e-IRG Workshop Summary

24-25th Nov 2015, Luxembourg

The Building Blocks of the e-Infrastructure Commons

For the recordings of the talks and copies of the presentations used at this workshop, please visit: http://e-irg.eu/workshop-2015-11-programme

Contents

Executive Summary

The main objective of this workshop was to identify the building elements of the e-Infrastructure Commons, a platform uniting all e-Infrastructures into a single service point for all research needs.

Conclusions

The main **conclusions** of the work carried out at this workshop are:

- The definition of the e-Infrastructure Commons needs to be further developed and agreed on, taking into consideration the current draft (http://e-irg.eu/documents/10920/290578/e-Infrastructure+Commons+summary.pdf) and also the outcome of the workshop (below).
- In this process, an emphasis needs to be placed on the following elements:
    - The services to be provided
    - The operational mechanism to ensure the provision of these services
    - Clear interfaces to the current service provision mechanisms (i.e. the current infrastructures)
    - Definition of a funding and legal model .

The Workshop Proceedings - Summary

DAY 1 - 24 November 2015

11:00 – 11:30 Welcome addresses and invited presentations, Chair: Antoine Barthel

- Antoine Barthel (Local host)
- Welcome address e-IRG; Local Host
- Robert Kerger (Luxembourg Ministry of Education and Research)
- Short Political address from the Luxembourg Ministry of Education and Research

Rudi Balling (LCSB) - Short Keynote
The Luxembourg Centre for Systems is an interdisciplinary biomedicine research centre. It requires IT to drive its research activities but it is also the progress of IT that defines the potential of the research activities of the centre. The biggest challenge of all is **Data Integration & Data Standardisation**. The exact position of the commercial service providers needs to be defined – whether they are drivers, partners or competitors.

Sverker Holmgren (e-IRG Chair) Welcome address; The theme of the workshop.
The Chairman of e-IRG introduced the main objective of the Workshop – a discussion on the building block of the e-Infrastructure Commons.

11:30 – 12:30 Evaluation of the progress of the e-Infrastructure Commons (1st session), Chair: Sverker Holmgren

11:30 Arjen van Rijn (e-IRG), The view of e-IRG
The current definition of the e-Infrastructure Commons is (according to e-IRG 2013 White Paper):

- An ecosystem of ICT services for scientific research attained through a joint strategic effort between users and primary strategic actors and suppliers, …
- in which providers have the freedom to innovate and ...
- where users enjoy the freedom to choose the services they need from a mix of public e-Infrastructure and commercial services, so that ...
- users can focus on doing science, through (international) research collaborations, whilst avoiding spending effort on the requirements to access various services;

The Commons would span over both the ESFRI Research Infrastructures and Other Research Infrastructures and ensure the provision of Tools and Services concerning Data, Computing and Networking/Connectivity.
Our recommendations for the strategic actors of this arena are:

- International user groups (ESFRIs and other RIs): “organize yourself”
- International e-infrastructure organizations: “team up!”
- National governments and funding agencies: “make strong building blocks”
- European Commission: “strengthen, regulate, coordinate”
- Existing e-infrastructure providers: “innovate or perish”

