



e-IRG Workshop Summary

8-9th of June 2017, Malta

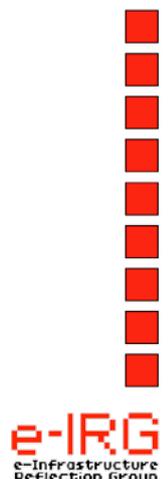


Progress of the e-infrastructure commons

For the presentations and copies of the recordings of the talks from this workshop, please visit:
<http://e-irg.eu/workshop-2017-6-programme> and <https://vimeo.com/album/4322392>

Glossary of Terms

AAI - Artificial Artificial Intelligence
CapEx – Capital Expenditure
CERN – the European Organization for Nuclear Research
CIO – Chief Infrastructure Officer
CoS – Catalogue of Services
CSA – Coordination and Support Action
DANS – Netherlands Institute for Permanent Access to Digital Research Resources
DevOps – Development Operations
EDISON – EDISON Project: Building the Data Science Profession
EDSF – Data Science Framework
EMBL- EBI – European Molecular Biology Laboratory- European Bioinformatics Institute
EOSC – European Open Science Cloud
ERIC – European Research Infrastructure Consortium
ESFRI – European Strategy Forum on Research Infrastructures
ESIF – European structural and investment funds
FAIR – Findable, Accessible, Interoperable, Reusable
FP7 – Framework Programme 7
GEANT – the pan-European data network for the research and education community
H2020 – Horizon 2020
HEAnet – Ireland's National Education and Research Network
HEP – High Energy Physics
HPC – High-Performance Computing
HPTC – High Performance Technical Computing
IaaS – Infrastructure as a Service
IoT – Internet of Things
KPI – Key Performance Indicator
LTDP – Long Term Digital Preservation
LTS – Long-term sustainability
NREN – A National Research and Education Network
OpEx – Operating Expenditure (Cost)
OSPP – Open Science Policy Platform





RI – Research Infrastructure
SD – Science Demonstrator
SKA – Square Kilometre Array
SNIC – Swedish National Infrastructure for Computing
SRC – Swedish Research Council
SWOT – Strengths, Weaknesses, Opportunities and Threats (Analysis)
WG – Working Group
WLCG – Worldwide LHC Computing Grid

Contents

Glossary of Terms	1
Executive Summary	3
The Workshop Proceedings - Summary	3
DAY 1 – 8 th of June 2017	3
10:00 - 10:30 Sergio Andreozzi (EGI, OSPP), "Governance and financial schemes for the EOSC"	3
10:30 - 11:00 Laurence Lenoir (BELSPO, Belgium and ESFRI), "Report from the ESFRI Long Term Sustainability (LTS) Working Group (WG)"	4
11:30 - 11:50 Benjamín Sánchez Gimeno / Juan Miguel González-Aranda (MINECO,Spain), "LifeWatch ERIC and sustainability plans"	4
11:50 - 12:10 Serge Bogaerts (PRACE), "PRACE: Sustainability of an ESFRI Landmark"	5
12:10 - 12:30 Hans Karlsson (SNIC Sweden), "Scandinavian model on e-Infrastructure long-term sustainability"	5
14:00 - 14:15 Andres Steijaert (GEANT), "Clouds on the ground - making clouds accessible through pan-European framework agreements"	7
14:15 - 14:25 Garvan McFeeley (HEAnet), "Case study: GEANT IaaS procurer"	7
14:25 - 14:35 Dario Vianello (EMBL-EBI), "Case study: procurer from the HNSciCloud project"	8
14:35 - 14:45 Marc-Elian Begin (SixSq), "Case study: SME cloud service provider in HNSciCloud PCP - via video"	9
14:45 - 14:55 Jurry de la Mar (T-Systems), "Case study: cloud service provider in GEANT IaaS and HNSciCloud PCP "	9
14:55 - 15:10 Bob Jones (CERN), "Cross-border procurement of e-Infrastructure services: Opportunities, Barriers, Use cases, Best Practices (based on the study supported by the EGI-Engage project)"	9
16:00 - 16:20 Paolo Budroni (e-IRG), "Assessment framework for the development of KPIs for the evaluation of e-Infrastructures"	11
16:20 - 16:40 Rafael Jimenez (ELIXIR), "ELIXIR KPI framework and alignment of KPIs between e-Infrastructures and research infrastructures"	11
16:40 - 17:00 Alasdair Reid (eInfraCentral), Ad Emmen (e-IRGSP5), "e-Infrastructure assessment strategy"- joint presentation by eInfraCentral and e-IRGSP5	11
DAY 2 – 9 th of June 2017	12





09:30 - 09:45 Ann Harding (Switch), "Current status and sustainability plans of Authentication and Authorisation Infrastructure" [pdf]	12
09:45 - 10:00 Yuri Demchenko (EDISON), "Training for sustainable skills development" [pdf]	13
10:00 - 10:15 Jamie Shiers (CERN), "The user perspective: An EOSC service pilot view" [pdf]	13
10:15 - 10:30 Prof. Kristian Zarb Adami (SKA), "SKA report on long term storage preservation requirements"	15
11:30 - 11:50 Juan Miguel Gonzalez-Aranda (e-IRG), "Long term preservation of research data (e-IRG draft document)"	15
11:50 - 12:10 Ingrid Dillo (DANS), "Data Archiving and preservation - Sustainability perspective"	15
12:10 - 12:30 Erik Fledderus (SURF), "The Go-FAIR initiative"	16

Executive Summary

The workshop had four main topics.

