e-IRG Workshop Summary

3rd June 2015 – Riga, Latvia

Open Science and e-Infrastructures

Executive Summary

The *e-Infrastructure Commons* needs to be part of Europe's strategy to create the **Digital Single Market (DSM)**[1]. The impact of Open Science on e-Infrastructure and the role of e-Infrastructure in this process should be assessed in order to enable that inclusion.

Main conclusions:

The following issues need to be tackled:

- The funding of Open Access to scientific publications, research data, software, methods, educational material and other resources needs to be addressed, in particular the provision of support for access and storage/maintenance
- New ways of measuring the impact of Open Science are required.
- The creation of the European Open Science Cloud for Research needs to be addressed (as an incarnation of the e-Infrastructure Commons).
- A framework to be able to share infrastructure costs is essential.
- There needs to be a close working relationship with ESFRI to address the coordination process of Member States investment strategies in e-Infrastructures: A joint ESFRI-e-IRG Working Group needs to be established.
- The e-Infrastructure Commons has another dimension that is Global, EU, regional, institutional, etc.

Glossary

AAI - Authentication and Authorisation Infrastructure

EC – The European Commission

e-IRG – e-Infrastructure Reflection Group

ESFRI - European Strategy Forum on Research Infrastructures

ICT – Information and Communication Technology

[1] A Digital Single Market Strategy for Europe: http://ec.europa.eu/priorities/digital-single-market/ and related EC Communication http://ec.europa.eu/priorities/digital-single-market/docs/dsm-communication_en.pdf

KPI – Key Performance Indicator

PPP - Public-Private Partnership

RI – Research Infrastructure

ROI – Return on Investment

Introduction

A total of around 80 persons attended the e-IRG Workshop in Riga, which was organised part of the Latvian EU presidency of the European Union.

This workshop addressed the Policy Aspects of Open Science for e-Infrastructures. Its objective was to discuss the transition to Open Science, with an emphasis on the requirements it will impose on and the consequences it will have for the e-Infrastructure policies. The topics discussed were the policies on governing the access to data, computing and networking resources, sharing of resources and the transfer of knowledge, the perspectives of the Latvian EU presidency, the EC Consultation on Science 2.0 results, and the policy implications for e-infrastructure when Open Science is introduced.

09:00 - **Sverker Holmgren (e-IRG Chair)** opened the workshop and introduced its main concepts. The main challenge ahead is the preparedness of the e-Infrastructures to accommodate the **Openness of Science** (and thus Research Infrastructures) and, conversely, the impact of Open Science on e-Infrastructure. The idea of e-Infrastructure Commons as a *one –stop shop and an innovating system for new services* has been widely accepted, but further work is required in order to define this concept. A simplified view of the e-Infrastructure Commons places this layer at the bottom of all the RI's and other international projects, with a multitude of actors and providers. The objective of the e-Infrastructure Commons is to open science and research through the use of ICT, make science more efficient and transparent and achieve a higher ROI on the effort put in by society. The key challenges to be addressed include:

- How to secure the funding needed, in particular in the provision of access and storage.
- How to ensure that all researchers are duly credited

09:10 - **Lauma Sika, Counsellor, Attaché (Research and Space questions), Permanent Representation of the Republic of Latvia to the European Union.**

Lauma Sika introduced the priorities of the EU under the Latvian presidency, having set an agenda for a competitive, digital and engaged Europe, along with the news from the Competitiveness Council. The Council held a policy debate on open and excellent science, as a follow-up to the Science 2.0 public consultation carried out by the Commission. And one of the key outcomes was the idea of a European Open Science agenda. Europe is in the process of creating a **Digital Single Market,** and the development of a European Open Science Programme is also one of the priorities. This process will contribute to the further development of the ERA. .Open Access could further increase the use of public funding. In addition, science metrics need to be reflected upon including an upcoming revision of the EU

copyright law. Investment in open access research data intensive research should be also improved. Finally, in its recent meeting, the EU Competitiveness Council invited ESFRI (and e-IRG) to explore mechanisms for better coordination of Member States investments in e-Infrastructures.

09:10 - **Ilmars Slaidins (e-IRG Co-Chair), Riga Technical University, Latvia**

Ilmars Slaidins made a reference to the tradition of technological research and teaching at the Riga Technical University, i.e. the location of the event.

