

# **EOSC and Research Infrastructure: Legal Challenges of HPC**

Ludovica Paseri

University of Bologna

**Abstract.** The European Open Science projects are strictly related to the wider European Cloud Initiative. This Initiative also takes into account the development of the research infrastructures, and, in particular, the High Performance Computing for scientific research purposes is highly encouraged. The implementation of HPC platforms for the benefit of universities and research centres entails several advantages but also poses some legal challenges. The paper aims to describe the link between Open Science, the European Open Science Cloud (EOSC) and HPC, focusing, finally, on the legal challenges of the latter

**Keywords.** Open Science; European Polices; High Performance Computing; Scientific Research; European Cloud Initiative; EOSC

## **Introduction: The European Open Science Projects**

The European institutions, and in particular the European Commission, are engaged in supporting scientific research by fostering Open Science projects. The Open Science paradigm advocates the openness of every phase of the research process, from data collection to publication of results (Fecher, Friesike, 2014), following the principle «as open as possible, as closed as necessary». The openness of research data promoted by the Open Science, in fact, does not aim at a blind and indiscriminate sharing of data. On the contrary, such openness is always the result of balancing the different interests involved (e.g., the protection of personal data; national security; or public interest).

On 23 November 2018, the European Open Science Cloud (EOSC hereinafter) was launched in Vienna: a trusted and federated environment that aims to coordinate the Open Science projects, implemented in recent years, to facilitate the sharing of scientific research both in terms of research data and in terms of research results, namely publications, for the benefit of 1.7 million European researchers and 70 million science and technology professionals, by breaking down geographical barriers, and boundaries traditionally imposed by different disciplines (Ayris, et al., 2016; Burgelman, 2021).

In the Communication of the European Commission entitled “A European Strategy for Data” (European Commission, 2020), the intention to support the Open Science is stressed and, in particular, the importance of EOSC is underlined. EOSC has been represented as one of the ten domain-specific data spaces for which the European institutions intend to develop specific actions and policies. An effective implementation of EOSC would be an example of digital sustainability, able to produce a concrete impact on the world of scien-

tific research and, more generally, on society.

In this context, characterised by the emergence of the Open Science paradigm, it is relevant to focus on the High Performance Computing (HPC, hereinafter), as a means by which foster scientific research (Section 1). Clarified this scenario, and the interplay between the Open Science projects and the HPC, attention will be drawn to the legal challenges which have arisen (Section 2). Finally, Section 3 will provide the conclusions of the analysis.

## 1. High Performance Computing for Scientific Research

EOSC should be considered as a part of a wider and more ambitious European project: the European Cloud Initiative (European Commission, 2016). This initiative, which aims to build a competitive data and knowledge economy in Europe, is based on two pillars: (i) EOSC; and (ii) the European Data Infrastructure (EDI, hereinafter). While EOSC is, as mentioned above, a trusted and federated ecosystem for sharing science in Europe, EDI is, instead, an infrastructure of highest quality, which will underpin EOSC, based on supercomputing capacity, fast connectivity and high-capacity cloud solutions (European Commission, 2016). An extremely relevant aspect of EDI architecture is represented by the High Performance Computing strategy (European Commission, 2012). HPC is a form of supercomputing that entails a complex processing system, involving multiple processors capable of providing high performance, previously unseen, such as real-time analysis of huge amounts of data, the so-called Big Data. The HPC strategy is the means by which the value of Big Data in Europe can be fully exploited, and by linking up with EOSC project, the main beneficiaries may become researchers, universities and research centres. The EDI structure takes into account the need of IT infrastructures by European researchers in order to take full advantage of the digital services available today.

The link between the development of Open Science projects (EOSC in particular), and the implementation of HPC platforms, as infrastructures to foster scientific research in Europe, is emphasised by the recent European Strategy for data: «The aim is to help common data and world class cloud infrastructures for the public good to emerge, enabling secure data storage and processing for the public sector and research institutions» (European Commission, 2020).

In order to foster HPC in Europe, the European High-Performance Computing Joint Undertaking (EuroHPC JU, hereinafter) has been created: it is a legal and funding entity which will empower pooling of EU and national resources in HPC. The EuroHPC JU project is explicitly related to the European Cloud Initiative «[...] in order to gather the necessary resources and capabilities [...] to close the chain from research and development to the delivery and operation of the exascale High Performance Computing system», as stated in the Council Regulation 2018/1488 of 28 September 2018. It is also claimed that «[...] a mechanism should be set up at Union level to combine and concentrate the provision of support to the establishment of a world-class European High Performance Computing infrastructure and for research and innovation in High Performance Computing by Member States, the Union and the private sector. This infrastructure should provide

access to the public sector users, users from industry, including small and medium- sized enterprises (SMEs), and users from academia, including the scientific communities of the emerging European Open Science Cloud» (Council Regulation, 2018).

## 2. HPC Legal Challenges

Providing exascale supercomputing capability to universities and research centres and boosting the European digital economy starting from research and innovation bring evident benefits. The novelty of these initiatives, however, also entails some previously unseen challenges. The major legal issues are summarised below in three macro categories: data protection; commercial trade secrets; ownership of data.

