

5 yrs EOSC. All about everything (adapted from Shakespeare)

Lightning talk: On the reasons why EOSC is still needed (20')

Jean-Claude Burgelman, Em. Professor Free University of Brussels; Editor In Chief Frontiers Policy Labs

Panel discussion with Barend Mons (CODATA), Michel Schouppe (EC) and Ilire Hasani Mavriqi (designated Chair of Austrian EOSC Mandated Organisation) (10')

1. Intro “On the reasons why EOSC is still needed. Such a long way ahead of us”

The title is not a reference to a Greek drama¹, but to a book written by a famous South African writer, Andre Brink, where he tells the story about a black and white guy, escaping from somewhere in SA and by helping each other, getting out of it. However they conclude that even though the benefit of collaboration got them out of the desert, it was still a long way to go for South Africa to become apartheid free.

If I translate this to today and to what I, as an observer, want to say about 5 years of EOSC.

Today there are even more challenges and needs for EOSC then when we – all of the stakeholders - started this project 5 or more years ago on the one hand, but on the other hand we still have so much to do to make it a truly pan European service.

In the discussions to set up this meeting I noticed with pleasure that most of us shared this and I would like to congratulate the Austrian team – and in particular Paolo Budroni and Stefan Hanslek – for taking this initiative which should not only be a celebration but must also become a reset for EOSC to be delivered.

2. 5 years of EOSC: Is there anything to celebrate?

Yes.

Compared to when the idea of EOSC was launched - officially 5 years ago, but the work on it started 9 years ago² - Europe has several hundreds of academics and experts working on EOSC, thousands across Europe make up an academic EOSC community (though it might melt away quickly if the money of the EC would dry up), most science policy makers in Europe are aware of the need for something like EOSC (and of open science), Europe has inspired the US to move ahead³, big and well respected research institutes like CERN are fully aligned with the need for open science and EOSC, Europe also inspired China (they have a kind of COSC in the meantime) and most importantly: the need for the web of fair data and services in science – the *raison d’être* of EOSC - is now widely accepted. In fact this web is being rolled out and that in itself is a major achievement

¹ J.C. Burgelman (2021) Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story) *Data Intelligence* (2021) 3 (1): 5–19.

² P. Budroni et al (2019) Architectures of Knowledge: The European Open Science Cloud. *ABI Technik* 2019; 39(2): 130–141 (with Michel Schouppe);

³ see the recent, June 2023, AI bill, which in a “disguised” way lays the foundation for an US version of EOSC)

So – it deserves a drink to have achieved that in 5/9 years.

But I hesitate to open a magnum champagne. Why?

To start with, most of the end users, the main clients, the first, second, third and last reason why EOSC was started, ai. the 2 million scientists of Europe, still don't know it, let alone are familiar with the principles of FAIR and open data⁴.

A lot can be said why this is so, but in my view and why I hesitate to be too joyfull, it has to do with the fact that EOSC still isn't a service available like e.g. Eduroam.

When I open my lap top in an academic environment, I can, as an authenticated user, connect immediately to Eduroam. The same smooth entry level is what EOSC needs we are far from it in Europe. Only now a procurement is launched for the EU Node for 35 million; 5 years after launching the EU wide agreed start of EOSC.

I had expected this to be much more advanced. Because when we started to work on EOSC, in 2014, there were obvious reasons to hurry up. And these reasons are still there at an even more pressing mode:

3. The reasons for EOSC

The need for EOSC was justified as for 2 categories of reasons⁵.

First as a strategy to maximize return on investment of publicly funded science - the money the public gives to the world of science - by making the results of it, in particular, the underlying data, FAIR. In doing it was expected to keep European science at the top of world in science production given the drive to data driven science (and the scientific challenges facing us).

Secondly right from the start EOSC was seen as the only way to avoid that the big cloud providers (all of them were and still are non-European) would monetize the hosting and service building of the European publicly funded data production.

Very simple and very straightforward ambitions and today none of these urgencies have disappeared. Au contraire.