09:30 - **Jean-Claude Burgelman, HoU A6, DG RTD, European Commission**

Keynote: EC Consultation on Science 2.0 results

There is a paradigm shift in the ecosystem of research infrastructures and research data. The Science 2.0 public consultation launched by the EC aimed to assess the following: 1/the degree of awareness amongst the stakeholders of the changing modus operandi, 2/ the perception of the opportunities and challenges, 3/ possible policy implications and actions to strengthen the competitiveness of the European science and research system. The term that has been accepted to name the change is **Open Science**[2]. New ways of measuring the impact of Open Science are required. Most respondents recognise the trends included in the consultation paper.

Availability of digital technologies and their increased capacities has been identified as the most powerful drive of this change, followed by 'Researchers looking for new ways of disseminating their output.'

The top barriers for 'Science 2.0' at the level of individual scientist are as follows: 1/ Limited awareness of benefits of 'Science 2.0 for researchers, 2/ Lack of integration in the existing infrastructures, 3/ Lack of credit-giving to 'Science 2.0', 4/ Concerns about quality assurance and 5/Lack of financial support.

The top implications of Open Science on for the economy, society and the research systems are: 1/ Data-intensive science as a key economic driver, 2/ Faster and wider innovation, 3/ Science more efficient, 4/ Science more reliable (e.g. re-use of data) and 5/ Greater scientific integrity.

Policy intervention is needed in the following areas: 1/ Open access to publications, 2/ Open access to research data, 3/ Research Infrastructures, 4/ Assessment of research quality and 5/ Alternative reputation system.

The objectives of a possible future policy initiative should be: 1/ Support big data (infrastructure) needs – includes governance, 2/ Improving Framework Conditions (removing barriers, creating incentives) for fostering Open Science, 3/ Making science more efficient

[2] A question was asked at this session on the sensitivity of some of the data that might be disclosed within 'Open Science'. The response of the EC was that work is taking place on exclusions and exceptions from that policy.

(better use of and sharing of resources), reliable (replicability/re-use of data) and more responsive to societal challenges.

The European Open Science Agenda should include the following potential actions (under consideration):

- Fostering Open Science: Creating incentives and removing barriers, e.g.
  - Establish a stakeholders forum at European Level and a self-regulation/ clearinghouse mechanism for addressing Open Science issues
  - Propose a European "code of conduct" setting out the general principles and requirements of how Open Science should affect the roles, responsibilities and entitlements of researchers and of their employers
- Mainstream Open Access to publications and data, e.g.
  - Consider extending the Horizon 2020 pilot on Open Access to data
  - Develop EU guidelines for addressing IPR issues and the funding of data-management
- Introducing Open Science actions to address common societal challenges under the European Research Area and under Horizon 2020
- Develop data infrastructures for Open Science, e.g.
  - Mandate the development of common interfaces and data standards
  - Coordinate at European Level the funding/ maintenance and interoperability of research infrastructures
  - Support the development of a European Open Science Cloud for data, protocols and methodologies

EC will also support the development of a **European Open Science Cloud** for research, which will:

- Provide all EU researchers a virtual environment with free, open and seamless services for data storage, management, analysis and re-use, across disciplines.
- Federate existing and emerging horizontal and thematic data infrastructures, effectively bridging today's fragmentation and ad-hoc solutions

10:00 - **Anders Flodström, EIT ICT Labs**

Keynote: Open Innovation and Open Science - Brainchildren

Anders Flodström tackled the topic of "Education & Innovation & Jobs" - Still Silos'.

'Higher Order Skills' are needed in order to secure a harmonised growth. We experience an explosion of communication and data – big business, big data, big science and also a disruption of our industries in the form Industry 4.0 (Cyber-Physical production), Internet of Things, 3D "printing", clean technology, renewables, etc. Today's technical progress is a lot faster than social and cultural advances. The new paradigm involves the move from a producer to a consumer economy, from efficacy to innovation, and from quality to dream. Innovation is becoming the leading factor in the economy, generating 75 % of the economic growth.

