**CNRS-EOSC** (internal note)

**Aim of this document**

The EOSC, which is going to provide researchers in Europe access to e-infrastructures, services, and FAIR data, is becoming a reality. The CNRS needs to urgently develop a vision and decide on a strategy to address the transformation that modern society in general, and scientific and engineering research in particular, are undergoing. The CNRS needs to engage in this process and to propose a science-driven strategy to the EOSC initiative and to the national open science plan and roadmap for data and computing infrastructures.

The aim of this document is to motivate and to initiate this process.

**EOSC from the point of view of CNRS**

The European Open Science Cloud (EOSC) is a vision of the European Commission of a large e-infrastructure to support and develop FAIR data, open science and open innovation in Europe and beyond. The EOSC aims to enable FAIR data and their seamless use by European scientists and users through a shared distributed platform of services and tools spanning a continuum of data and computing resources in edge and centralized (e.g., Cloud and HPC) environments.

The goal is not to create genuinely new services and e-infrastructures, but to build upon the existing landscape and create a shared distributed software platform of services that can address the logistical challenges of data use by application-workflows along the entire trans-national continuum, a critically important challenge for the scientific communities and the society in the coming decade. This should be achieved by providing a model of distributed services and resources, a thin scalable spanning layer and a sustainable governance and business model allowing the access to the EOSC free-of-charge at the end users point. In this sense, the EOSC can be understood as a system-of-systems. The EOSC is an important component of the [European Cloud Initiative](https://ec.europa.eu/digital-single-market/en/%20european-cloud-initiative) and an essential element of Europe’s Open Science and FAIR data strategy.

Although the official launch of the EOSC took place in November 2018, the system is in a critical phase. While the EOSC begins to provide European wide services, the system in itself is and must be an evolving community-driven one. In the next two years important decisions about the future of the EOSC will be taken, such as the definition of core functions, rules of participation, the governance and business model, and other vital aspects.

**Why should CNRS as a research organization play a role in defining the EOSC?**

The open science initiative of the French government, and the national roadmap of computing and data infrastructures, mandates an EOSC-type approach.

The CNRS as the largest research organization in Europe touches almost all areas of scientific research today. The vision of the EOSC is similar to what should be that of the CNRS: providing a shared software platform of services that enables the CNRS researchers to seamless access to data and a continuum of edge and centralized data and computing infrastructures, across the scientific domains and research practices of the CNRS institutes. CNRS and EOSC face similar challenges, although at different levels: how to set up a community-driven shaping strategy; how to provide sufficient human support for the activities in a shared environment; how to set up a persistent business and governance model for shared services and resources (software platform of distributed services, data and computing infrastructures, and human capacity building).

CNRS is already a service provider and one of the main users of EOSC services, e.g. through [EOSC-hub](https://www.eosc-hub.eu/), [XDC](http://www.extreme-datacloud.eu/), and other EOSC-related projects. Funding schemes for the different services are diverse: some are funded by CNRS as being part of a IR or TGIR, others are part of the activities of a single institute and are funded by CNRS through the institute budget, some are financed through European or ANR grants. Also governance structures for the various services and infrastructures participating in the EOSC vary over a wide range.

[France Grilles](http://www.france-grilles.fr) is an IR and a *Groupement d’Intéret Scientifique* (GIS), led by CNRS-IN2P3 and involving a number of the CNRS institutes and other national partners (e.g., CEA, INRA, INRIA), is the main French contribution and driver behind the development of a European Grid and Cloud distributed infrastructure, and the provision of [EGI](https://www.egi.eu/) and [EUDAT](https://eudat.eu/) services.

Similarly CNRS with GENCI is a major stakeholder and provider of HPC services in [PRACE](http://www.prace-ri.eu/) and EUDAT for the benefit of French and European scientific communities.

Concerning EOSC-hub, one prominent example of added value of the EOSC is demonstrated by the [DIRAC](http://diracgrid.org/) project enabling workflows instantiation that was originally developed by the research informatics and the particle physics community. With the help of EOSC, the service is now being expanded to give easy Grid and Cloud access to other communities through web portals.

The CNRS through its different institutes is also strongly committed and contributing in several ESFRIs from the European roadmap, which are intense data producers and data infrastructure providers. These ESFRIs are already largely involved in FAIR data services complying with the EOSC, e.g., ICOS ERIC, EPOS, ACTRIS, eLTER, KM3NeT, CTA, HL-LHC, SPIRAL-2, DARIAH, CLARIN and many more.

Other examples of EOSC in action are the [B2-services](https://eudat.eu/b2-service-suite) of EUDAT, made available through the [EOSC porta](https://www.eosc-portal.eu)l[[1]](#footnote-1)) . These allow storing, accessing, distributing and managing data. The portal gives some more examples of [specific CNRS-related science use cases of the EOSC](https://www.eosc-portal.eu/eosc-in-practice/use-cases).

