. We should communicate requirements to users and get feedback.

DARIAH – Our recommendation for researchers is to have a clear focus on their objective

ELIXIR – We recommend focusing on the existing ESFRI and complete their implementation.

4/ Do you think that ESFRI projects could benefit from and contribute to innovation in the e-Infrastructure area? If yes how will this be achieved?

ESS - We are still rolling out. The answer is a yes, but not sure where we could contribute.

EISCAT – Offering a link to user communities, e.g. coordinated observation in our case and we can offer that to the RIs (as they want to use some tools in coordination with other observation); RI's are a link to user communities, they need to stay flexible. An e-Infrastructure is a service, and each e-Infrastructure should define their own customers. Keeping the RIs flexible.

SKA – Yes, we should not work in isolation and cooperation is needed.

DARIAH – Contribution should happen through user communities.

ELIXIR – use of RIs and beyond that – we are service providers and we need to adopt technology.

5/ Is the harmonization of the scientific assessment between ESFRI and e-Infrastructure providers (e.g. PRACE) thinkable to accelerate the application procedure?

ESS – It depends on the scope of the ESFRI project. E.g. we are not sure whether asking for a data plan would accelerate the application procedure.

EISCAT – Science should come first. Policy makers tend to underestimate the effort required. They should stay in contact with user communities and listen to their specific requirements. Users are scientists and not used to questions on RI's – it takes a lot of time and effort to communicate that.

Other Comments and Questions:

1/ What is the part of your RI project that could be offered to other RIs? – **(ESFRI)** – We went through complex prioritisation process. Two projects are given opening for a large extra support – a model to help other RIs.

2/ (PRACE/Sergi Girona) – e-Infrastructures are RI's as well and the principle of open access should apply here as well – there should be no difference. The most important thing is the peer review.

3/ Scientific collaboration is key.

4/ (Françoise Genova) We need to find a way to work it out in harmony.

5/ (INRIM) – We should work together on standards.

6/ (ELIXIR) – We should capitalise on the diversity of all RI's – share experiences, use cases, and solutions among the different science areas.

7/ (Chalmers, Sweden) - e-Infrastructures should pick up innovation from RI's and turn it onto a clear service that can be really provided.

8/ (PSNC, Poland) - Is there any special policy on data openness within ESFRI? – **(EISCAT)** It depends on the level of data products there, the policies of research councils are followed. **(DARIAH)** The policies are discipline-specific. They use a critical mass approach.

9/ (PSNC, Poland) - Do we need a data policy and how do we establish that assuming that RI's are supposed to be open? **(EISCAT)** The real question is what the needs for data being open are. It is not enough just to place it in an archive – there is lots of work involved. Who will pay for that? **(ELIXIR)** They have 21 data sharing policies and all data will be open but there some data that is not. The issue is where those data will end-up. Public storage might disappear due to sustainability problems.

10/ (Françoise Genova) – We should collect use cases from a wide range of communities,

11/ (DARIAH) - Do not be afraid of being collaborating with others. It is important to have places to discuss that

12/ (ELIXIR) – Projects not being RI's are not influencing the policies enough. More funding is needed for that sort of collaboration.

13/ (EISCAT) – The roadmap process requires a certain structure. Research is based on diversity and the process does not allow for that diversity.

Day 2

The Chairman of e-IRG highlights the importance of e-Infrastructure Commons. This discussion is only just a start and more coordination and discussion needed.

Chair: Ilmars Slaidins Track 3: European e-Infrastructures for Data Management and Computing.

09:00 Sanzio Bassini (CINECA, Italy - Chair of the PRACE Council) The PRACE Supercomputing Research Infrastructure

Workshop Notes

An Introduction about PRACE: PRACE is an AISBL with PRACE BOD as the management body and its four pillars: SSC representing science, Access Committee carrying out the peer review, User Forum representing users and Industrial Advisory Committee (IAC) representing industry.