1. **Long-term sustainability.** The structure and financing/funding issues were discussed by representatives of the EOSC and ESFRI.
2. **Procurement and financial schemes of digital services for research.** Both procurers and providers were invited to present their case studies.
3. The third topic concerned the **evaluation of e-Infrastructures and related Key Performance Indicators (KPIs)**. A generic framework for the development of KPIs and specific KPIs from ELIXIR were presented.
4. **Long-term sustainability in relation to services and data** (similar to 1.). Related issues such as the training of new professionals, the user perspective, initiatives such as Go-FAIR, etc., were tackled.

The Workshop Proceedings - Summary

The following are the main points of the discussion that took place.

DAY 1 – 8th of June 2017

9:30 - 10:00 Opening and welcome address by **Gabriele von Voigt** who is the new chair of the e-IRG since 1 January 2017. A welcome address to the participants by the CIO of the Ministry of Education followed.

Session 1: Long-term sustainability: structure and financing/funding

- [Sergio Andreozzi](#): Governance and financial schemes for the EOSC, [pdf]
- **Laurence Lenoir**: Report from the ESFRI Long Term Sustainability (LTS) Working Group (WG), [pdf]
- **Benjamín Sánchez Gimeno / Juan Miguel González-Aranda**: LifeWatch ERIC and sustainability plans [pdf]
- [Serge Bogaerts](#): PRACE: Sustainability of an ESFRI Landmark [pdf]
- [Hans Karlsson](#): Scandinavian model on e-Infrastructure long-term sustainability [pdf]

10:00 - 10:30 Sergio Andreozzi (EGI, OSPP), "Governance and financial schemes for the EOSC"

Sergio Andreozzi initially defined the European Open Science Cloud (EOSC). The European Cloud Initiative is based on three pillars: 1. EOSC; 2. European Data Infrastructure (EDI), and 3. Widening access and building trust.





The European Commission applies a multi-stakeholder approach. The EC will launch a new High Level Expert Group EOSC in 2017-2018. The Open Science Policy Platform (OSPP) has 8 priorities for a system change: 1. Reward system; 2. Measuring quality and impact: altmetrics; 3. Changing business models for publishing; 4. FAIR open data; 5. EOSC; 6. Research integrity; 7. Citizen Science; 8. Open education and skills.

The OSPP has published an EOSC Report¹. The main recommendations are:

- 1) EOSC should rely on multi-level (integrate from local to global and national and European levels) and multi-stakeholder (RIs - including e-Infra, researchers, policy makers, funders, private sector) governance.
- 2) Facilitate access to the EOSC across borders and disciplines. Analyse all aspects of interoperability and translate into a common model and rules of participation.
- 3) European countries and EC should ensure long-term funding of the services needed to enable the integration and access to federated resources of EOSC. Develop federation of systems to address the need for shared services to work together. Long-term funding (no project lifecycles) is needed.
- 4) Different and innovative funding schemes should be considered to support users in availing of the services of EOSC-certified providers that are approved based on a commonly-agreed European certification scheme. Different business models need to be explored. Develop incentives to open up infrastructures, data and services.

10:30 - 11:00 Laurence Lenoir (BELSPO, Belgium and ESFRI), "Report from the ESFRI Long Term Sustainability (LTS) Working Group (WG)"

The EC has launched a public consultation on long-term and sustainable investments. ESFRI has created a working group long-term and sustainable investments (LTS) with dedicated Terms of Reference. The final report will be published in autumn 2017.

The main recommendations of the LTS report are:

- Establish and maintain excellence through the entire life cycle of RIs by all appropriate means, by securing adequate framework conditions, and by opening the RIs up to the world.
- Ensure that RIs have the right people in the right place and time by strengthening and harmonizing national research and educational systems to make sure that all essential skills are available.
- Harmonise and integrate a vision for convergent operation of RIs and e-infrastructures in Europe to ensure cost-effective service provision to the user communities.
- Fully exploit the potential of RIs as innovation hubs by incorporating strategies for their development into national and European innovation policies.
- Set up effective means of determining the economic and wider social value of RIs, and incorporate these benefits into science-policy-society dialogues.
- Establish adequate framework conditions for effective governance and sustainable long-term funding for RIs at every stage in their life-cycle, together with effective management.
- Foster broader coordination at national and European levels when designing processes for planning and supporting national and pan European RIs and so enhance their strategic value.

Regarding these main recommendations, the report added several specific recommendations towards the European Commission and national governments.