There is a vital interest for the CNRS in becoming one of the main stakeholders of EOSC and in promoting a community-driven shaping strategy behind EOSC. The CNRS urgently needs to develop a vision and a strategy to guarantee a co-definition, co-development and co-implementation of the future EOSC in synergy with research needs and practices.  This is critical for the 2019-2020 phase that must develop policies, rules of participation, governance and business models of the EOSC, but also on the long term, as the EOSC must be a science-driven system, with an open [service catalogue](https://eosc-hub.eu/catalogue), evolving with the research needs and practices.

In this sense, the CNRS and the EOSC share common goals: providing a distributed software platform of services to the science communities enabling seamless access to a continuum of edge and centralized data and computing resources; developing a community-driven shaping strategy to ensure that these services and continuum of resources evolve (in quality and quantity) with research needs and practices; developing persistent solutions with reliable and acceptable governance and business models supporting European open science. This has a cost and must include human resources and expertise.

It is clear that a large number of CNRS communities will participate in the EOSC through the international scientific collaborations they are part of. The risk is here that this might result in participation of the research community without an active strategy and vision of the CNRS.  It can also be anticipated that national and international funding agencies will increasingly request open access to the data generated through their funded projects and supporting publications, which requires an appropriate e-infrastructure for the scientific community. Finally, through an efficient software platform of distributed services rendering research data available - such as may be expected for the EOSC - CNRS will leverage the full potential of the research outcomes produced by its researchers.

The CNRS needs today to operate in a fast changing landscape and to respond to the disruptive wave of change, over the past four years, wrought by BigData and computing technology that transform research practices and raise new scientific challenges. The EOSC is a timely opportunity to reflect at the CNRS level on the right strategies and policies.

**The CNRS needs to urgently develop a vision and build a community-driven shaping strategy to address the watershed transformation that modern society in general, and scientific and engineering research in particular, are undergoing. The CNRS needs to engage in this process and to propose a science-driven strategy to the EOSC initiative and to the national open science plan and roadmap for data and computing infrastructures**.

This is becoming critically needed to support new research practices in and across the CNRS institutes, to keep the CNRS internationally competitive and attractive to young researchers, and for the CNRS to contribute to science-informed policy making.

**In what ways can CNRS provide important advice for the development of the EOSC?**

The CNRS as a cross-disciplinary research organization can be seen as a research-driven use case and test bed for the EOSC. The CNRS faces similar challenges with different levels of: FAIR data maturity and services; interoperability capabilities across its institutes. This reflects a large variety of domain specific ‘cultures’ and research practices in terms of: data management, curation and sharing; data logistics of scientific applications/workflows; as well as different organization and funding models for data stewardship and data-driven research activities. Thus, the need of a software platform of distributed services spanning the continuum of edge and centralized data and computing resources, of FAIR data services enabling science-driven data interoperability, and the questions of reliable and acceptable governance and business models, are similarly important in the CNRS and EOSC ecosystems.

CNRS must therefore develop a community-driven shaping strategy providing important input from its research communities on research-driven requirements, needs, and challenges in deploying and using the EOSC services to ensure that EOSC evolves in synergy with the research practices and challenges of the CNRS communities. In return EOSC provides a unique European opportunity to address these disruptive issues and to reinforce the international visibility and attractiveness of the CNRS.

CNRS is also strongly involved in shaping the French data infrastructure required for large instruments and observation systems, e.g. in high-energy physics (LHC), environmental sciences (System-Earth), astronomy and astrophysics (CDS Strasbourg), in satellite observations together with the CNES, in life science with the French Institute of Bioinformatics (IFB - CNRS INSB, joint unit with INRA, Inserm, CEA, INRIA) and in social sciences with Huma-Num and Progedo, but also with services for scientific publications like OpenEdition, CCSD, Persée and METOPES. It has therefore every science-based and experience-based asset to play a leading role in a community-driven shaping strategy and implementation of the EOSC.

**What is the current engagement of CNRS in EOSC?**

CNRS participation and/or membership in the EOSC:

* [EOSCpilot](https://www.eoscpilot.eu/); a project (2017-2019) to explore how EOSC can be put into action
* [EOSC-hub](https://www.eosc-hub.eu/); making a significant number of services EOSC-compliant and accessible through the EOSC portal (2018-2020); participation on CNRS side e.g. through [DIRAC4EGI](https://www.eosc-hub.eu/keywords/dirac4egi), [OpsPortal](https://www.eosc-hub.eu/catalogue2/Operations%20Portal), [ARGO](https://www.eosc-hub.eu/catalogue2/Argo%20Service%20monitoring), [EISCAT-3D](https://www.eosc-hub.eu/research-communities/eiscat3d-agile-data), [Marine research](https://www.eosc-hub.eu/research-communities/marine-research-competence-centre), [LOFAR](https://www.eosc-hub.eu/research-communities/radio-astronomy-competence-center), [ELIXIR](https://www.eosc-hub.eu/research-communities/elixir), [EPOS](https://www.eosc-hub.eu/research-communities/epos-orfeus-competence-center), and [ICOS](https://www.eosc-hub.eu/research-communities/icos).
* [PaNOSC](https://panosc-eu.github.io/): The Photon and Neutron Open Science Cloud (PaNOSC) is a European project (financed by the INFRAEOSC-04 call, 2018-2022) for making FAIR data a reality in 6 ESFRI Research Infrastructures (RIs), developing and providing services for scientific data and connecting these to the EOSC. Partners are the [European Synchrotron Facility](https://www.esrf.eu/) (ESRF; coordinator), [ILL](https://www.ill.eu/fr/), [XFEL.EU](https://www.xfel.eu/), [ESS](https://europeanspallationsource.se/), [CERIC-ERIC](https://www.ceric-eric.eu/), [ELI-DC](https://eli-laser.eu/), and [EGI](https://www.egi.eu/).
* ExPaNDS: This [INFRAEOSC-05](http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/infraeosc-05-2018-2019.html) project is designed to be the national counterpart of PaNOSC. It is coordinated by DESY and PSI, and SOLEIL is responsible for the work package on training and is co-leader for the one data analysis.
* EOSC-Pillar: another successful INFRAEOSC-05 proposal to advance national services to be included in the EOSC and to develop an EOSC governance model and to implement it (2019-2022).
* [OpenAIRE](https://www.openaire.eu/): Shift scholarly communication towards openness and transparency and facilitate innovative ways to communicate and monitor research.
* [XDC](http://www.extreme-datacloud.eu/): developing scalable technologies for federating storage resources and managing data in highly distributed computing environments. Solutions for Life Science, Astrophysics, High-Energy Physics, Photon Science and Clinical Research.
* [ESCAPE](https://indico.in2p3.fr/event/18279/): address the Open Science challenges shared by ESFRI facilities and pan-European research infrastructures in astronomy and particle physics (2019-2021).
* [EGI](https://www.egi.eu): federated e-Infrastructure set up to provide advanced computing services
* [EUDAT](https://eudat.eu): e-infrastructure of integrated data services and resources
* [RDA](https://www.rd-alliance.org): the Research Data Alliance builds the social and technical bridges to enable the open sharing and re-use of data
* [EOSC-LIFE](https://www.bsc.es/research-and-development/projects/eosc-life-providing-open-collaborative-space-digital-biology): through the participation of the French Institute of Bioinformatics (INSB)
* [Co-OPERAS](https://www.go-fair.org/implementation-networks/overview/co-operas/): Implementation Network (GO FAIR initiative) to develop standards and FAIRification of SSH data — lead by TGIR Huma-Num.
* [SSHOC](https://www.sshopencloud.eu/news/sshoc-kick-utrecht-march-11-12-2019%20): address the Open Science challenges shared by ESFRI facilities and pan-European research infrastructures in Social Sciences and Humanities (2019-2021). TGIR Huma-Num and UMR MAP 3495 are involved here.

**What next steps should be taken by CNRS vis-a-vis the EOSC?**

1. Sign the [EOSC declaration](https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf) (this is not a formal commitment to provide resources to the EOSC, but a declaration to support the EOSC vision in general).
2. Ensure CNRS representation in crucial activities, such as EOSC Working Groups, Summits, and the Stakeholder Forum to reflect the research practices of its different institutes. In addition, explore what steps can be taken to increase the influence of CNRS on the future EOSC program?
3. Develop an open science, FAIR data policy and services within the CNRS in synergy with the national and European open science plans, RDA together with other international organisations of relevance for the CNRS research communities.
4. Rapidly discuss the vision of CNRS at large and build a community-driven shaping strategy for the development of a software platform of distributed services that enable seamless access to its edge and centralized data and computing infrastructures, and their evolution in general and in particular within the EOSC.
5. Identify potential issues for CNRS in its role as provider of services and resources and develop a clear position and a constructive science-based strategy (vis-à-vis the ministries and other partners).
6. Analyse issues of integrating existing distributed services with the future EOSC (e.g. interoperability between HAL and related data repositories for example, edge and centralised data and computing infrastructures, data stewardship).
7. Develop and apply a strategy to inform and engage the potential users of the EOSC within CNRS.
8. Coordinate EOSC proposals and their collaboration between the different institutes well in advance.
9. Increase the visibility of CNRS in EOSC. For example, as of today, CNRS is not listed among the EOSC members on the main [EOSC portal](https://www.eosc-portal.eu/).
10. Contribute to the definition of a viable and acceptable business model for the EOSC since e-infrastructures and the provision and support of services are, in most cases, not funded through the EOSC. Support at national level (e.g. from the CNRS itself, from MESRI or other ministries) should be discussed.  **Without additional resources for EOSC-related activities, the CNRS participation in the EOSC will be marginal.**

GT CNRS-EOSC:

V. Beckmann (IN2P3), F. André (DIST), M. Baaaden (INC), C. Berthon (INSMI), C. Blanchet (INSB), D. Boujard (INSB), Ph. Briand (INSMI), M. Daydé (INS2I), S. Dumouchel (INSHS), B. Gagey (Soleil), D. Girou (IDRIS), F. Godeferd (INSIS), L. El Khouri (DIST), S. Lamare (INEE), L. Lellouch (INP), P.-E. Macchi (CC-IN2P3), L. Maurel (INSHS), B. Pier (INSIS), S. Rousset (DIST), D. Veynante (MICADO), J.-P. Vilotte (INSU)

1. https://www.eosc-portal.eu [↑](#footnote-ref-1)