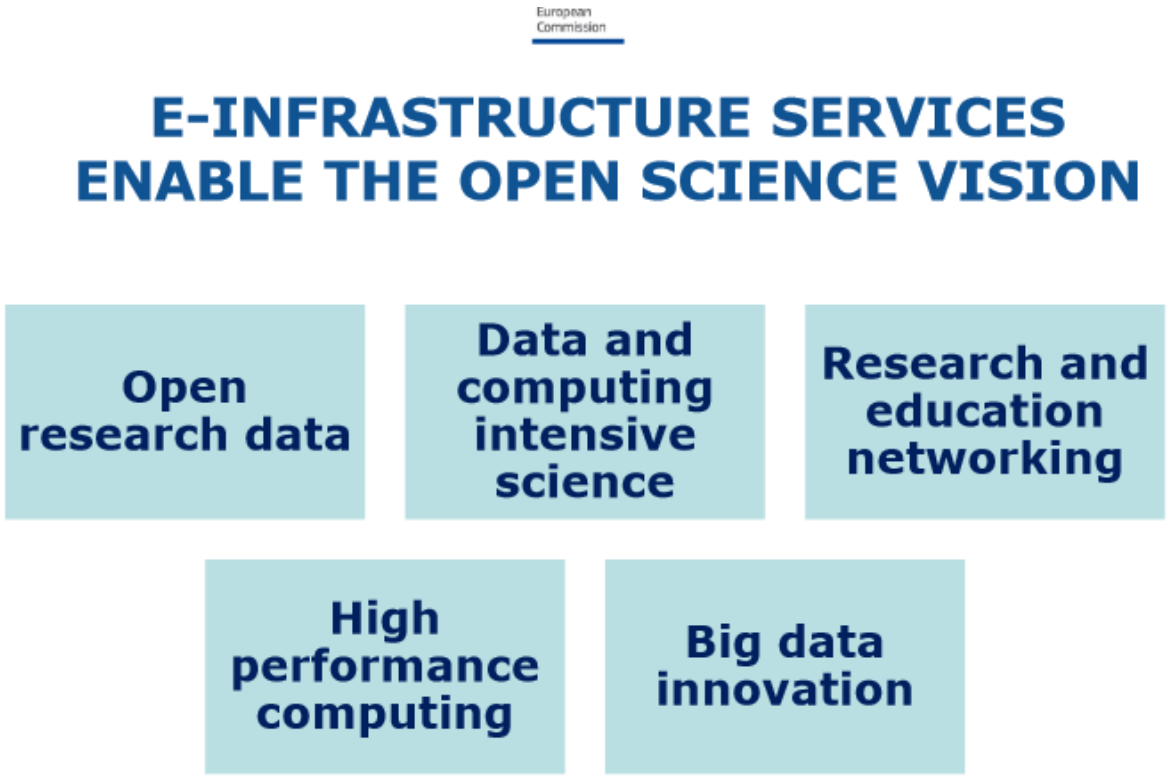
Our WP16/17 recommendations (June 2014) are:

- National e-infrastructure coordination
- European e-Infrastructure coordination
- The availability of Horizontal services
- Transnational ease of use
- The development of appropriate Commercial services

- Well organized user communities

12:00 Augusto Burgueño (EC, DG CONNECT), How the e-infrastructure Workprogramme 2016-2017 can support the e-Infrastructure Commons

This talk introduced the 2016-17 Work Programme of the EC and its relationship with the e-Infrastructures and the Commons. The current landscape of the European research support services is shown here:



The following are the Implementation Principles of this landscape:

- Service orientation through a Catalogue of services
- Maximizing the impact of e-Infrastructures - Use of KPIs for operational, technical and socio-economic impact assessment
- Innovation through co-design
- Open research data
- Participation in international fora such as the Research Data Alliance (RDA)
- Horizon 2020 as a catalyst for growth and jobs

The e-Infra WP2016-17 EC Work Programme consists of the following three themes:

The various aspects of the e-Commons are included in these Themes, e.g.:
1. A platform for the coordination of the services building the Commons– Theme 3/Policy support, KPIs and cost models
2. The provision of sustainable and interoperable services within the Commons– Theme 1/Integration and Consolidation through e.g. a Service Catalogue
3. Innovation – Theme 2/Innovation (Open, Co-design, etc.)

14:00 – 15:00 Evaluation of the progress of the e-Infrastructure Commons (Break-out sessions)
In this session, the attendees were divided into groups to discuss the definition of the concept of the e-Infrastructure Commons. The basis of the definition provided was:
An ecosystem of ICT services for scientific research, where users enjoy the freedom to choose the services they need from a mix of public e-Infrastructure and commercial services and where providers have the freedom to innovate.

The following 3 questions were asked:

- What could we do to improve the concept of the e-infrastructure commons? And: what is the difference with the European Science Cloud?
- What do you believe are essential elements of an e-infrastructure commons in terms of e-Infrastructure services? What do we mean with 'interoperable and integrated services'?
- How should we coordinate that on the European level? What about a single coordinating 'European e-infrastructure organization'?

The Working Groups were asked to relate these issues to the following stakeholder groups and develop recommendations for them:

- National governments / research funding agencies
- European Commission
- National (public) e-infrastructure providers
- European (public) e-infrastructure providers (organizations and projects)
- European / international Research Communities (RIs)
- Commercial e-infrastructure providers

The recommendations from this part of the event were presented on Day 2 and are included at the end of this document.