There is a huge lack of ICT skills in Europe, while most European economies are experiencing high unemployment rates. Innovation creates and destroys jobs. Job change cycles are much

faster than educational cycles. Re-education and multiple competences are required now; jobs change faster than people can change. Universities must become a stakeholder among others in the new knowledge society. This is a shift from a Knowledge Society to a **Skill & Competence Society**. The new concept of Knowledge Triangle includes Research, Education and Innovation and as the key factors adding value to the economy, society and science. Massive Open Online Education (MOOC) could be a mechanism to address the skills gap. Examples of the job areas in demand in the new World are: A Digital World – the Cloud and DNA, A World of Seamless and Ubiquitous Communication and Internet of Things. In the coming ten years millions of new jobs will be created and millions of old jobs will disappear, while jobs will change faster than education.

11:00 – 12:30 **Session 1: Open Science - Directions and main issues**

Objective: Present main directions on Open Science along with main issues or areas that may need policy actions.

11:00 - **Sami Niinimaki, Ministry of Education and Culture, Finland**

Finnish Open Science and Research Initiative 2014–2017

Finland now has an updated roadmap for Open Science and RI, mapped onto the ESFRI roadmap, which is included in its **Open Science and Research 2014-2017** strategy. This initiative sets to increase the quality and competitiveness of Finland's research and innovation system. By increasing openness in research, we will simultaneously be improving reliability, transparency, and the impact of research. Openness also creates opportunities to participate in scientific advancement, and enables easier and more effective utilisation of research results. Promoting open science and research requires not only extensive involvement from the research community, but also cooperation and coordination, internalising new ways of working, and developments in research environments, researcher services and research infrastructures.

Its main recommendations are: 1/ Improvements of access to and collaborative use of research infrastructure and 2/ Shoring up of the funding base of research infrastructures.

Finland's main objective is to incorporate open science and research to the whole research process to improve the visibility and impact of science and research in the innovation system and society at large. The main challenges to be addressed are: Creating ownership, Availability of infrastructures, Harmonization of metadata, Open access, Licence policy, Cultural change towards openness, International collaboration.

In an Open Science maturity assessment basic information from openly available material on the web has been collected to assess the open science operating culture. This methodology has five maturity levels.[3]

[3] A question was asked on the different levels of RI maturity and the methodology used. Sami Niinimaki explained that there was still little methodology in this metrics, although other countries were interested in adopting a similar system.

11:30 - **Wolfram Horstmann, University of Goettingen, Germany and LIBER**

Beyond Big Bata — The Long Tail of Research

Take Home Message: Open Science needs to be on the ground with people.

Universities and research institutions are identifying research data as an asset and invest in collecting and providing access to those datasets generated at their institution that do not fall within the scope of other discipline-based, or government repositories. This "long-tail" (e.g. FigShare, My Experiment, '50% of research is done on data sets that are less than 1 GB)' can pose challenges for those institutions, including the diversity of disciplines, the variety of standards, the multitude of projects and persons involved. Yet, the support of the long-tail is paramount —not the least because the highest innovation capacity in research is in the emerging fields — not in the already established mainstream. The challenges are data quality, discoverability (with 50% of data in the university lab and 39% on the university server) incentives, and creating a feasible business case and improving access to datasets. RDA has established a Long Tail of Research Data Interest Group[4] to tackle this issue. The main conclusions in this area are: simplification to Big Data – otherwise, Europe's potential will not be utilised, data diversity, the institutional perspective needs to be fostered.[5]

Q: Health science – data sets along a long time, each group has access to a small part, the assumptions underlying the foundations of data are key, how do you define the process, A: health is very compliant, the stakeholders should produce the standards for big data,

12:00 - **Mark Parsons, Rensselaer Polytechnic Institute, USA and RDA Secretary General**

Data Policy for Open Science

Open science requires open data. While that is a truism, it can lead us to consider more specific considerations around data policy. Data policy should serve the interests of the organisations sponsoring the data collection. With research data the sponsor is often governments so the interests should be those of the public, and it is the public that ultimately benefits from open science. The Vision of RDA is: Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society. As its Mission, RDA builds the social and technical bridges that enable open sharing of data. RDA as a Policy Test Bed – it is not a policy organisation, but it can help implement policy

As such, at a first level, data policy should require data to be as **open as ethically possible**, but policy needs to address others issues as well. These issues include **data access requirements**, **data preservation** and **stewardship requirements**, **standards and compliance mechanisms**, **data security issues**, **privacy and ethical concerns**, and potentially even **specific collection**

---

[4] e-IRG has initiated a similar task force to tackle the broader issue of the Long Tail of Science.