11:00-11:30 Coffee break

11:30 - 11:50 Benjamín Sánchez Gimeno / Juan Miguel González-Aranda (MINECO, Spain), "LifeWatch ERIC and sustainability plans"

LifeWatch ERIC (European Research Infrastructure Consortium) is a pan-European e-Science distributed

¹ <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>





Infrastructure focused on how to measure the impact of Global Change issues on Earth Biodiversity and Ecosystem Research. This RI handles Big Data from many diverse origins via the construction and operation of Virtual Research Environments.

This RI has been built by using synergies among its 'satellite' components – in order to accomplish a distributed and federated RI. The countries participating contribute in-cash and in-kind, while other funding arrives from supporting projects and initiatives (e.g. Structural Funds and H2020). Since March 2017 LifeWatch has been operating as (the 14th) ERIC.

LifeWatch cooperates with "distributed" Centres in cooperating countries (including their regions). These Centres develop and operate virtual and physical media and other components.

This sustainability approach requires the following elements:

- "Realistic" Construction and associated Business Plan based on "real" needs and users' requirements.
- "Feasible" Governance Scheme and Funding Strategy in order to guarantee the RI Sustainability based on an early identification of "in-kind" and "in-cash" by Member States (including their Regions-RIS3) contributions, consisting of at least a 5-years period "Cash Flow" Analysis.
- The Business Plan follows a sequence of releases, starting with a down-scaled level and an incremental & iterative construction methodology ("Agile"-based) with a proper Quality Assurance & Risk Management Plan associated. Establishment of the proper Working Methodologies, including Technical Advisory and Operational Committees.

Conclusions – the following should aid "realistic" constructions and granting the sustainable operation of new-born facilities.

- It is very important to establish efficient multi-layer (at regional, national and European level) governance mechanisms that include the inter-regional perspective to better coordinate the different initiatives related to RIs, and above all, to guarantee their SUSTAINABILITY.
- European Commission and National Authorities should set specific guidelines for the regional authorities interested in mobilizing European structural and investment funds (ESIF) in the construction and operation of ESFRI RIs and their use, combined or sequential, in different stages of the RIs construction and operation.
- The creation of new RI must be performed through an iterative and incremental update process involving the existing distributed facilities (e.g. "Agile" Methodology).
- It is essential to perform a periodical assessment of the impact of international cooperation processes related to the development of Research Infrastructures

11:50 - 12:10 [Serge Bogaerts \(PRACE\)](#), "[PRACE: Sustainability of an ESFRI Landmark](#)"

Seven years after its establishment, PRACE, the pan-European HPC research infrastructure, only e-Infrastructure listed as a Landmark in the ESFRI Roadmap 2016, has just entered its second phase. Meant to create a fertile basis for the sustainability of the infrastructure, we will present the underlying principles and agreements that have led to the ratification of the so-called PRACE 2 Programme. We will then share some considerations on a longer-term sustainability based on the on-going work of bodies of the organisation. This preparation of the future phase of the infrastructure is performed in the perspective of the European Cloud Initiative launched by the European Commission on 19 April 2016.

12:10 - 12:30 [Hans Karlsson \(SNIC Sweden\)](#), "[Scandinavian model on e-Infrastructure long-term sustainability](#)"

The Swedish National Infrastructure for Computing (SNIC) is developing its future long-term funding model – SNIC 2.0. SNIC makes available resources for large scale computing and storage, as well as provides advanced user support to make efficient use of the SNIC resources. SNIC is being funded by the Swedish Research Council (SRC) and six partner Universities. It is the second largest national research infrastructure.





SRC funds it with 120 MSEK/year and the six partner universities co-funds (in-kind) with approximately 50 MSEK/year.

This model has the following advantages:

- Efficient use of resources by high-level coordination of investments nationally
- SNIC has the Swedish mandate in the international collaborations
- Some amount of pooling of competences within SNIC
- A distributed infrastructure provides redundancy
- The six partner universities are involved in the activities and projects
- The user support is naturally close to the users at the partner universities

This model has the following flaws:

- Lack of long term funding decisions makes investments difficult
- Large scale investments depends strongly on in-kind co-funding by the partner Universities
- Some Universities with large use of SNIC resources are not part of and do not contribute to SNIC
- Strong SNIC-centres may sometimes promote local interests and local solutions which can make the organization less flexible
- The small SNIC office makes SNIC very dependent on the availability of competences at the centers

The current funding period from the Swedish Research Council ends in 2017 giving a natural opportunity for renewal and change of governance, funding and collaboration within SNIC. A new funding model is needed to:

- provide a long term sustainable e-infrastructure for large scale computing and storage
- handle changing and increasing needs for the current users
- be flexible enough to cater for new user groups and workflows
- make advanced user support for an efficient use of the resources available to all users
- involve all ten main universities in SNIC, both with regards to local support and funding

According to SRC, the new SNIC should:

- be governed by a consortium of (at least three) Universities providing 50% of the budget. The remaining 50% of the budget should be provided by the Research Council.
- provide generally available national services for large scale computing and storage, that cannot be made available on a local scale
- provide advanced user support for an efficient use of the resources

The components of this funding plan are:

- A clear vision has to be developed to gain acceptance for SNIC 2.0.
- SNIC 2.0 should take on a national leadership role to support excellent science.
- A broadened scope is necessary, particularly to addressing different aspects of the data life cycle.
- A risk analysis needs to be developed to demonstrate an awareness of the risks in moving from SNIC to SNIC 2.0.
- The governance structure has to be developed to ensure a sustainable and stable organization that is able to take strategic decisions, make prioritizations and provide sufficient commitment in case of difficult economic conditions.

