

"European Open Science Cloud implementation - scenarios for EOSC architecture and impact on governance aspects",

Tuesday 9th May between 14h30 to 16h00 within the EGI conference in Catania.

PANEL OVERVIEW:

This is a session organised by the European Commission's EOSC team in unit RTD A6 to discuss state of play of the EOSC pilot project on architecture and governance, as well as EC policy development concerning possible technical and organisational architecture of the EOSC. This panel is intended to be very interactive with audience while counting on your expertise to drive meaningful discussions on key aspect of the EOSC.

BACKGROUND CONCEPTS & POINTS:

3 "O"s (Moedas) Open Science, Open Data, Open to the world

Key Challenges:

- Lack of awareness of the value of data and incentives for data sharing
- Lack of common standards to ensure interoperability of data
- Not enough hardware capacity for scientific computing, storage, connectivity,
- Fragmentation and lack of coordination over different scientific community and countries. Not just member state specific but national.
- Need to translate recent changes in data protection and copyright rules.

*DATA FLOWS ACROSS EU BORDERS currently the data localisation measures prevent data from being transferred easily between EU countries. Fundamental changes are required in order to let **data move across EU borders and not hampering an industry valued at around €272 billion in 2015.**¹*

Some highlights on RDA

RDA Europe project has great added value to achieve the goals of Open Science and Open Data in Europe.
RDA is THE place for data practitioners to discuss and implement data sharing and reuse.
RDA helped to advance the conversation between policy makers and high-level stakeholders on how to implement data sharing policies and practices.
The RDA recommendations have global impact on research and on the Grand Challenges.
RDA Europe has a positive impact to bring research closer to European citizens.

1. What main functionality should the EOSC provide?

VISION: to create a trusted open environment for storing, sharing & re-using scientific data & results and supporting Open Science practices

Therefore, the EOSC needs to give seamless access to Europe's wealth of data and the products, services and solutions to facilitate research & innovation and contribute to the data economy. **INTEROPERABILITY & TRUST** are two key elements. There are many, many requirements coming from different communities. **The services**

¹ http://www.euractiv.com/section/digital/news/fifteen-countries-ask-for-eu-legislation-on-data-flows/?nl_ref=36953490

deployed should be interoperable, orchestrating the platform and allowing users to pick and choose resources they need from the one stop shop.

Here are a couple of reusable results that could be part of the EOSCpilot from now:

- 1) the result of PICSE (H2020 CSA project) was the HNSciCloud Pre-Commercial Procurement that seems to be a suitable model for public sector organisations to jointly procure services. All the recommendations and lessons learnt coming out from PICSE (see PICSE roadmap) were taken into consideration in the writing of the HNSciCloud tender. Currently, many EU projects are asking for such tender material and they are reusing it in their PCP. An example is the **SELECT for cities** PCP – a project in the Internet of Everything area;*
- 2) EUDAT pilot users – there are 20+, rather than signing the SLA contracts provided by service providers or data centres, they could be drawn to review the common reference model (CRM) delivered under SLA-READY.eu – (all info is there) to make sure all the correct legalise & 26 terms of use are included & secondly the service providers & data centres themselves could draw upon this & match their content to that provided by the SLALOM project that pulled out actual content & cross-check that the info is relevant & updated.*

2. EOSC Ecosystem: where are we now in terms of its implementation and how will it evolve?

The EOSCpilot is currently contributing in a very practical way with 30 applications received for science demonstrations for 5 slots. The vision is to have a further 5 that should be funded at the end of the year 2017 beginning 2018 with the aim of reaching 15 in total for the whole project.

Moreover, EOSCpilot has a task on industry engagement – led by Corallia (Athena RC) & **RDA is currently working with ITU on IoT & data related requirements.**

5 science demonstrators from project start
5 to be selected from April 2017 call (30 applications being evaluated now) - due to start in 2nd half 2017
5 more to be funded through a 2nd open call due to be published in Sept 2017
Total EOSCpilot Science demonstrators.