[5] The following observations were made in this part of the discussion: 1/ data is alive and there should be processes in place to accommodate its evolution and 2/ 'long-tail' is not small data but rather the data produced by a small research group that does not need huge resources, hence the issue of creation and sustainability and 3/ health scientists should produce their own standards for big data in order to define the foundations of data and all the underlying assumptions.

protocols and defined data flows. It is also helpful to clarify and recognize **specific roles** in the data ecosystem.

Preservation requirements are well defined in the Open Archive Information System (OAIS) Reference Model, but no similar model for access requirements (Not even a common definition of "access" and what restricts it). Some areas (e.g. biomedical) require unique access. The key recommendations are: Mind your preservation and access—your stewardship, Clarify and credit roles, Promote and empower the champions—those who add generative value, Look for consensus and emergent norms from the data science community, Iterate.[6]

13:30 - 15:00 **Session 2: Open Science - Use Cases**

Objective: To present some concrete use cases on Open Science, including best practices, challenges and plans. New ESFRI/Research Infrastructure cluster projects and e-Infrastructure data projects will be introduced focusing on their approach regarding Open Science. The issue of open software licenses will also be tackled.

13:30 - **Mark Allen, CDS Strasbourg, France**

Open Science in the framework of the ASTERICS Astronomy ESFRI cluster

Astronomical research is facing the challenges of combining data from many different telescopes. The Horizon 2020 'Astronomy ESFRI & Research Infrastructure Cluster' (ASTERICS) addresses the cross-cutting synergies and common challenges shared by the various Astronomy ESFRI facilities (SKA, CTA, KM3Net & E-ELT). ASTERICS will enable astronomers to have broad access to the data products of the ESFRI telescopes via a seamless interface to the Virtual Observatory framework. Embracing the concepts of Science 2.0, open access, and open innovation, ASTERICS aims to increase the scientific impact of the telescopes and greatly encourage use (and re-use) of astronomical 'big data' in new and novel ways.

The 'Virtual Observatory' will include archives and databases form a 'digital sky'. It will open new possibilities via data discovery, efficient data access and interoperability, driven by exploding data rates as well as multi-wavelength, time domain & survey science.

The benefits of being open and interoperable within Science 2.0 are:

- Transition in the way Astronomy is done
- Opening up the research process
- Access, Interoperability
- Engagement – scientists, data providers, citizens

The astronomical research community is willing to lead the way with biggest infrastructures as participants in defining the VO framework. The main challenges in this process are: Sustainability, Support for openness, Keeping things simple while enabling complex

[6] The following observations were made following this talk: 1/ Access to data should not be based on ROI (which could be questioned by some stakeholders) but on ethical issues. It has been proved that Open Access generated more ROI and security/confidentiality issues are often overstated. Those who do not participate will not benefit from Open Access and there are huge advantages of Open Data.

capabilities, Interface between domain-specific & generic infrastructure, Community awareness, visibility, recognition.

13:50 - **Stuart Lewis, Edinburgh University Library, UK**

Research Library Support for Open Science - supporting the transition

With the move towards open science, research libraries can play a key role in aspects such as open access publishing, supporting research data management, and providing access to digital collections of information.

The Library has a role in coordinating infrastructure and services that have been made available to every researcher at the University of Edinburgh. It assists with undertaking well-managed research, having been driven by local University and national policies towards open science. This includes a suite of services and tools to create data management plans, store and sync data, archive it safely, publish it openly, and share it with the world. In particular, these tasks include: Facilitating Open Access publishing, Support Research Data Management, Providing access to digital collections.

14:10 - **Natalia Manola, University of Athens/ATHENA Research Center, Greece**

Text mining: the next data frontier. An infrastructural approach

Recent years witness an upsurge in the quantities of digital research data, offering new insights and opportunities for improved understanding. Text and data mining is emerging as a powerful tool for harnessing the power of structured and unstructured content and data, by analysing them at multiple levels and in several dimensions to discover hidden and new knowledge. However, text mining solutions are not easy to discover and use, nor are they easily combinable by end users.