SNIC 2.0 will adopt a Service-Oriented Approach:

- a balanced set of national services for large-scale computing and storage/management of active data sets,





- a coordinated effort within user support, facilitating the efficient use of the national and relevant international services by existing and new research communities.
- National services and interfaces to international services are acquired by SNIC 2.0 under SLA-like agreements
- Complemented by an extended effort on support towards use of national/international services (not connected to physical computing/data resources)
- It will include 1/ General services, aiming at supporting all areas of research and 2/ Specialised services, set up in collaborations with other research infrastructures or large-scale research collaborations

Other remarks:

- Governance by the ten largest Swedish Universities will provide a long term sustainability for SNIC
- The 50/50 funding model between the Consortium and the Research Council will give incentives for a cost efficient infrastructure catering for all users
- The members of the Consortium are also responsible for the large data generating research infrastructures. This opens up for the possibility to form a coherent e-infrastructure for Swedish research.

Session 2: Procurement and financial schemes of digital services for research

- **Andres Steijaert**: [Clouds on the ground - making clouds accessible through pan-European framework agreements](#) [[pdf GEANT+HEAnet](#)]
- **Garvan McFeeley**: Case study: [GEANT IaaS procurer](#) [[pdf GEANT+HEAnet](#)]
- **Dario Vianello**: Case study: [procurer from the HNSciCloud project](#) [[pdf](#)]
- **Marc-Elian Begin**: Case study: [SME cloud service provider in HNSciCloud PCP - via video](#) [[pdf](#)]
- **Jurry de la Mar**: Case study: cloud service provider in GEANT IaaS and HNSciCloud PCP
- **Bob Jones**: Cross-border procurement of e-Infrastructure services: Opportunities, Barriers, Use cases, Best Practices (based on the [study supported by the EGI-Engage project](#)) [[pdf](#)]

14:00 - 14:15 **Andres Steijaert (GEANT)**, "[Clouds on the ground - making clouds accessible through pan-European framework agreements](#)"

And

14:15 - 14:25 **Garvan McFeeley (HEAnet)**, "[Case study: GEANT IaaS procurer](#)"

Cloud services are becoming core to research and education institution's business. More and more institutions want to use cloud services on a larger scale. However, the cloud brings a paradigm shift in distribution and usage models. This requires a new approach to connect the supply and demand sides and establish the right conditions of use.

In GÉANT, the European NRENs (National Research and Education Networks) collaborate on clouds. They share expertise and resources, align roadmaps and jointly engage the market, with the goal to enable Research and Education institutes in Europe to adopt cloud solutions in an easy and safe manner. GÉANT deliver to 10.000 organisations with 50 million users. It has adopted a collective hybrid multi-cloud approach. The organisations associated establish together right conditions of use: services that are safe and easy to use, accessible, affordable and interoperable. It is a single digital single market with many cloud services, scaling up national community cloud solutions to a European level.

In this presentation, Andres Steijaert and Garvan McFeeley summarise this GÉANT cloud collective and zoom in on a pan-European tender for IaaS (*Infrastructure as a Service*) solutions.

The inputs in the network bear the following requirements:

- Data needs to be handled safely and meet European and national regulations
- The acquisition and use of services takes place through the institutions' structures:



- There needs to be an affordable and predictable cost and purchasing model
- Network traffic costs need to be limited
- Network integration: Log-in to happen with institutional account

The expectation of the user is to **be able to move data** (to another provider).

AWS and GÉANT are bringing the AWS Educate program to the European Research and Education community.

AWS will make available free AWS credits (vouchers) to students and educators, for hands-on experience of AWS cloud, access to AWS technology, open content for courses, training resources, and a community of cloud evangelists.

Access to AWS Educate takes place through the GÉANT InAcademia service.

Example: HEAnet (Ireland's National Research & Education Network). Its main focus is the Higher Education Sector, it has 21 Main Clients, 70 Clients in total and it integrates schools broadband.

HEAnet began IaaS Preparations in 2016. IaaS required a comprehensive business model – a definition of the service, cost structure, pricing and billing.

14:25 - 14:35 [Dario Vianello \(EMBL-EBI\), "Case study: procurer from the HNSciCloud project"](#)

Cloud computing is deeply re-shaping how institutions - and individual teams and researchers - can meet their compute requirements. However, especially when important amounts of possibly sensitive data are involved, the entry barriers to the clouds are still quite high. The "Helix Nebula - The Science Cloud" (HNSciCloud) pre-commercial procurement aims to foster the uptake of cloud computing by the European scientific community removing - or at least reducing - these entry barriers. During this short presentation, we will cover the base ideas behind the HNSciCloud project and how EMBL is engaging with it and with public clouds more in general.

