

Comments to Jean-Claude Burgelman's article *Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story)*

—“EOSC is a bigger ME” and the Dunning Kruger effect

Barend Mons^{1,2†}

¹Leiden University Medical Center, Leiden 2333 ZA, The Netherlands

²GO FAIR International Support & Coordination Office (GFISCO), Leiden 2333 AA, The Netherlands

Citation: Mons, B.: Comments to Jean-Claude Burgelman's article *Politics and Open Science: How the European Open Science Cloud became reality (the untold story)*—“EOSC is a bigger ME” and the Dunning Kruger effect. *Data Intelligence* 3(1), 32-39 (2021). doi: 10.1162/dint_a_00074

PREAMBLE

This personal reaction is written from multiple perspectives. First and foremost, as the corresponding author of the original FAIR article. Second as the chair of the first High Level Expert Group (HLEG) of European Open Science Cloud (EOSC) (which is how I met Jean-Claude) and third from my current GO FAIR and CODATA perspective. None of what I write below is to be seen as a formal position of any of the organisations I am associated with.

Let me start by stating that, after some periods silent of hope and of deep despair, I now strongly feel that, with the governance of the EOSC Association in place, EOSC will become a success after all. It will still be critical that the Association involves the member states (MSs) and actual researchers in an agile and non-bureaucratic manner, for which we need bottom-up mechanisms such as operated by the Research Data Alliance (RDA) and GO FAIR. But a balancing formal entity operating along the formalised Strategic Research and Innovation Agenda [1] and the Partnership proposal as well as the various “declarations” including the recent one under the German presidency [2] are an excellent guiding roadmap to a successful EOSC, obviously in global context.

That said, at the risk of sounding like broken record, this reaction should also look at the points where it went “almost” wrong, as we should try and learn from our mistakes. I may make some enemies—or

[†] Corresponding author: Barend Mons (E-mail: barend.mons@go-fair.org; ORCID: 0000-0003-3934-0072).

strengthen the opinion of existing ones—in the process, but then, a wise old friend, who also wrote one of the reactions once told me: “Barend, unless you made some enemies you probably lived in vain.” So I will speak my mind (“what’s new?”). I also like to say that “EOSC” brought me some real new friends for life!

First of all, the fact that quickly after its inception FAIR became a hype term[Ⓞ], which was probably partly even accelerated by the prominent role it played in early EOSC discussions with EC’s Director General, also has its downsides. Like for the term “AI”, everyone co-opts the term and some start watering the concept down to a bloodless caricature from what it originally meant. In the case of FAIR this includes removing the central notion of *machine actionability*, mis-characterising it as a *standard*, conflating it with “open”, only linking it to *data sensu stricto*, ignoring *software*, *algorithms* and more. In general terms, people that sometimes seem to have never read the original article [3], the most flagrant abuse of the term I have heard (obviously not from an active researcher) is this: “If data are Findable, Accessible and Interoperable it is “automatically” Reusable.” This is of course “swearing in FAIR church” as the R (principles R1-3) [3] clearly state that rich provenance and reuse conditions are critical and in particular the provenance. The decision whether (even high quality) data are *fit for purpose* (reuse in a particular study) is a critical step and is imho (in my humble opinion) at the basis of the reproducibility problem we currently face. Therefore, I would like to re-emphasize here my current one liner to summarise the aim of the FAIR guiding principles: “The Machine Knows what I mean”. Those who feel that FAIR is too ambitious and for instance promote that “achieving F and A is enough for now” in my humble opinion fail to see the disruptive character of the solutions we need to make EOSC and its sister around the globe a real paradigm shift towards Open Science (OS). Or they are just trying to preserve the *status quo* and move incrementally at a pace they can follow.

This nicely bridges to the first observation on EOSC as such. I indeed think that the first “Communication” that needed 126 iterations mentioned by Jean-Claude, which happened in the same time frame as our “HLEG-1” period, was symptomatic for a basic flaw in the discussions, which haunts us still today. Conflating the “ICT”/HPC (or basic e-infrastructure) with the *data and end user applications for analytics*, has caused an enormous hurdle. In the entire journey of the HLEG we had to carefully navigate around this cliff and it is still a highly controversial issue today. This part was the “Dunning Kruger effect” [4] *pur sang*: The “other side is easy” (*because I am not hindered by any knowledge about it*) and is “more or less already done” (*because I do not understand the complexity*). This is not only true for the active researchers who cannot use the current e-infrastructure efficiently (and naturally that is “entirely the fault of the nerds who build things I do not understand or cannot operate”), but also for e-infrastructure engineers who know everything about ICT and “thus” (?) *also about data* (*because “that is just ones and zeros”*) as Jean- Claude also noted. I also believe however, that it is a mistake to completely separate e-infrastructure for the data and services layer, as the e-infrastructure should route (and understand at least at middleware level) what processes are needed on the data and how the FAIR services “run”. Nowadays (after many iterations) I use

[Ⓞ] Apart from the 3,500 + citations on the mothership paper, many declarations contain the word FAIR in many contexts.

the diagram below (Figure 1) to explain that all three basic elements of the “Internet of FAIR Data and Services” are needed. Each of them should be adorned with FAIR (machine actionable) metadata to seamlessly form a Web of FAIR Data and Services on top of the current, proven Internet backbones, thus forming the “Internet of FAIR Data and Services”, eventually creating an “Internet for Social Machines” [5] where people and machines can both efficiently use all services, independently and in collaboration.