OpenMinTeD aspires to enable the creation of an infrastructure that fosters and facilitates the discovery and use of text mining technologies and interoperable services. It examines several use cases identified by experts from different scientific areas, ranging from generic scholarly communication to literature related to life sciences, food and agriculture, and social sciences and humanities. OpenMinTeD text mining tools, services and associated resources will run on the cloud, requiring an in-depth optimization of service deployment and execution via scalable VM-based service distribution and use of distributed storage. One of the needs identifies is the need to **share infrastructure cost.**

14:30 - **Neil Chue Hong, Software Sustainability Institute, UK**

Open software for (open) science

Modern research relies on software. However, it can be difficult to understand the implications that the way in which software is licensed can have on its impact and uptake.

Open source is now de facto for infrastructure software. Open source encourages Exploitation, Reproducibility and Robustness, Reuse. Open source helps support open science. “Publicly

funded research […] is a public good, produced in the public interest, which should be made openly available" – (UK Research Council). Open Source should fulfil the following conditions:

- Non-discriminatory - Research should not be restricted or siloed
- Access to source code - Research should be transparent, robust, and accessible
- Redistribution of software - Providing access to the widest possible community
- Removing barriers to reuse - Research should encourage building on the work of others, and giving them credit

15:00 - 15:30 **Coffee Break**

15:30 - 17:00 - **Session 3: Open Science - The policy perspective and relevance to e-Infrastructures (PANEL)**

Objective: To present Open Science policy perspectives at EU/regional/national level and relevance/implications on e-Infrastructures (again at EU/regional/national level). The viewpoints of national initiatives, EC, ESFRI and of key e-Infrastructure data projects will be explained.

Francoise Genova, CDS, France, Panel Chair

Panellists: Jean-Claude Burgelman DG RTD-EC, David Bohmert, ESFRI representative, Swisscore, Bob Day, Janet and JISC (replacing Josh Howlett, Chair of the NREN Trust & Identity Committee, JISC), Iveta Gudakovska, Latvian representative in OpenAIRE project, Library University of Latvia, Sami Niinimaki, Ministry of Education and Culture, Finland

Panel: short statements from member who did not make a presentation

- David Bohmert, ESFRI:
    - There is a need to move from a project-approach, towards a whole landscape approach and work closely with the EC.
    - More a distributed approach, rather than single-sited centralised approach. The income model for such distributed infrastructures is worth looking for.
    - Sufficient use of scarce resources is needed.
    - Re. Gold Open Access to publications, publishers will get extra income.
        - Stuart Lewis: If institutions are going to pay Gold OA charges, subscription costs *must* go down.
        - David Bohmert: All stakeholders need to work closely together on this.
- Iveta Gudakovska, Latvian representative in OpenAIRE project:
    - Creation of data infrastructures is needed at national or other levels.
    - We should work together at different levels, defining clear policies, clear support for scientists and establish a system of license.
    - OpenAIRE membership is very positive.
- Bob Day, Janet and Jisc:
    - The UK government strongly supports the Open Science model. JISC is strongly involved in that, as already reported by Stuart Lewis.
        - Trust and Identity need to be worked out and involve these frameworks beyond their initial intention.

        - Security is another important area, in particular in terms of IPR protection, privacy, ethical frameworks.
    - We need to make effective use of what we already got. Awareness raising, training, and adaptation to user needs.
        - For software an e-Infrastructure leadership board is required.
        - Effort also needed to make e-Infrastructures compliant.
    - Effort is needed towards EU integration, so as to be able to do actions once, rather at all different levels.
        - Exploit the funders' mandate, so that EC and Member States work closely to bring together these infrastructures seamlessly.
        - Sustainability is crucial for e-Infrastructures.

