

Delivering the European Open Science Cloud (EOSC): Principle and Practice in delivering Open Science

Author: Paul Ayris - Pro-Vice-Provost (UCL Library Services), Co-Chair of the LERU INFO Community (League of European Research Universities) & Adviser to the LIBER Board (Association of European Research Libraries)

Email: p.ayris@ucl.ac.uk

16.1. THE EUROPEAN OPEN SCIENCE CLOUD

In 2016, the European Commission's High Level Expert Group (HLEG) on the European Open Science Cloud published their Report.¹ I was privileged to be a member of this HLEG. The Report is designed to establish a vision for the future of Research Data, particularly Open Data, in Europe. The main findings were:²

- The majority of the challenges to reach a functional EOSC are social rather than technical.
- The major technical challenge is the complexity of the data and analytics procedures across disciplines rather than the size of the data per se.
- There is an alarming shortage of data experts both globally and in the European Union.
- This is partly based on an archaic reward and funding system for science and innovation, sustaining the article culture and preventing effective data publishing and re-use.
- The lack of core intermediary expertise has created a chasm between e-infrastructure providers and scientific domain specialists.
- Despite the success of the European Strategy Forum on Research Infrastructures (ESFRI), fragmentation across domains still produces repetitive and isolated solutions.
- The short and dispersed funding cycles of core research and e-infrastructures are not fit for the purpose of regulating and making effective use of global scientific data.
- Ever larger distributed data sets are increasingly immobile (e.g. for sheer size and privacy reasons) and centralised HPC alone is insufficient to support critically federated and distributed meta-analysis and learning.
- Notwithstanding the challenges, the components needed to create a first generation EOSC are largely there but they are lost in fragmentation and spread over 28 Member States and across different communities.
- There is no dedicated and mandated effort or instrument to coordinate EOSC-type activities across Member States.

The Report proposed 4 Recommendations for Policy development, 4 for Governance and 7 for the next phase of Implementation.³ Policy Recommendation 4 captures the essence of the HLEG vision for the EOSC: 'Frame the EOSC as the EU contribution to an Internet of FAIR Data and Services underpinned with open protocols.' The Recommendations on Implementation spell out the need for a Roadmap for future development of the EOSC, with rules of engagement and a light-touch framework for governance. How easy will it be to deliver this bold vision? Let us consider two challenges: Funding and FAIR research data.

¹ European Commission: <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; last accessed 8/1/17.

² Ibid., p. 6.

³ Ibid., p. 7.

16.2. CHALLENGES: FUNDING AND FAIR RESEARCH DATA

Funding is an obvious challenge. The Report emphasises that the need is not so much one to build new infrastructures, but rather to make what Europe already has less silo-based and more interoperable:

Based on the consensus that most foundational building blocks of the Internet of FAIR data and Services are operational somewhere, but that they operate in silos per domain, geographical region and funding scheme, we recommend that early and strong action is taken to federate these gems. Optimal engagement is required of the e-infrastructure communities, the ESFRI communities and other disciplinary groups and institutes. Several of these cross-ESFRI building blocks begin to operate in individual Member States. Simultaneously, the wealth of small and large industrial players in Europe should be engaged. All partners and stakeholders that adhere to standards and sign off on the Rules of Engagement (RoE) should be eligible.⁴

Even if this is true, the costs for delivering the EOSC are for significant sums of money. Overall, it is estimated that €2 billion is needed from the Commission's Horizon 2020 funding pot, as well as additional public and private investment of €4.7 billion to develop further the European data infrastructure. Of this €4.7 billion, €0.2 billion is needed for widening the user base to the public and private sectors, €1 billion for the EU-wide Quantum technologies flagship and €3.5 billion for data infrastructure.⁵

Stakeholder Group / Need	Funding requirement to deliver EOSC
European Commission	
Horizon 2020	€2 billion
Public / Private funds	
Widening User Base to public and private sectors	€0.2 billion
EU-wide Quantum technologies flagship	€1 billion
Data infrastructure	€3.5 billion
TOTAL	€6.7 billion

Figure 16.1: Perceived funding requirement to deliver the EOSC

The financial requirement is significant indeed.

What can be said about the Report's central insistence on FAIR data? The vision of the Report is for the EOSC to be technically conceived as an Internet of FAIR data and Services. It points to a parallel with the early development of the Internet:

The creation of NSFNET, choice of the TCP/IP standard and the authorised development of Domain Names enabled the boom of the Internet in the 1990's, where the development of the HTTP and HTML drove its major application domain, the largely textual WWW. This combination of authorisation, key support by a major leading agency (NSF) and a dedicated community (W3C) setting and enforcing minimal standards allowed virtually everyone to start building standard-compliant tools and services in the ecosystem.⁶

⁴ Ibid., p. 14.

⁵ European Commission: <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>; last accessed 8/1/17. For Quantum technologies, see European Commission: <https://ec.europa.eu/digital-single-market/en/news/european-commission-will-launch-eu1-billion-quantum-technologies-flagship>; last accessed 8/1/17.

⁶ European Commission: http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf; last accessed 8/1/17.

FAIR research data is **F**indable, **A**ccessible, **I**nteroperable and **R**eusable.⁷ A move to FAIR data as the default principle for research data management is a significant step. The 2016 UCL study of research data analysed the reasons why UCL researchers would not share their data.⁸

Question	Result/Response
Nobody asked me to	40%
Confidentiality / Intellectual Property / Data Protection	25%
Ethical Issues	10%
Other	7%
Time / effort required to collect them	6%
Possible misinterpretation of data	6%
Commercial Issues	3%
Licence agreement prohibits sharing	2%
Data no longer readable	1%

Figure 16.2: Question 55. Why didn't you give access to your data? (118 respondents)

Whilst FAIR research data is essential to underpin the EOSC as an Internet commons of data available for sharing and re-use, it is clear that the academic community has some way to go to see Open Data as the default position for research data they are creating/using, as the UCL survey shows.

16.3. CONCLUSION

The publication of the EOSC Report represents a watershed in the vision for the creation of a European, and in the long term a global, commons of FAIR research data. Johannes Gutenberg's invention of moveable type printing in the West at the end of the fifteenth century revolutionised the way ideas were recorded and disseminated. The Protestant Reformation, and the Counter Reformation, would not have been possible without the aid of the printing press. FAIR, Open Data and developments such as the European Open Science Cloud have the potential to have a similar impact in the 21st century. Studies such as the one undertaken by UCL, however, underline the challenges that need to be overcome to deliver the EOSC vision. Researchers are engaged on a journey, and it is the mission of the LEARN Toolkit of Good Practice to help them arrive at their chosen destination.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654139.

⁷ FORCE11: <https://www.force11.org/group/fairgroup/fairprinciples>; last accessed 8/1/17. For an early discussion, see M.D. Wilkinson, M. Dumontier, B. Mons et al., 'The FAIR Guiding Principles for scientific data management and stewardship' at Nature: <http://www.nature.com/articles/sdata201618>; last accessed 8/1/17. Published in *Scientific Data*, 3:160018, DOI: 10.1038/sdata.2016.18.

⁸ For a full description of this survey, see chapter 9 of this Toolkit, pp. 47-58.