15:00 - 15:45 Towards a "one-stop shop" for e-infrastructure users: national perspectives, Chair: Yannis Ioannidis

15:00 Jan Gruntorad & Jan Oppolzer (CESNET) - CESNET's approach to e-Infrastructure Users
In this talk the Infrastructure Environment for R&D in the Czech Republic was presented. The country has developed a Roadmap for Large Research, Development and Innovation. Within this framework, the task of CESNET is to develop and operate the network communication infrastructure in the Czech Republic to support science and research (as a non-public operator). CESNET is a one-stop shop point for use of all services (except HPC). All academic users have open access to e-infrastructure services. There is a process for feedback from users e.g. via CESNET Users Advisory Board. Co-funding (ca 15%) is provided by user organizations based on their budget (currently no relation to infrastructure use).

15:20 David Salmon (JISC) - e-Infrastructure convergence in the UK (A Federated Approach)
The UK e-Infrastructure components include National R&E network Janet plus high-capacity peerings with:

- National commercial Internet Service Providers
- International: NRENs via GEANT & commercial Internet
- Compute & Data Storage services – internal and external plus national frameworks
- Research Council National Laboratories and Facilities
- Community resources – University/Organisation level compute & storage resources & local facilities

Some examples of the national/international members are: STFC (Science and Technologies Facilities Council) or EMBL/EBI Hinxton. On the National & Regional level such examples are: EPSRC Archer HPC or HPC Wales.
The coordination of this system takes place through the following bodies:

- E-Infrastructure Leadership Council (UK)
- UK cross-Research Council e-Infrastructure Working Group
- HPC Special Interest Group – self organised
- Big Data Special Interest Group – self organised

The Stakeholders & contributors are:

- Research Councils
- JISC
- Mature research communities

JISC activities with Research communities:

- Networking
- Research Data management work
- Security & access management Working Group
- Framework contracts for access to commercial providers
- HPC access portal - Arcus
- Datacentre framework contracts
- Collaboration – EGI, EUDAT, e-IRG

£4M e-Infrastructure funds “reserved” for industry support.

Conclusions:
A “One-stop-shop” approach in the UK is premature - too simple, the real world is complex. Evolution is needed to accommodate the “long tail” of science. A federated approach could work where appropriate.

16:15 - 17:15 Evaluation of the progress of the e-Infrastructure Commons (Re-using existing e-Infrastructures)

16:15 Steven Newhouse (ELIXIR) ELIXIR and its use of e-Infrastructures

ELIXIR is a sustainable European infrastructure for biological research data. Its tasks are to facilitate research, safeguard data and build sustainable data services. It delivers services through ELIXIR Nodes building on national strengths and priorities. The coordination is delivered through an Elixir Hub. The members of ELIXIR connect to the Hub with their national centres.
The ELIXIR infrastructure is both scientific case driven (i.e. containing solutions required by various scientific projects or areas), and platform orientated (e.g. training, tools, data, compute). It has developed links to other existing infrastructures in the areas of e.g. storage (working closely with EUDAT) or networks (working with GÉANT).

Conclusions:
Currently, the potential components of e-Commons are not integrated, not user friendly, not accessible and vary in terms of quality/capacity, etc. However, there are plenty of useful components to build on, e.g. in the field of authentication and authorisation. The European Open Science Cloud (EOSC) seems to be a potential tool for the implementation of the e-Commons.

16:45 Kay Graf (KM3NET) Projects that submitted proposals to the ESFRI Roadmap 2016

KM3NeT is a deep-sea research infrastructure in the Mediterranean Sea. Astro particle physics uses detection techniques from particle physics to study astrophysical phenomena and different messenger particles: gamma, Cosmic Rays, neutrinos.
KM3NeT is part of the ESFRI Roadmap 2016. It is built as a Tier-like structure with mixed access (GRID + direct (batch)). KM3NeT has a Data Management Plan and an Open Data Policy in place.
KM3NeT supports the creation of e-Commons in Neutrino Astronomy – a seed of such a system is already in operation (e.g. through exchange of sky maps).

17:15 - 18:00 Evaluation of the progress of the e-Infrastructure Commons (Report from the break-out sessions), Chair: Sverker Holmgren

This item was cancelled due to time constraints. A summary of the discussion that took place is available at the end of this document.

DAY 2 - 25 November 2015

09:00 – 10:30 Networking Aspects, Chair: Valentino Cavalli

09:00 David Fergusson (UK) User view on information assurance and AAI (health sector)

The development of genomic and personalised medicine is having profound effect on the bio-medical research process. Increasingly biomedical researchers have to link data relating to different aspects of a disease and representing a broad range of scales (i.e. genomic, phenotypic, physiological, population). This means accessing, synthesising and analysing disparate data sources, some of which may have significant access restrictions. To do this, a new paradigm of results as a service is required.