Panel discussion

Questions and Answers-Discussion

- Norbert Meyer: How can we convince people to be involved in Open Science?
    - David Bohmert: This is what we have been doing in ESFRI since 2000. Now the Competitiveness Council has asked us in its conclusions to assess the investments strategies in e-Infrastructures and the coordination among Member States. A confederated approach is very promising.
- Jean-Claude Burgelman: Principle of reciprocity[7] applies also in the case of the Open Science Cloud: be part of the integrated cloud only if you are contributing.
    - Bob Day: There is the subsidiarity principle also, which says that only the tasks that cannot be performed at the local/national level, should be performed at the EU level. So, when it comes to the users we have to check first at national level and sometimes local level.
- Mark Parsons: Is there a role for industry in the e-Infrastructure environment?
    - [There is no answer from the panel] Per Oster: Yes, industry has a role, primarily for the long tail of science. But it also brings complications, mainly because of the terms of use. They can turn down a service without any explanation. While for e-Infrastructures sustainability is very important. So the key message is that a cooperation between research e-Infrastructures and industry is needed. A blend of the two to guarantee sustainability and innovation!
- Jean-Claude Burgelman: There are barriers in the universities and research centres. They should allow new models such as the cloud to be used by their users.
- Sverker Holmgren, referred to the presentation by Wolfram Horstman on the long tail of science, questioning that there is no Nobel prize coming out of big data; yet there is a Nobel prize in software[8]. Still, it not only about developing the software but also maintaining that software (which is very challenging).

[7] In international relations and treaties, the principle of reciprocity states that benefits that are granted by one state to another, should be returned. Source: wikipedia

[8] http://www.wired.com/insights/2013/10/and-the-winner-of-the-nobel-prize-in-software-is/

- Bob Day: To capture these assets (such as software) is one part of the problem; we also need understand them (discover, understand what they are supposed to do and what they are doing and manage all this).
- Jean-Clause Burgelman: If all the fragmented needs of the long tail of science are aggregated, then they will become visible.
- Dimitrios Koureas: The technical challenges are not so important. We have e-infrastructures. Do we have users that are using these e-infrastructures? Yes! Do we have pioneering communities? Yes! Do we have users not using them? Yes! Can we afford going ahead without them? No! Two ways of involving them: One way is an institutional way and other way is a more thematic way (domains/disciplines). We do not talk enough about these and we need to fund enough of these!
- David Bohmert: We need to go ahead of institutes and domains! Towards education and knowledge society!
- Tiziana Ferari: There are 7 ESFRI RIs talking about their own clouds! It is important to understand how these initiatives are funded at national level. This may be an action for e-IRG! In this way confusion will be avoided. It is maybe important to create an e-IRG working group about the national level funding and the coordination with other levels.

16:45 - **Panel Chair/Rapporteur**

Main summary points:

- There is always diversity: no one size fits all.
- There is a need for a system view covering all levels.
- People are always key! In addition, training, awareness raising and incentives for rewarding users are also crucial!
- The e-Infrastructures need to engage all types of users: individual researchers, communities, citizens, even kids from schools! So e-Infrastructures need to be built with the data user needs in mind. In this way they will become relevant and usable.
- Things are forgotten when they work, and this should be the case with e-Infrastructures that should gradually become transparent to its users being able to use them for their research.
- Sustainability of e-Infrastructures is key for the users to invest in them/use them.
- The subsidiarity principle applies in EU affairs and e-Infrastructures. Users should look first at the local level before going at the EU level.
- Research e-Infrastructures should be combined with services from industry; the first to serve demanding users and guarantee sustainability and the second to serve commodity needs and the long tail.
- The distributed/con-federated approach and related business/funding models need to be studies.
- Education –kids is the future for everything! Involve them in e-Infrastructures.
- Data/software stewardship is crucial. It is not only about developing software/storing data, but maintaining them in order to be fully usable.

16:55 - **e-IRG Chair, Sverker Holmgren - Wrap-up**

The e-Infrastructure Commons has a third dimension that is Global, EU, regional, institutional, etc. There needs to be a close working relationship with ESFRI to address the coordination process of Member States investment strategies in e-Infrastructures: A joint ESFRI-e-IRG Working Group needs to be established.

The next workshop is on 24th Nov and 25th Nov (ending at lunch) in Luxembourg. Organised by e-IRG/RESTENA (NREN)/LCSB (System Bio). The topic is 'The future landscape of e-Infrastructures. '