EMBL-EBI is Europe's home for biological data services, research and training. It is a trusted data provider for the life sciences and part of the European Molecular Biology Laboratory. It operates as an intergovernmental research organisation with 600 members of staff from 57 nations. It is also the seat of the ELIXIR Technical hub.

The (HNSciCloud) Hybrid Cloud model brings together:

- Research Organisations
- Data providers
- Publicly funded e-infrastructures
- Commercial cloud service providers

In a hybrid cloud with procurement and governance approaches suitable for the dynamic cloud market

In the HNSciCloud Pre-Commercial Procurement process 3 EMBL use-cases have been identified:

- PanCancer - Builds a dataset to enable researchers to compare data across cancer types, ~ 2800 samples
- EuroBioImaging - Image data repositories and analysis tools
- ELIXIR - Long Tail of Science - To simplify access to quality-controlled data, services and tools for researchers in all life-science disciplines.

Why the Cloud?

- As data increases, have the community bring their compute to data but not all their compute to our data centre!
- Need to push relevant data sets and services out to cloud providers: EMBL-EBI Embassy Cloud → ELIXIR → EOSC → AWS/GCP/MSA → ?
- Hybrid cloud an approach to optimise EMBL-EBI CapEx vs. OpEx (allow CapEx to lag demand & use OpEx to manage peaks)
- The Cloud is a cool place: virtually unlimited resources & unlimited scalability



- Build a playground, monitor everything, and then iterate

HNSciCloud - what next?

- Now in Prototype phase (end December 2017)
- Requires an interesting amount of involvement from all the parties
- Once completed, will hopefully lower the entry barriers to the Cloud:
- Data transparency layer
- Federated AAI
- Procurement frameworks
- Expand the breadth of our pilots
- Training, training, and some more training
- Fully port applications to the cloud: investment needed!
- Contribute & pull in any outstanding results out of:

14:35 - 14:45 [Marc-Elian Begin \(SixSq\)](#), "Case study: [SME cloud service provider in HNSciCloud PCP](#) - via video"

Lessons learned in building, offering and operating interoperable multi-cloud brokerage service for scientific communities". Marc-Elian will share challenges and successes in building this type of service, and highlight the unique role SMEs can play in this field.

SixSq is a company of software artisans, expert in Cloud Computing and Application deployment automation. It provides solutions for Edge Computing, DevOps, Big Data, Smart City and IoT

This SME is involved in the following H2020 projects:

- CYCLONE: Complete dynamic multi-cloud application management
- PaaSword: Secure storage in the cloud
- SCISSOR: Security in SCADA and Smart Grids
- mF2C: bring cloud computing capabilities closer to the end-device and users
- EU-SEC: studying ways to improve approaches to trust, assurance and compliance in the ICT market
- Cloud for Europe (FP7): cloud brokerage for public administration

The company is heavily leveraging this investment in HNSciCloud with a view to contribute to the European Open Science Cloud.

Conclusions:

- HNSciCloud PCP is a key stepping stone towards EOSC
- Our SME status provides high rate of innovation and necessary independence between buyers and providers
- Volume is required to make the eco-system sustainable and
- Early adopters have a unique opportunity to shape it as anchor tenants

14:45 - 14:55 [Jurry de la Mar \(T-Systems\)](#), "Case study: cloud service provider in GEANT IaaS and HNSciCloud PCP "

T-Systems is a provider of IT solutions (e.g. Lufthansa check-in and some public services. e.g. EU – running parts of their infrastructure). T-Systems is the GEANT contract provider for Germany. They develop prototypes in the area of cloud computing and HPC (High-Performance Computing). T-Systems has prepared a SWOT analysis on their strategy (and tenders) in relation to the Open Science Cloud.

14:55 - 15:10 [Bob Jones \(CERN\)](#), "Cross-border procurement of e-Infrastructure services: Opportunities, Barriers, Use cases, Best Practices (based on the [study supported by the EGI-Engage project](#))"

This report is based on the following use cases:

- Helix Nebula Science Cloud Pre-Commercial Procurement





- GÉANT tender for Infrastructure as a Service solutions
- The Commons Credit Model

Three concepts have been used in this process:

- Cloud Credits Pilot: A Business Model to Support the Use of Cloud Computing for the Commons
- Service Lifecycle:
 - Defines platform service lifecycle that will govern how services enter into production and are eventually retired
 - Different phases will require different procurement approaches
- The Hybrid Cloud Model

Analysis led to an identification of a set of potential opportunities for cross-border procurement:



Key questions for stakeholders:

- who pays?
- what risks are the participating parties ready to accept?

The 2016 edition of the e-IRG roadmap recommends a service marketplace approach for implementing an e-Infrastructure commons. As part of the e-needs gathering process for the ESFRI 2018 Roadmap, e-IRG can help drive the market analysis by quantifying the scale and variety of services that RIs will need over the medium term

Key conclusions:

- There are a number of on-going procurement activities which are breaking new ground, gathering valuable experience and lessons learned that can serve the community for the future
- Collective procurement has advantages of reducing the time, effort and risk while increasing cost effectiveness for individual organisations in the community which often lack procurement experience
- Bringing together this valuable experience in the context of the EOSC will help organise our community and increase the added value of the EOSC itself.