In order to support this demand, networks and data source providers need to be able to present a seamless access model.

There is an increase in data requirements as sequencing is expected to increase rapidly and generate 100s of Pb per annum within the next 5 years. Data can be divided into Sensitive and Identifiable Data and Complex Data.

There is a change in the dynamic – solutions need to become data centric, not compute centric. This is because data problems are harder to deal with than compute problems, or data is hard (expensive) to move. Moving into the Cloud could be an option.

Collaborative Data Centre –eMedLab – is a unique, powerful centre to build, test, deploy new infrastructure tools between Organisations (it includes HPC where the data resides)

09:30 Ann Harding (SWITCH) Need for higher information assurance and AAI to support data intensive research (NREN perspective)

SWITCHaai 2005 – is a classic identity federation. An Identity Provider (IdP) asserts authentication and identity information about users. Service Providers (SP) check and consume this information for authorization and make it available to an application. A group of organizations are coordinating IdPs and SPs that agree on a common set of rules and standards to build trust.

No researcher works in isolation – at CERN a single paper can easily have 3000 authors.

Assurance and Trust in Identity Federation Today:

- "Classic" Level of Assurance (LoA) approaches on national scales
- Cost, scalability and demand are concerns

Observations about assurance:

- Key principle decisions about access need to be made by the service
- Even where no LoA is officially supported, many criteria required by research groups are supported
- It is difficult to identify person and roles in a unique way

Conclusions:

- Federated Identity delivers scalable trust by having the responsibility for providing information closest to where it is created and managed
- Swiss edu-ID approach takes this further, putting identity in the hands of the user and integrating attributes from other parties
- Scaling assurance with detailed vetting processes over multiple jurisdictions remains a challenge
- But one with solutions now grounded in research needs

10:00 John Dyer (GÉANT) Acceptable Use Policies and Connection Policies. Help or hindrance for end-to-end research, innovation and the Open Science Cloud.

The Connection Policy:

- Defines the conditions under which organisations are eligible to access and use the network and its services
- Not yet undertaken an extensive survey & comparison of European NREN CPs

There are two observations:

- If you connect to someone, they connect to you, too
- Once connected, you normally accept all the traffic.

There needs to be an Acceptable Use Policy: e.g. 'The network may not be used by a User Organisation or its Members for any activity that may reasonably be regarded as unlawful or potentially so'.

Conclusions:

- The world & networking landscape is evolving rapidly
- R&E networks need to adequately support all its user communities
- The existing policies are significantly diverse
- RISK: following the path of the lowest common denominator
- Need solutions that serve all users, including high-end users that cannot be satisfied by the market

11:00 - 12:45 e-Infrastructure coordination and panel with e-Infrastructure providers, Chair: Arjen van Rijn

11:00 Fotis Karayannis Vision for a one-stop shop marketplace

The idea of a marketplace for e-Infrastructure services is based on the following premises:

- A concrete step towards e-Infra services integration
- Towards the e-Infrastructure Commons 2020
    - Integrated living ecosystem of resources & services
    - Open, user friendly, accessible to EU researchers
    - Continuously adapting to user needs (e.g. RIs)
- EU Open Science Cloud (instantiation of e-Infra Commons) part of EU Digital Single Market related EC Communication

What is a marketplace for researchers?

- A place for researchers to find all the services they need
- A service catalogue and/or store
- Both from e-Infra and Commercial Service Providers
- Integral part of the e-Infra Commons/Open Science Cloud

Examples:

- UK GovDigital Marketplace
- HelixNebula AppStore on cloud services
- Major e-Infrastructure projects (such as EGI, EUDAT) also working towards this direction

<u>Marketplace main features:</u>

- Single point of access (one-stop shop)
- Catalogue of services–central registry
- Search facility for easily finding relevant services
- Use common identity, authentication and authorisation schemes
- Both research and industrial ones
- Service Level Agreements

Conclusions:

- Marketplace: a first step towards the Commons and "lightweight" integration of e-Infra services

- Main features: one-stop shop, searchable catalogue, common access, research and industrial services, etc.
- Some convergence already
- Great opportunity to deliver something beneficial for the users!

PANEL DISCUSSSION

Steve Cotter (GÉANT)
The e-Infrastructures Commons should be: 'An ecosystem of ICT services for scientific research, where users enjoy the freedom to choose the services they need from a mix of public e-Infrastructure and commercial services and where providers have the freedom to innovate'.