There will be PRACE DAYS (annual PRACE scientific and industrial event) in Dublin in May 2015 and then the 2016 event in Prague.

PRACE also provides access to Tier – 1. It is involved in Data Management with the second pilot in this area undergoing technical evaluation, integrating EUDAT technology, in accordance with EUDAT.

PRACE Scientific Impact – 242 papers by April 2014, The PRACE h-index is 16.

PRACE will grant access to future CoE's and some resources will be reserved so that CoE's will not have to compete.

PRACE Training is implemented through PRACE Advanced Training Centres (PATCs), Code Enabling, and Summer of HPC.

The Call process includes: Preparatory Access, Project Access, Tier 1, Access for CoE's, and SHAPE.

PRACE aims to provide high quality service, attract and train talent and lead the integration of the HPC Ecosystem and also to further develop the PRACE association.

Question:

0.25% reserved for CoE's – Why would they access PRACE? – This will not be forced. It will fully depend on each CoE's decision.

09:25 Damien Lecarpentier (CSC - IT Center for Science, Finland - EUDAT Project Director) EUDAT Collaborative Data Infrastructure: future perspectives

Abstract

Damien Lecarpentier

EUDAT Project Director

EUDAT – European Data Infrastructure – is a pan-European research data infrastructure initiative funded by the European Commission. EUDAT's vision is to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment, as part of a Collaborative Data Infrastructure (CDI) conceived as a network of collaborating, cooperating centres, combining the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centres. Currently, EUDAT is working with more than 30 scientific communities and has built a suite of five integrated services – B2SHARE, B2DROP, B2FIND, B2SAFE, and B2STAGE – to assist them in resolving their grand challenges. The presentation will provide an overview of the project status and highlight some of the main challenges being addressed to achieve the CDI vision.

Workshop Notes

An update on EUDAT. It is a solution for research data challenges faced by RI's. It links community specific repositories. There are 26 partners and it is looking for more research communities. There are discussions with RI' (EUDAT started with 5, and has grown to 30).

EUDAT addresses the full life-cycle of research data:

- B2Drop – Drop Box, to be used also for sensitive data, hosted at JSC – Sync and Exchange Research Data.
- B2SHARE – Store and Share Research Data, small scale data, deposit data
- B2SAFE – Replicate Research Data Safely, optimising data access, archiving and preservation.
- B2STAGE – Get data to computation Moving data closer to HPC, access data through RESTful
- B2FIND – searching metadata

What is coming next?

- Policies, Data Access and Reuse working group – DARUP
- Business Model – who pays, etc. looking for the right revenue models
- An organisational model – using EUDAT vs. Joining EUDAT. Developing SLAs.

Issues:

- Integrating e-Infrastructure and Research Infrastructures
- e-Infrastructure Commons – single access is needed
- Bridging National and European Solutions – this needs to be better synchronised

Questions:

1/ What % of available is covered by users – We are only starting, it is quite small. TBs are being used at the moment. The resources committed are in the PB range.

2/ What is the format of data? Is any user free to store any format? How do you search? – Any format is possible, EUDAT try to map metadata on all data.

Comment: (EC) We are in favour of open access but innovation prevents open access and we understand that. Users could act as providers and we are looking into that.

09:50 Tiziana Ferrari (EGI.eu - Technical Director) Open Science Commons for the European Research Area

Abstract

Tiziana Ferrari

EGI.eu Technical Director, EGI-InSPIRE Project Director

The presentation introduces the Open Science Commons, a vision to allow researchers from all disciplines to have easy and open access to the advanced digital services, data, knowledge and expertise they need to collaborate to achieve excellence in science, research innovation.

This vision requires the contribution of e-Infrastructures and Research Infrastructures for its realization, and it requires technical integration as well as organizational harmonization, cooperation and coordination of all the players to realize synergies in procurement and provisioning of services, development of integrated business models and policies of access.

The realization of the Open Science Commons is necessary to advancement of the implementation of the ERA, the sustainability and persistency of e-Infrastructures and Research Infrastructures, and requires the development of a coordinated roadmap both at the European and national levels.