15:10 - 15:30 Panel discussion with the speakers

Session 3: Evaluation of e-Infrastructures and relates Key Performance Indicators (KPIs)

- [Paolo Budroni: "Assessment framework for the development of KPIs for the evaluation of e-Infrastructures" \[pdf\]](#)





- [Rafael Jimenez](#): "[ELIXIR KPI framework and alignment of KPIs between e-Infrastructures and research infrastructures](#)" [[pdf](#)]
- [Alasdair Reid](#), Ad Emmen: "e-Infrastructure assessment strategy"- *joint presentation by eInfraCentral and e-IRGSP5* [[pdf eInfraCentral](#)] and [[pdf e-IRGSP5](#)]

16:00 - 16:20 [Paolo Budroni](#) (e-IRG), "[Assessment framework for the development of KPIs for the evaluation of e-Infrastructures](#)"

Generally, KPIs follow policies and resulting legislation and regulations. The community started as pan-European scholarly community but it is continuously expanding (geographical, application-specific coverage and industry/government, etc.). Regarding the governance for the e-Infrastructure Commons, several aspects are important: policy-aspect for e-infrastructures (access: open, restricted, closed; open science); all stakeholders are represented and have a say in definition of governance rules; open approach (not necessarily free).

e-Infrastructures are evaluated by certain KPIs: operational success (technical and operational indicators; scientific outcome); user expectations (user satisfaction, user development, service requests, e-Accessibility and barrier free indicators); general public expectations (knowledge transfer, socio-economic impact, innovation aspects).

As the next step, eInfraCentral and e-IRGSP5 are working on the topic (see also <http://e-irg.eu/kpi>).

Members of the KPI editorial Board are Erik Fledderus (delegate, chair WG), Françoise Genova (delegate), Paolo Budroni (delegate), Gabriele von Voigt (chair e-IRG), Jan Wiebelitz (member support project), Marcin Lawenda (member support project).

16:20 - 16:40 [Rafael Jimenez](#) (ELIXIR), "[ELIXIR KPI framework and alignment of KPIs between e-Infrastructures and research infrastructures](#)"

This presentation introduces the work in progress of ELIXIR on KPIs to assess and evaluate ELIXIR services focusing on 5 thematic areas: data, tools, compute, interoperability and training. This presentation aims to highlight relevant work on KPIs that could be reused and jointly developed in collaboration with e-IRG and other research infrastructures. ELIXIR is the European distributed Research Infrastructure for biological research, which provide data services essential to enable, sustain or enhance biological science.

Services are evaluated but also projects and groups. It is not just about performance but also about impact, quality, the adoption of best practices. Indicators for data, tools, training, interoperability and compute (in this order) are used to assess relevance, usage, reliability, sustainability and impact.

Five facets are assessed by several indicators: 1. Scientific focus and quality of science (indicators: archives vs knowledge base, scope statement, international dimension, staff effort); 2. Community served (indicators: overall and potential usage, usage in research, dependency of other resources); 3. Quality of service (indicators: use of identifiers, data throughput, technical performance, use of standards, data availability, customer service); 4. Legal and funding infrastructure and governance (indicators: scientific advisory board, legal framework supporting Open Science, Privacy policy, Ethics policy, sustainable support and funding); 5. Impact and translational stories (indicators: counterfactual analysis, accelerating science, translational data). The last one is difficult to measure.

16:40 - 17:00 [Alasdair Reid](#) (eInfraCentral), Ad Emmen (e-IRGSP5), "[e-Infrastructure assessment strategy](#)"- *joint presentation by eInfraCentral and e-IRGSP5*

As part of the process of developing a Catalogue of Services (CoS) and the eInfraCentral portal, the project team will collect and structure key performance indicators (KPIs) and related data from the e-Infrastructure





service providers. In doing so, the project will take account of both the current state of the art of monitoring performance by the e-infrastructure project and build on the work of, and collaborate with, the e-IRG. We are also taking account of the views of service providers and users currently being collected through an online survey (open until 19 June):

https://www.surveymonkey.com/r/e-infra_service_catalogue

Ad Emmen presented KPI activities in e-IRGSP5. Usually H2020 call KPIs will be reflected in the projects KPIs. e-IRG is working on KPI classification. The aim is also to do some recommendations to projects to standardise KPIs in order to be able to make comparisons easier.

e-InfraCentral is looking at users while e-IRGSP5 is not.

Alasdair Reid presented the eInfraCentral project, which is a 30-month H2020 CSA project. Its aim is to develop a harmonised service catalogue and a one-stop-shop portal, also monitoring service performance and quality.