Making sure the building blocks for the science of the future are in an "open/fair" mode will be as paramount today for the future of science as a global collaborative platform than in the last decade. In fact the stakes are even bigger than 10 years ago.⁶

On the one hand because the age of data driven science is there whilst on the other hand and at the same time, the age of data driven science production (data and article) has

⁴ Most data in this article can be found at <https://eoscobservatory.eosc-portal.eu/home>

⁵ J.C. Burgelman et al (2019) Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century *Frontiers in big data*. (with C. Pascu, K. Repanas, M. Schoupe, K. Szkuta, A. Karalopoulos, R. Von Schomberg);

⁶ J.C. Burgelman (2023) [Getting a grip on data and Artificial Intelligence](https://doi.org/10.25453/plabs.22787081.v1). *Frontiers policy labs* 2023. <https://doi.org/10.25453/plabs.22787081.v1>

arrived too. AI, and in particular the LLM that give it smartness so to speak, is therefore a game changer for the way science is made and thus for the data underlying it and will undoubtedly lead to an explosion in the volume of output of science and in the role machines will play to make our science.

It follows, that the need for fair data governed in an equitable way – which is nothing more than keeping the entry levels to science as low as possible - is even more acute than 10 years ago.

But – the more we insist on FAIR data -also meaning Fully AI ready - to make sure the science of the 21st century is “open”, the more interesting it becomes to monetize these data.

This explains why publishers are moving into this field too and this wasn't the case when EOSC was started.

They do this in 2 ways⁷. On the one hand by offering services an EOSC should/could do (mainly by buying start ups) and in doing so create a lock in for scientists using them. On the other hand by making sure there is a premium connected to the use of the data they sit on. The latter, again in 2 ways.

First, the data underlying science is a gold mine for a publisher. But as the world of publishing is increasingly an oligopoly this makes it very easy to make sure that no new players come in to join the gold digging and in such a context it remains to be seen if EOSC will be able to orderly govern the gold digging that is already going on.

Second, one of the preferred sources for LLM to be trained are clean texts and data. The producer by excellence of the latter is indeed science as its outputs are verified and validated as much as it can. So why would a big publisher, sitting on XX or so per cent of the world production of science, not monetize this resource by asking access fees to use them as training material. Here again it remains to be seen how EOSC as it is now can get a grip on what no doubt will become a big business. In fact it should be examined with urgency if already EOSC will be too late to play a significant role here.

I also can't see why the big cloud providers would not come into play here. Here too, as far as I can document it, most European universities already now host their data on big (non-European) cloud providers - what we wanted to avoid, by having a grip on it, in 2014. And it is highly predictable that these contracts stipulate exclusivity. There is a reason why the big publishers fear first and foremost not their direct competitors or the EC for that matter, but first and foremost players like Google.

In any case, the catalogue of services tailored to the needs of a researcher that the big cloud providers already offer is so exhaustive and attractive, that I can't see how we can still compete with that. Let alone why we should use public money to offer what is already available. In the set up discussions of EOSC it was always considered a win-win if EOSC would

⁷ J.C. burgelman (2021) Scholarly Publishing needs regulation. Creeping oligopoly threatens value, innovation, and the flow of knowledge. Research Professional Europe, 28-1-2021.

tender at maximum what private players provide anyhow, so that EOSC resources could be used to focus on core issues (like governance).

And finally, on the even bigger urgency for having something like EOSC: if machines – based on AI – write most of the science of the future – will these AI companies then own the articles/data too? This btw also underlines the urgency to make sure that in data driven science we can trace the provenance of the data – and there FAIR is a sine qua non.

In other words, looking at what is happening in the real world of cloud computing, data publishing and the monetizing of science – these 3 are the new “dynamic” data policies have to be developed in - we should speed up EOSC as much as we can.

But looking at what was done up till now, a lot of speed needs to be injected.

4. is the momentum for EOSC lost?

On the one hand it is tempting to say the momentum is may be lost (note that I don't say yes).

When eosC was launched 5 years ago, all was there to make it happen quickly. In 2018 EOSC had a clear top level mandate supported from all Member States, a clear business case (no exclusive foreign cloud hosting, better return on investment), a large support of the stakeholders (coalition of the willing) etc.

In fact the EOSC community developed then, avant la lettre, what today is the mantra of Europe technology policy: making sure has technological sovereignty.