The EGI current contribution to the Open Science Commons, the future strategy and the EGI recommendations to e-IRG, ESFRI and the EC are presented.

Workshop Notes

The main challenge is: Open Science Commons for ERA.

Milestones:

- 2006 – 1st ESRI Roadmap
- 2009 - ERIC legal framework
- 2010 - EGI.eu and PRACE AISBL

Current issues at ERA:

- Avoid the digital divide
- Harmonise policies and access
- Collaboration in the increase of capabilities and capacities
- National dimension of ERA
- Long tail of science, SMEs and Industry
- Incomplete national roadmaps. Fragmented national landscapes hinder the sustainability of European dimensions.
- e-Infrastructure Commons not achieved yet
- Lack of one backbone of European ICT Capabilities
- e-Infrastructures and RIs should be part of the same research systems

The ERA Vision:

- Researchers can access all digital resources
- Open Science Commons should include all elements of the research eco-system
- Principles of the Commons should be developed
- Open Knowledge for all RI's
- Distributed competence centres
- Integrating training programme with EUDAT and PRACE
- Future – Coordinated roadmap with other RI's

EGI Strategic Actions for the OSC:

- Open data platform on a community federated cloud
- Open e-Infrastructure – which should be user driven

Recommendations for e-Infrastructure – EGI Interaction

- Collaboration for service co-design
- Knowledge transfer/training
- Service procurement
- Federated operations

Recommendations for EC:

- Consolidation of national e-Infrastructures
- Review the capacity of a European backbone ICT/Governance
- Integrated actions for RI's to work with e-Infrastructures
- Conclusion: – Open Science Commons are needed

Questions and Comments:

1/ How will the integration of EGI and Centres of Excellence happen? –A federated approach is adopted with sites all across Europe. It is a network of computing centres. CoE's would be one component of the network while EGI would look after open access, data, etc. thus integrating all CoE's into one offer.

2/ Open Science Commons should be placed in the centre and the commons spreading all over the other areas in the model presented.

3/ (HELIX/SIMULA) How are we expected to contribute? What are the next steps? – Working together with others, having one front desk, having one roadmap, one strategy in approaching SMEs and Industry (in line with EC). EGI will have a new body – new governance for RIs, Strategy and Implementation Board, with presence of SMEs and industry, fostering clusters, Open Data Movement, Big Data Value Chain – involving SMEs and industry (conclusion – we do not need to do again).

10:45 Christian Grimm (DFN, Germany - General Manager DFN) and Bob Day (Janet, UK) Geant perspectives in H2020 era

Workshop Notes

GEANT provides Networking, Services and People. Terena and Geant together become the GEANT association. Its task is the development of advanced network and e-Infrastructure services. It is included in Horizon 2020 which provides the building blocks while the EU funding as a glue.

The priorities of the Framework Partnership Agreement:

- Sustainability
- Universality
- Reliability
- Innovation

Mission:

An open, innovative and trusted information infrastructure for the European knowledge economy.

Actions needed:

- We have not identified all of users
- Keep the horizontal funding glue for e-Infrastructure – avoid geographical and disciplinary disintegration
- Encourage vertical e-Infrastructure integration, generalise and commoditise RI service stacks. There is need for more formal partnerships
- Encourage user-led national e-Infrastructure coordination
- Most users are nationally funded and services
- GEANT will be governed as a national membership body. EGI moving in that direction as well

11:10 Panel discussion & discussion with audience

Prepared questions:

1/ Have the providers made progress in establishing an e-Infrastructure Commons? What are the concrete results?

GEANT – We have made progress. (please note that the GEANT panellist had to leave earlier and provided his answers before leaving)

2/ What are the next planned steps?

GEANT – We need to take a breath and understand what we have achieved to date.

3/ How have the users been involved?

GEANT - - Yes, not direct, but they are involved

4/ What are the user experiences? (question to the audience)

GEANT – We have regular meetings and we obtain good feedback.