17:00 - 17:30 Panel discussion with the speakers

DAY 2 – 9th of June 2017

Session 4: Long-term sustainability: services & data

- [Ann Harding: "Current status and sustainability plans of Authentication and Authorisation Infrastructure"](#) [pdf]
- [Yuri Demchenko: "Training for sustainable skills development"](#) [pdf]
- [Jamie Shiers: "The user perspective: An EOSC service pilot view"](#) [pdf]
- [Prof. Kristian Zarb Adami: "SKA report on long term storage preservation requirements"](#)
- [Juan Miguel Gonzalez-Aranda, "Long term preservation of research data \(e-IRG draft document\)"](#) [pdf]
- [Ingrid Dillo, "Data Archiving and preservation - Sustainability perspective"](#)
- [Erik Fledderus, "The Go-FAIR initiative"](#) [pdf]

09:30 - 09:45 [Ann Harding \(Switch\), "Current status and sustainability plans of Authentication and Authorisation Infrastructure"](#) [pdf]

This presentation looks at the current and future deployments of (Authentication and Authorisation Infrastructure) AAI supporting research, supported by insights from "Realising the European Open Science Cloud". The focus is on the complexity aspects of researcher use cases and the implications this has for interoperability between different e-Infrastructures and Research Infrastructures.

AAI as seen from "Realising the European Open Science Cloud" – this task presents the following challenges:

- Build on existing capacity and expertise where possible.
- The major technical challenge is the complexity of the data and analytics procedures across disciplines rather than the size of the data per se.
- Generic protocols – ICT standards of the e-Infrastructure communities
- Combine single sign-on issues with the connection of social and professional people oriented web applications resulting in a federated identity and credentials for all people in the EOSC.





09:45 - 10:00 [Yuri Demchenko \(EDISON\), "Training for sustainable skills development" \[pdf\]](#)

The education and training of Data Scientists currently lacks a commonly accepted, harmonized instructional model that should reflect the whole lifecycle of data handling in modern, data driven research and the digital economy. This talk will present the EDISON Data Science Framework (EDSF) developed as a part of the EU funded EDISON project that is intended to create a foundation for the Data Science profession definition and work as a basis for the whole ecosystem of the skills management and capacity building involving academia, research, industry, and public sector. The EDSF includes the following core components: Data Science Competence Framework (CF-DS), Data Science Body of Knowledge (DS-BoK), Data Science Model Curriculum (MC-DS), and Data Science Professional profiles (DSP profiles). The MC-DS is built based on CF-DS and DS-BoK, where Learning Outcomes are defined based on CF-DS competences and Learning Units are mapped to Knowledge Units in DS-BoK. In its own turn, Learning Units are defined based on the ACM Classification of Computer Science (CCS2012) and reflect typical courses naming used by universities in their current programmes.

The presentation will provide examples of the target use of the proposed EDSF for different components of the sustainable skills development and management: (a) defining Data Science and Data Analytics competences and skills required by organisation and building tailored staff development and re-

/upskilling programmes; (b) designing effective Data Science curricula and training programmes adjusted to university resources and needs of stakeholder organizations; (c) designing job advertisement and assessing candidates competences and skills;

The presentation will also provide information about EDSF adoption and EDISON project cooperation with multiple projects and initiatives in Europe and worldwide, including Champion Universities community piloting Data Science programmes in Europe.

10:00 - 10:15 [Jamie Shiers \(CERN\), "The user perspective: An EOSC service pilot view" \[pdf\]](#)

The goal of the EOSC Pilot Science Demonstrator (SD) on Data Preservation is to demonstrate “best practices” regarding data management and their applicability to long-term data preservation, “open” data sharing and re-use. It is modelled on the existing CERN Open Data portal but using “open” services rather than High-Energy Physics (HEP) specific ones.

The CERN time-line is based on the following principles:

- Robust, stable services over several decades
- Data preservation and re-use over similar periods
- “Transparent” and supported migrations

The Data Preservation Demonstrator:

- Goal is to demonstrate “best practices” regarding data management and their applicability to LTDP + “open” sharing + re-use
- Equivalent to CERN Open Data Portal but using “open” – i.e. non-HEP – solutions

For a user (community) to go “shopping around” to find the right services, resources and support is a challenge/impediment. More (and more complex) services needed to support data processing, distribution and analysis (full data lifecycle=WLCG4LHC).

HEP (high-energy physics) data characteristics:

- Digital information - The data themselves, volume estimates for preservation data of the order of a few to 10 EB
- Other digital sources such as databases to also be considered





What Makes HEP Different?

- We throw away most of our data before it is even recorded – “triggers”
- Our detectors are relatively stable over long periods of time (years) – not “doubling every 6 or 18 months”
- We make “measurements” – not “observations”
- Our projects typically last for decades – we need to keep data usable during at least this length of time
- We have shared “data behind publications” for more than 30 year, (HEPData)

“Data” Preservation in HEP

- The data from the world’s particle accelerators and colliders (HEP data) is both costly and time consuming to produce
- HEP data contains a wealth of scientific potential, plus high value for educational outreach.
- Many data samples are unique, it is essential to preserve not only the data but also the full capability to reproduce past analyses and perform new ones.

This means preserving data, documentation, software and “knowledge”.