**Data protection.** The compliance with GDPR is complicated by the different layers of which an HPC platform is generally composed. The main issues are related to compliance with the principles of lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, confidentiality, and accountability (Paseri, et al., 2021). Another interesting aspect is related to Article 22 GDPR, which requires that «The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her» (see, Pagallo, 2020). Although scientific research enjoys the derogations provided for in Article 89 GDPR, the complex context requires a more in-depth analysis. In addition, data protection issues in the specific field of HPC have been relatively little investigated so far. Considering this aspect and the typical data protection issues in the field of scientific research (e.g. discipline fragmentation, on the Italian case, see: Durante, Senior, 2019), the situation becomes more complex and deserves further investigation.

**Commercial trade secrets.** The topic becomes relevant when the research project is co-sponsored by private actors, who have economic interests: the participation of different kind of actors can lead to the situation in which researchers do not release research results until a patent has been obtained (Carroll, 2015). In this situation, the initial phase of the research project becomes crucial: it is in the setting of agreements between universities and private entities that the future of the project is defined. The hurdle is represented by the fact that very often research bodies have neither legal competence, nor time, nor bargaining power with private bodies. This lack of legal support can represent a limit for the subsequent sharing of research results.

**Ownership of data.** The data ownership raises two different levels of complexity: on the one hand there is a tension between those conducting the research and those running the platform (e.g., Sheehan, 2018); and on the other hand, in research projects involving private companies there may be a problem of ownership of the data processed and resulting by the research projects (Guibault, Wiebe, 2013).

## 3. Conclusions

In light of the analysis of the ecosystem designed by the European Union that involves Open Science projects, in particular the EOOSC and the HPC infrastructure, some legal

issues can be identified. This contribution aims to provide an overview of these challenges, which are: the protection of personal data involved, the management of trade secrets and the ownership of research data, in relation to the multitude of actors involved with different interests.

These issues need to be further investigated, considering that there are already a multitude of HPC platforms operating for scientific research, scattered across the EU. Usually, researchers and operators of these HPC platforms are dealing with the same legal issues, leading to a kind of general uncertainty. What is crucial to understand, however, is that very often this legal uncertainty does not stem from the lack of ad hoc legal provisions for the specific technology: more frequently the uncertainty concerns the application of existing disciplines. This uncertainty becomes a barrier to sustainable scientific research capable of exploiting the advantages of the most innovative technologies. Nevertheless, the commitment to the Open Science principles can transform this limit into an opportunity (Besançon, 2021), pursuing not only greater openness of science, but also greater collaboration between researchers and different fields of knowledge.

## References

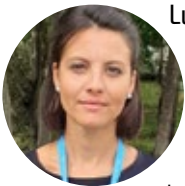
- Ayris P., et. al., (2016) Realising the European open science cloud, European Union.
- Besançon, L., et al. (2021) “Open science saves lives: lessons from the COVID-19 pandemic.” *BMC Medical Research Methodology* 21.1, 1-18.
- Burgelman, J.-C. (2021) “Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story).” *Data Intelligence* 3.1, 5-19.
- Carroll, M., W., (2015) Sharing research data and intellectual property law: A primer, in “*PLoS biology*” 13.8.
- Council Regulation (EU) 2018/1488 of 28 September 2018, “Establishing the European High Performance Computing Joint Undertaking”, ST/10594/2018/INIT, in OJ L 252, 8.10.2018, p. 1–34.
- Durante, M., Senor, M., (2019) “Report on the harmonization of Italian Law with the enforcement of The GDPR.”, 89-98.
- European Commission, (2012) “High-Performance Computing: Europe's place in a Global Race”, COM/2012/45 final.
- European Commission, (2016) “European Cloud Initiative - Building a competitive data and knowledge economy in Europe”, COM/2016/178 final.
- European Commission, (2020) “A European Strategy for Data”, COM/2020/66 final.
- Fecher B., S. Friesike, (2014) Open science: one term, five schools of thought, in: “Opening science”, Springer, Cham, 2014. 17-47.
- Guibault, L., Wiebe, A. (2013) “Safe to be open: Study on the protection of research data and recommendations for access and usage”, Universitätsverlag Göttingen.
- Pagallo, U., (2020) “Algoritmi e conoscibilità.” *Rivista di filosofia del diritto* 9.1, 93-106.
- Paseri, L., Varrette, S., Bouvry, P., (2021) “Protection of Personal Data in High Performan-

ce Computing Platform for Scientific Research Purposes.” Annual Privacy Forum. Springer, Cham, 123-142.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Sheehan, K. B., (2018) “Crowdsourcing research: data collection with Amazon’s Mechanical Turk.” Communication Monographs 85.1, 140-156.

## Author



Ludovica Paseri [ludovica.paseri2@unibo.it](mailto:ludovica.paseri2@unibo.it) - [ludovica.paseri@unito.it](mailto:ludovica.paseri@unito.it)

Ludovica Paseri is a PhD candidate at LAST-JD Joint International Doctoral degree program in “Law, Science and Technology”, at the Alma Mater Studiorum, University of Bologna, in cotutelle with the University of Luxembourg, faculty of Computer Science. She graduated in Law at the University of Turin and attended a postgraduate Master in Data Protection Law, at the Department of Law, University of Turin. Ludovica has completed the internship as trainee lawyer in an Administrative Law firm and her main research interests are Cloud Computing, EU Open Science policies, scientific data governance, privacy and data protection law.