If this mandate was so clear, why then do we still don't have it?

Here a lot can be said but in my view the core issue is that what was initially a simple agenda (open data in a fair format, allowing return on investmet etc) became a far too large to handle scope of ambitions for EOSC.

A less elegant way is to say that it EOSC became a Xmas tree for everyone in the European science data world to push his or her agenda – irrespective of the value it could have for EOSC

Scanning the many meetings and workshops held with under the EOSC label, I read, in my view, far too many other ambitions that are now linked to the initial policy agenda: realising trust, reviewing the assessment system, promoting open science, addressing the disparities at the global level, linking it to all the Research Infrastructures even federating with HPC⁸

⁸ At the latest Madrid symposium <https://eosC.eu/events/eosC-symposium-2023/> a slide was shown <https://symposium23.eosCfuture.eu/symposium/opening-plenary/> where an official stakeholder announced that EOSC will federate with EURO HPC . This was never the ambition, in fact we resisted to that at all costs as it is most likely a recipe for downplaying EOSC. There is, as far as I know, not 1 success story in the digital world where a service (EOSC) is successfully federated with an infrastructure (HPC).

In enlarging the scope of ambitions, one creates the perfect setting for a standstill because making all these ambitions happen is mainly to be done on a self regulatory basis. The latter implies that realising an enormous scope (the extended OS agenda as it is now) in a bottom up way, without no instruments nor institutions to leverage decisions or to be held accountable for, is most likely a recipe for achieving very little.

We are very good in Europe in making these Xmas trees. The same seems to happen now with the recent very ambitious AI plans. As recently stated by M. Irgens (board member of Adra, the European AI, Data and Robotics Association; and pro-rector of Kristiania University College) and E. Girardi (president of Adra and founder of POP AI)

“When Ursula von der Leyen took on the presidency in 2019, she made AI a focus and provided significant funding. But now almost four years later, the gap between Europe and the US has arguably grown. There are two primary reasons for this. First, EU research funding has been spread thinly across thousands of poorly aligned projects spread across Europe and carried out by researchers who are not committed full time the projects. The second, linked problem, is that member states are vying with each other for dominance in AI, leading to further fragmentation of funding and a lack of will to concentrate effort to concentrate effort. Overcoming national interests and working at scale are necessary components for a successful sector. As Europe slides further behind the US, it has turned its focus to regulation, seeking to set the global agenda through the proposed AI Act for the EU”⁹

Momentum lost?

May be not, if reset and speed up the machinery for EOSC. In the first place because the need for a distributed governance system and management layer for fair data is much bigger now and much more acute given the rent seeking behavior of the private players (on top of all the other arguments to support EOSC).

It is therefore essential to start designing and working on EOSC – a Minimal Viable/valuable EOSC? - as an implementation project (not a scientific problem) which addresses the European ambitions of technological sovereignty, but in such a way it suits science best: tech sovereignty without becoming a fortress¹⁰.

To reset successfully I would suggest to focus on

- Deciding on a clear and very ambitious delivery date for EOSC as an Eduroam alike reality;
- Setting up an accountable governance system which needs to be performance based;
- Consider turning the Association in to not for profit company to make it happen and harness the advantage of think public but acting as a private business.

⁹ Science Business 22/8/2023.

¹⁰ We explored this in depth in L. Soete and J.C. Burgelman (2023) Reconciling Open Science with Technological Sovereignty: Can the European Union do it? in JOAL, Journal for OA and Law, Vol. 11 No. 1 (2023): Special Issue on "Open Science and Data Protection".

To conclude and go back to the title: looking at how science develops (data and AI driven); at how this will make science even more attractive for business; at the changing geopolitics where it is our responsibility as scientists to make sure science can remain global; it's a no brainer we need EOSC even more than before, but there is still a long way ahead of us to go.

But if we stop making the road to walk longer ourselves (the Xmas tree) and if we start walking the talk rather than talking all the walking we want to do (by simply start implementing EOSC as a not for profit business) we can still do it.

To quote Nike - btw a Greek Goddess too – “just do it!” so that when we come back here in 5 yrs we can open our laptop and are automatically logged in into a pan European service, called EOSC.