5/ What are the major obstacles to fulfil user's service requests? Financing, service not in portfolio, etc.

GEANT – Changing your requirements too often but we know how to address this.

6/ Do you see the necessity of an overall, centralized service desk?

GEANT – It would be good to know what this is like.

Other Questions and Comments:

- HELIX - There are no commercial industrial providers involved. Can they be included?
- EGI – There are various levels of maturity of the SMEs involved. There are opportunities in supply and use. We could use RI's to provide commercial services,
- HELIX – There will be a business model issue. We need a reliable business model on top of the policies. RI's serves researchers and we should achieve better sustainability.
- Audience Comment – Some companies receive EC funds to build e.g. computing centres.
- Enzo Valente – € 500 million on top of the € 25 million given by the EC, 5% enough to build RIs, € 10 per person per year. People will pay when things are needed and other things like PRACE will have to cost more.
- Enzo Valente – No one has succeeded to build a global system. Only physics has been successful but where are the other disciplines. Why give recommendation if you are not successful? The real difficulty is that each country has difficulties in supporting pan-European RI's. There is no consensus on shared systems. Do we really need the CoE's.
- EGI – The ESFRI roadmap enables communities to plan their RI's. There is a huge progress on the ESFRI roadmap implementation and the demand is there.
- PRACE/Sanzio Bassini – There is need for a certification approach with three main elements: HPC, Data and Infrastructure. A process is needed to commit to creating a persistent RI based on Solidarity, Transparency, and Ethics. There should be a mechanism to manage the shared resources, joint procurement in order to guarantee persistence, etc. We must involve users, based on the drivers of users.
- GEANT –The EC 'funding glue' should be in place. An example of industry involvement: T-Systems do not have a business model to serve RIs. They use us to provide with those services. Users do not want to deal directly with T-systems. There are benefits on both sides. Many other suppliers compete in this area.
- EUDAT – Why do we need e-Infrastructure Commons? Research is changing and there is dependence on various components. There needs to be a panel on what to put into the Commons.
- GEANT – GEANT has a country based membership. Is there room for European communities to become a full member? - Voting members are countries. Others can join

without a vote. We can invite two non-ERA members to the Steering Board with a vote. We are thinking of reserving two seats for users.

- Panel leader/ Ilmars Slaidins – How can we approach users?– (EGI) There are user communities which should be organised at pan-European level. We should also involve industry and SMEs and we should approach international communities together.
- PLAN-E/Patrick Aerts – Should e-Infrastructure providers come with a great master plan to have a seamless cloud of e-Infrastructures to compete with commercial providers? – (PRACE/Sanzio Bassini) – The most important element is the value added, i.e. the provision of digital e-Infrastructure, managing priorities and interaction with scientific community.

12:00 Sverker Holmgren Wrap up and conclusion

12:10 Sverker Holmgren, Enzo Valente, Ilmars Slaidins Final words, Remarks, Actions

The main conclusion: The intersection between vertical and horizontal research infrastructures needs further discussions. Please refer to the beginning of the document for further detail.

From ESFRI projects

- Relevance for science is imperative
- We want to take part in the development and delivery of e-Infrastructure services for our projects!
- Sustainable e-Infrastructures are needs. Who pays?

From e-Infrastructures

- An extensive set of services are available and operational today
- Wide/open access to data puts new (significant) demands on e-Infrastructures
- A discussion on the implementation of the e-Infrastructure Commons has started

Other Conclusions in relation to e-Infrastructure Commons:

- There is an agreement on the importance of the e-Infrastructure Commons but there are still somewhat different understandings of the concept and different initiatives (of different scope) are taken by several actors.
- The e-IRG plenum will continue to discuss and facilitate the implementation of the Commons as described in the e-IRG White Paper 2013.
- Further discussions on governance, access modes, funding streams etc. needed
- National dimension is essential for the further development of European e-Infrastructures (and other research infrastructures)