User requirements / expectations:

- (Large) user requirements often exceed available resources /budgets (and existing resources typically fully utilised)
- Service expectations (e.g. max 10’ downtime) quasi-impossible to achieve

Benefits of collaboration: LTDP

1. The elaboration of a clear “business case” for long-term data preservation
2. The development of an associated “cost model”
3. A common view of the Use Cases driving the need for data preservation
4. Understanding how to address Funding Agencies requirements for Data Management Plans
5. Preparing for Certification of HEP digital repositories and their long-term future.

What are the right metrics?

- As easy to use as Amazon?
- Cheaper (and better) than doing it in-house?
- A majority of ESFRIs use it as their baseline?
- How about a 5-star scale for “Open Science: Open to the World”?

LTDP Conclusions:

- As is well known, Data Preservation is a Journey and not a destination.
- Can we capture sufficient “knowledge” to keep the data usable beyond the lifetime of the original collaboration?
- Can we prepare for major migrations, similar to those that happened in the past? (Or will x86 and Linux last “forever”)
- For the HL-LHC, we may have neither the storage resources to keep all (intermediate) data, nor the computational resources to re-compute them!
- You can’t share or re-use data, nor reproduce results, if you haven’t first preserved it (data, software, documentation, knowledge)





10:15 - 10:30 [Prof. Kristian Zarb Adami \(SKA\), "SKA report on long term storage preservation requirements"](#)

The Square Kilometre Array (SKA) project is a radio interferometer spread over two continents which will provide unprecedented views of the radio sky. The scale of this project is such that the data rates produced by the telescope are more than 10x the global internet traffic today. These large data rates necessitate the development of new 'real-time' processing algorithms as well as novel data storage architectures and facilities. The SKA project has developed some approaches to dealing with these data rates and is very much looking forward to having fruitful discussions with the panel as to how the community can begin to deal with these large scientific instruments.

10:30 - 11:00 Panel

11:00-11:30 Coffee break

11:30 - 11:50 [Juan Miguel Gonzalez-Aranda \(e-IRG\), "Long term preservation of research data \(e-IRG draft document\)"](#)

e-IRG has developed a document containing guidelines for the long-term preservation of research data. These guidelines are intended to show a set of technical recommendations, methodologies and standards, providing technical details on the recommended practical implementation. The document addresses five main "themes" consisting of "guiding principles" that should be applied to guarantee the preservation, accessibility, and usability of research data in the long term:

- State-of-art, what is long-term preservation
- What is available
- What is needed
- Costs associated to long term preservation
- Possible solutions and further recommendations

The importance of long-term preservation of data is fast becoming one of the main concerns of large research initiatives (including associated Infrastructures). It goes beyond the data, and extends to their (meta-) data preservation and curation, and therefore, including the quality of research (meta-)data, as data are often accessible via metadata, and thus ensuring metadata quality is a means to provide long term accessibility.

For that purpose, preservation of data for long-term use will require data management strategies that include curation and preservation planning and implementation. While data management and curatorial activities have been an integral part of some scientific domains for years (see for example, astrophysics and high energy particle physics), these are new concepts in other areas of science (including inter-disciplinary such as Climate Change ones-related). Concepts such as provenance, representation for re-use, and work-flow capture are rarely understood, let alone addressed.

Therefore, preservation of research data for long-term use requires careful planning, and would benefit from some new approaches, which are presented in the document.

11:50 - 12:10 [Ingrid Dillo \(DANS\), "Data Archiving and preservation - Sustainability perspective"](#)

"The what, why and how of long-term data preservation". Governments are developing open science and open data policies and funders are mandating data sharing. What are the incentives for researchers to comply with these policies? And how can they be enabled to comply with the new data management requirements? One important aspect is the availability of an infrastructure for the preservation of and long-term access to data. We need a global network of financially





sustainable and certified trustworthy data repositories.

12:10 - 12:30 [Erik Fledderus \(SURF\)](#), "The Go-FAIR initiative"

The aim of GO FAIR is to implement a coherent approach to treating data. It is a collection of good practices on content, process and implementation. GO FAIR influences curricula (training) for new students & Life Long Learning. It works with partners to create FAIR-framework for existing and new data.

GO FAIR is intended to be a self-coordinating, board-governed organisation drawn from the stakeholder community, taking decisions driven by community consensus and consideration of different interests.

Main messages:

- Open Science IS NOT (the same as) Open Access
- Open data is about more than disclosure FAIR
- FAIR data IS NOT (the same as) Open / Free data

GO FAIR has developed several recommendations regarding governance, implementation, rules of engagement and policies.

GER/NL will start creating a support office – open for other contributors – focusing on:

- Assisting implementation networks/hubs to align themselves with the GO FAIR Rules of Engagement in order to prepare them to join the GO FAIR initiative.
- Supporting the implementation and further development of GO FAIR, e.g. by organising GO FAIR events such as workshops and meetings of the governing board.
- Fostering compliance with the self-prescribed Rules of Engagement by supporting the establishment and implementation of a monitoring mechanism for the GO FAIR implementation networks.

12:30 - 13:00 Panel

