The Open Science Cloud Challenges

EGI Input to the Stakeholder Community Workshop[1], 30-10-2015

27 Nov 2015

1 About the Open Science Cloud

Nowadays, research practice is increasingly and in many cases exclusively data driven. Knowledge of how to use tools to manipulate research data, and the availability of e-infrastructures to support them, are foundational. Along with this, new types of communities are forming around interests in digital tools, computing facilities and data repositories. By making infrastructure services, community engagement and training inseparable, existing communities can be empowered by new ways of doing research, and new communities can be created around tools and data.

In the EGI vision, **the Open Science Cloud offers researchers from all disciplines seamless, open access to the advanced digital capabilities, resources and expertise they need to collaborate and to carry out data- and computing-driven science. Secure and trustworthy, the Open Science Cloud engages researchers in governing, managing and preserving resources for everyone's benefit. The Open Science Cloud is an open, service-driven endeavour, inclusive of all stakeholders**[2].

The Open Science Cloud is not just a data federation. Cloud computing creates an element of data virtualization, takes computing to the data, and may help to solve some of the problems facing big data. We propose to exploit these opportunities by creating a 'commons' as one possible sustainable model. The Open Science Cloud will be collections of public and private resources (including cloud resources) for storing data and computing with those data. Research objects in the commons — for example, data, software, pipelines, experiments or papers — must be uniquely identified, discoverable, sharable (taking into account privacy issues), and resolvable to its source by using a unique identifier. The commons will be offered by different service providers, including e-Infrastructures for the delivery of baseline technical solutions, Research Infrastructures and research communities for the delivery of thematic and discipline-specific solutions.

[1] This paper reflects the views of the EGI Council Participants and of the EGI-Engage project consortium, including ESFRI research infrastructures and international communities: BBMRI-ERIC, DARIAH, EISCAT, EPOS, INSTRUCT, WLCG (high energy physics) and the MoBRAIN Virtual Research Community (structural biology and neuroscience).

[2] European Open Science Cloud, Nov 2015, EGI, EUDAT, GEANT, LIBER, OpenAIRE Joint Position Paper [http://dx.doi.org/10.5281/zenodo.32915]

As defined in the joint position paper, EGI believes the European Open Science Cloud for Research must be:

1. Open in design, participation and use.
2. Publicly funded & governed as a Commons.
3. Research-centric with an agile co-design with researchers and research communities.
4. Comprehensive in terms of universality and inclusiveness of all discipline
5. Diverse & distributed empowering networks effects.
6. Interoperable with common standards for resources and services.
7. Service-oriented as well as protocol-centric.
8. A socio-technical endeavour that connects diverse communities and promotes the development of human networks.

The Open Science Cloud services should offer different access modes depending on the type of service: not only free at point of use, but also policy-based (users are granted access based on policies defined by the providers to meet some national or EU objectives) and market-driven (access requires the payment of a fee).

1.1 EGI Mission and Vision

EGI is an international collaboration that federates the digital capabilities, resources and expertise of national and international research communities in Europe and worldwide. Our main goal is to empower researchers from all disciplines to collaborate and to carry out data- and compute-intensive science and innovation.

The organisational and governance structure of EGI builds on EGI.eu, a not for-profit foundation established under Dutch law in the Netherlands. The foundation has participants and associated participants drawn from NGIs, EIROs, ERICs, and other legal entities. These entities participate in the foundation independently or as the representative of a national e-infrastructure consortium.

Our vision[3] is that r*esearchers from all disciplines have easy, integrated and open access to the advanced digital capabilities, resources and expertise needed to collaborate and to carry out data/compute intensive science and innovation*. The mission of EGI is to create and deliver open solutions for science and research infrastructures by federating digital capabilities, resources and expertise between communities and across national boundaries.

1.2 EGI contributions to the Open Science Cloud

- The capability to federate digital capabilities and resources services worldwide. Currently the federation comprises 620,000 CPU cores, 350 providers, 500 PB of storage, data, > 500 virtual appliances and open source software. Participation is open to providers – both publicly funded and private – who meet a set of community-defined requirements.
- Abstraction of service capabilities through the development, implementation, adoption and promotion of open standards and adoption of different interoperable technologies.