5 science demonstrators that are an integral part of the project and started 1st Jan 2017

- Environmental & Earth Sciences - ENVRI Radiative Forcing Integration to enable comparable data access across multiple research communities by working on data integration and harmonised access
- High Energy Physics - WLCG: large-scale, long-term data preservation and re-use of physics data through the deployment of HEP data in the EOSC open to other research communities
- Social Sciences – TEXTCROWD: Collaborative semantic enrichment of text-based datasets by developing new software to enable a semantic enrichment of text sources and make it available on the EOSC.
- Life Sciences - Pan-Cancer Analyses & Cloud Computing within the EOSC to accelerate genomic analysis on the EOSC and reuse solutions in other areas (e.g. for cardiovascular & neuro-degenerative diseases)
- Physics - The photon-neutron community to improve the community's computing facilities by creating a virtual platform for all users (e.g., for users with no storage facilities at their home institutes).

3. How Research Infrastructures dedicated to serve the needs of a particular domain participate in this ecosystem? What is their ROI in making available their services?

This is a very complex question & clearly requires further development to which the EOSCpilot is contributing to, some of the standing point sin our vision are in the following, we appreciate that they are general considerations where we would need to build upon.

General considerations only:

- Research Infrastructures provide access to data but also to services and all working on community building. They have already invested many, many years to interacting with their stakeholders. They understand the domain specific as well as cross & multidisciplinary challenges as many of them are already integrating services in different communities;
- creation of a trusted environment for hosting and processing research data to support European science in its global leading role will help overcome many key challenges currently facing scientific disciplines;
- end users must be taken into account and their expectations should be carefully managed and quality of services deployed by EOSC is very important;
- while the complexity would need to be worked out the domain specific services & access to data should continue to be managed and deployed by the RIs themselves who understand the needs of their end-users clearly;
- KEY ROI ELEMENTS: access to scientific data should be easier, cheaper and more efficient.
- EOSC should reduce the duplication of efforts and allow RIs to invest their financial and human resources in the development or customisation of domain specific solutions, giving them access to and use of horizontal or heterogenous services and resources.

4. EOSC is expected to support open, collaborative, multidisciplinary and innovative scientific work. This means that scientists of different domains, working in different Member States, that decide to collaborate for a specific period of time to address an emerging and challenging scientific question will be supported and facilitated in performing their activities . Who should pay for the services and resources that this dynamically aggregated teams will consume?

Models of sustainability can be distributed onto the member states, enabling a true pan-European service: the 6th May Economist article: *“The Data Economy– the world’s most valuable resource”*² interestingly states that governments could encourage the emergence of new services by opening up more of their own data vaults or **managing crucial parts of the data economy as public infrastructure**, In the Future RDA 4 proposal a consideration was provided in its business plan to receive contributions from member states, that would be acquired by the national/regional nodes for services rendered by the RDA Europe entity. Quick figures: considering 27 member states at 30k€ per year we are looking at over 750k€ per year that could be replicated in the EOSCpilot model;

Hybrid payment models for communities / individuals could / should be foreseen – basic and premium levels or part paid by institutes / organisations (**like journal subscription model**) and a part that comes from the public funders and commercial users . . . again the one size fits all payment model will not work. Furthermore, **donation schemes** seem to be promising for these types of environments.

5. How to set basic level of quality of FAIR data and of services within the EOSC? How to implement certification schemes for data repositories?

Research data repositories across the world and scientific domains can only benefit from a single set of internationally recognized requirements to demonstrate their status as a Trustworthy Data Repository.

Certification schemes are clearly difficult to establish. For instance, with our experience in RDA we would like to suggest that light certification schemes of membership/compliance will be developed with no

² <http://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>

technical validation involved for organisations, this could be a potential source of revenues that can fuel the movement but concertated with RDA global. We believe that this schema is a good starting point of discussion.

(STANDARDS: RDA and WDS (world data system) (International Council of Science Unions (ICSU) in collaboration with Data Seal of Approval (DSA) has created harmonized Common Procedures for certification of repositories at the basic level, drawing from the procedures already put in place by the Data Seal of Approval (DSA) and the ICSU World Data System (ICSU-WDS) – in the process of ICT technical specifications (2nd cycle) – RDA is recognised as an organisation that can issue ICT tech specifications.)

6. Federating existing resources and facilities: for which functions you see centralisation needed?

Users First

The involvement of research infrastructures and the fact that they should continue to develop & deploy services for their users directly. **Science is and will continue to be domain specific.** Scientists will continue to conduct their research in their own specific field and those designing and deploying the services for scientists need to be centrally involved. On the basis of this consideration centralisations should probably be limited to the common functions where personalised & localised services should be left a national / community level.