Steven Newhouse (Helix Nebula)
The Integrated Cloud Services should include:

- e-Infrastructure Commons:
  - Commercial cloud providers able to serve researchers
  - Rationalised and integrated services
- European Open Science Cloud:
  - Build on e-Infrastructure Commons
  - Commercial cloud providers for different SLAs
  - Services that enable Open Science

Tiziana Ferrari (EGI.eu)
The e-Infrastructure Commons are a necessary component of the Open Science Commons.

- Not only "ICT services" but also policies and rules, e.g. for access and funding
- Complemented by Data as a Service, instrumentation, knowledge

The e-Infrastructure Commons Challenges:

- Bridge data preservation infrastructures and computing
- Open Science platforms: sharing of open tools, applications, scientific software, research data
- From "services" to "solutions" involving multiple providers (e-Infras, RIs, research communities, data providers…)

Florian Berberich (PRACE)
PRACE's position on the e-IRG Recommendations:

- Outreach to and involvement of user communities
  - PRACE is prepared to join forces in outreach and involvement of user communities; already in place:
    - Networking session with EGI and EUDAT at ICT
    - PRACE Scientific Steering Committee and User Forum
- Services registry, discovery and provisioning
  - PRACE is prepared to work on joint services to simplify usage of resources across infrastructures
  - The work should be based on real-life use-cases
  - Examples: Pilot Call for joint Tier-0 access and data storage services and resources, joint training events, Security for Collaborating Infrastructures
- Financial, legal, business development and procurement issues
  - PRACE has a well-defined, unique business and funding model
  - Exchange of experience is possible

Per Öster (EUDAT)
There are two challenges: Integration and Sustainability.
Integration:

- EUDAT CDI - compose & combine EUDAT technical services ("B2 Enterprise Edition")
- European e-infrastructure - agree protocols, interfaces, identity management with HPC, cloud, and networks (open EU data/compute platform for research)
- European Research - agree policies, API, and methods with universities, libraries, digital publication actors, and service companies (open science)

Sustainability:
EUDAT CDI - Create partnership of sustainable organizations and develop it (utilize the full potential of "B2 Enterprise Edition")
Financial - ensure multiple revenue streams for partners and partnership (open EU data/compute platform for research)
Societal - follow policies of governments, universities, libraries, digital publication authorities (open science, open society)

Paolo Manghi (OpenAIRE)
OpenAIRE support the creation of an Infrastructure for Open Knowledge that would:

- Foster and facilitate the shift of scholarly communication towards making science Open and Reproducible
- Collaborative and participatory approach at European and Global level
- Link people, ideas, and technologies

The Commons should provide End-User Services such as: Repository analytics and usage stats or Open Access & DMP Helpdesk, for examples.
OpenAIRE has already a registry of services with different abstractions for different users.

Panel discussion

12:45 - 13:00 Wrap-up, Final Words and Closing

12:45 Sverker Holmgren (e-IRG Chair)
Wrap-up The e-IRG Chair Sverker Holmgren, summarised the discussions bringing up the following points:

- The building blocks of the e-Infrastructure Commons have been presented several times. It is clear they are there! However, they need to be built further and become integrated towards the users.
- It is thus now time for action! The providers in the panel share the same goals to best serve the users. The users are not interested in governance or funding models and they don't care if it is one or many organisations serving them. They want to solve their problems and currently it is still difficult to find their way. They want some sort of navigation in the services.
- e-IRG is developing its new Roadmap which will include a landscape analysis and also the way to the future. It builds on top of the e-Infrastructure Commons, but also on other components and new developments/discussions including the European Open Science Cloud. e-IRG will also consult ESFRI, as ESFRI consulted with e-IRG for its Roadmap. But e-IRG is not only about big infrastructures, but also researchers, i.e. the full span of e-Infrastructures. Lastly, e-IRG is composed of delegates from Member and Associated States. And what is done at EU-level depends on the national-regional level. So there is a clear connection. This is also true for the users: as a researcher you want a coordination between national and EU systems.

12:50 Sverker Holmgren, Antoine Barthel, Arjen van Rijn

Final words Sverker Holmgren closed stating that the workshop was very helpful and the discussions will continue in Amsterdam in spring. He concluded thanking everybody involved: the organisers, the speakers and the audience.

Arjen van Rijn, as the host for the next e-IRG workshop, stated that it will take place in Amsterdam 9-10 March followed by the ESFRI Roadmap 2016 launch event.

Antoine Barthel thanked all involved for participating and for actively contributing to lively discussions.