[3] http://www.egi.eu/about/

- EGI Federated Cloud: the first open standard-based research cloud federation allowing the sharing of software appliances through a library of tools, their portability across the federation; brokering in the federation allows bringing computing to the data;
- Sharing of services across research communities for increased efficiency of use and greater return of investment, for the development of an international multi-disciplinary resource commons that is available for access to any researcher worldwide through different access policies, including pay for use.
- A service-oriented approach in which research communities are not only consumers but also providers of innovation and services for community-specific capabilities in one integrated catalogue.
- Co-design through Competence Centres involving Research Infrastructures and research communities that contribute to the development of human networks of experts, services providers and technology providers.

1.3 Grand Challenges

Challenge 1. Long-term availability of national e-Infrastructures to support multidisciplinary science ubiquitously in Europe, to ensure that sharing of digital services is possible and enabled both at a national and international level.

Recommendations for the Open Science Cloud implementation

- Incentivize national investment plans of a *national* Open Science Cloud accessible to all researchers and disciplines, which is federated in a global Open Science Cloud integrating digital services beyond Europe, to provide the necessary capacity and capabilities of data-driven science. Involve governments in the definition of the sustainability plan of the Open Science Cloud and its cost models, in order to converge towards a sustainable business plan in terms of technical solutions, legal, financial and governance models.
- Enforce the availability across Europe of generic national e-Infrastructures services able of meeting the needs of different user communities so that duplication of services and silos are removed.
- Foster public-private partnerships to develop capacity and capabilities with industry and SMEs, establish a market generating business opportunities; remove national legal barriers that prevent such a market from existing.
- Develop new skills in data science and develop human networks.

Challenge 2. The procurement of a European storage and computing backbone that jointly provides data preservation and computing capabilities, offers research data as a service to promote sharing, allows for scalable access to research data, and provides an amount of capacity sufficient to face the computing and storage needs of the research data deluge.

Recommendations for the Open Science Cloud implementation

- Define a multi-stakeholder European governance harmonizing e-Infrastructures and Research Infrastructures controlling: policies, compliance to values and mission, performance, enforcing rules, demand-supply relationships, the definition of strategic/tactical/operational layers etc.
- Establish a network of physically distributed research data hubs that federates heterogeneous archiving systems, and datasets from multiple providers with data peering agreements and data geo-replication across different world regions and involving multiple data providers. Develop a business model that promotes sharing.
- Establish European cross-border procurement of an Open Science Cloud that serves all disciplines and research communities and offers on-demand capacity to resource-bound users; coordinate funding of digital services at a national and European level to avoid duplication and silos.
- Funders should encourage the development of new metrics to ascertain the usage and value of data, and persuade data resources to provide such statistics for all of the data they maintain.
- Develop new services for Open Science, for example to support repeatability of science, for IPR management. By increasingly sharing models and modelling tools researchers and research communities can capture the steps of the digital research processes they carry out for excellent science. With suitable abstractions and robust provenance capabilities such models and tools would enable the repeatability, and therefore the incremental improvement of research practices and processes within and across research teams.
- Ensure the Open Science Cloud provides fair access to sensitive data in compliance with different research data policies and EU regulations: the data custodian of hosted sensitive datasets will always retain the control of access to ensure compliance to the informed consent for the given data.

Challenge 3. The realization of a integrated service portfolio with both thematic and baseline cross-cutting "solutions" – jointly provided by e-Infrastructures and research communities – to address the whole research lifecycle, open to any provider and research discipline, through increased technical interoperability, standards adoption, harmonized access policies, federated service management processes involving multiple service providers.

Recommendations for the Open Science Cloud implementation

- Remove competition for funding across different communities and different infrastructures by providing grants that reward services according to usage; reorganize national and European funding so that the creation of partnerships is rewarded (both e-Infrastructures and Research Infrastructures are service providers); incentivize the use of existing services.
- Establish credits to access services, in which a grant recipient is given credits instead of funding to pay for computational time and other digital services. A principal investigator would be able to spend those credits at any commons-compliant resource. Researchers whose work involves extensive computation on small amounts of data may spend their credits at a different commons-compliant resource to investigators who do minimal computing on large amounts of data. Such credit system would help to promote sharing of research data and methodologies.

- Involve research communities and infrastructures that are service providers in the Open Science Cloud in the technical support and develop human networks, software development and data science skills and the knowledge commons.

This work is supported by the EGI-Engage project, co-funded by the European Union (EU) Horizon 2020 program under Grant number 654142 http://go.egi.eu/eng