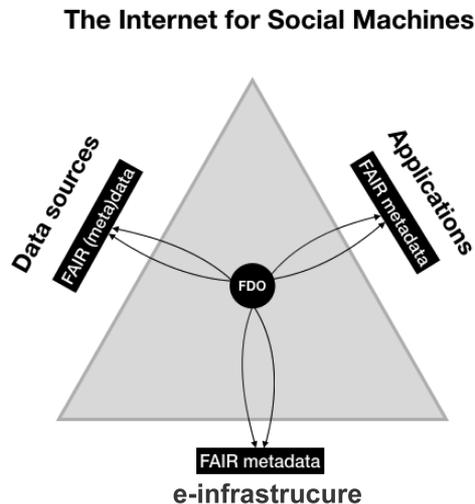


Figure 1. The final aim: The Internet for Social Machines, which enables seamless collaboration of people and computers should be based on minimal, but rigorously required protocols and agreements. The current infrastructure that supports the Internet and the Web applications we know, should be reused as much as possible, including its basic operation on TCP/IP and domain names. What needs to be added to realise the Internet for FAIR Data and Services on top of the current Internet is a Web of FAIR Data and Services. Applications (regardless of whether they work under parental guidance of people or not) should be able to Find, Access, Interoperate and (if relevant) Reuse data (and associated applications). Increasingly (virtual) machines will operate largely independently from direct human interaction and therefore two basic elements are absolutely critical to make this all happen (comparable to the centrality of TCP/IP). 1. All elements (including the e-infraestructure) should be adorned with rich and machine readable (FAIR compliant) META-data, and 2. All elements of the Web of FAIR Data and Services should be composed of *FAIR Digital Objects (FDOs)* [6,7,8].

This does absolutely not mean that the foundation (e-infraestructure) of the triangle is “trivial” or “can be reused as is”. Not only middleware, but also the crucial and fundamental concept of FDOs needs to be developed in close collaboration between data and computer experts and is largely domain-agnostic.

The seamless combination will become the principle “package” of information that machines (and also people) can understand and act upon. Major infrastructure builders should actually co-lead this, while domain scientists need to decide on which data formats and metadata schemes (i.e. FAIR Implementation Profiles [9]) should be built on this basic schema.

“EOSC IS A BIGGER ME”

Together with the Dunning Kruger effect, too many overlapping and redundant projects supporting the talking/meeting/landscaping, re-landscaping and re-re landscaping' has resulted in what I became to call the “EOSC is a bigger Me syndrome”. On the one hand, countless people voluntarily invested (and still invest) their time in the development of the EOSC, but others seem to only see EOSC as “yet another way to collect EC funding for their current solutions that are in my opinion not future- and OS proof. This misbalance between people investing their own time and effort based on intrinsic motivation and vision and on the other hand the “reliance on EC subsidy” caused a dichotomy during the scoping years of EOSC between disruptive and “preservative” approaches. The heavy reliance on EC subsidy also largely ignored the *subsidiarity principle* [10] and the fact that 90% of the eventual infrastructures and services that we need for EOSC will be paid by the MSs. Also data and research intensive industry was largely kept out of the loop, which was another mistake I have frequently pointed out. This helped to create and sustain the “Brussels Bubble” that Jean-Claude described. The Association will hopefully reverse that trend.

Finally, the influence on the HLEG report of the then-commissioner was rather profound. The report was not only delayed almost 6 months after its proposed publication version, but there is also a nice additional “untold story” here: The originally proposed title of the report was: “**A Cloud on the 2020 Horizon**”. In my original foreword I explained the slightly “glooming” connotation of that title. When the report was finally approved, it appeared that the title had been unilaterally changed into “*Realising A European Open Science Cloud* [11]”. Not only did I have to hastily change my foreword (because it made no sense anymore) but also, my notorious statement that the “result” should neither be “European” (only), nor Open (only) nor (only) for Science and certainly not (just) a “Cloud” was entirely ignored in changing that title. But it again emphasises the “This is an EC thing” context, with the associated risk for confiscation of the concept by the “usual suspects” in EC subsidy land. However, I feel after three years of intensive deliberations, which may be considered lightning fast on the geological time scale, see George's reaction, we can conclude that most of the original HLEG recommendations are well-represented in the basic guiding documents of the EOSC Association, which makes me a happy man at the end of this crazy year.

That leads me to the final observation: As a result of the (quote from Jean-Claude): “non-paper seen as the political turning point in support of EOSC” [12], GO FAIR (Global Open FAIR) [13] was started, originally by Germany and The Netherlands and soon joined by France as a temporary “kick-start”, bottom-up approach to accelerate EOSC (see also recommendation I-2.1. in the HLEG report, annex 1).

Soon, GO FAIR became really global and the agile *modus operandi* of practical Implementation Networks yielded a number of crucial approaches to speed up the adoption of the FAIR guiding principles and the hourglass approach [14]. Now, late 2020, when the EOSC Association is a fact, GO FAIR (1.0) has achieved its goals (early implementation steps) and we need to reflect on its future. Next to the intrinsic value of the active GO FAIR IN community [15] as such, several particular *assets* that I need to mention here are the development of the FAIR Implementation Profile and Metadata4Machines approach, the development of easy to install FAIR data points for open, FAIR metadata publication and indexing, and last but not least the international effort (involving many players, also outside the direct GO FAIR initiative) to develop the

minimal specs of the FDO framework [7] in a more specified form than when coined in the FAIR expert group report [5]. These assets (all open source and open access) can be carried over, not only to EOSC, but will have much wider, international, impact most likely leading to a continuation of GO FAIR (2.0) beyond its original time scope, namely three years, the predicted time it would take to complete the international policy and bureaucracy process to reach the status of a formal association as we have today. I hope the leaders of the Association will optimally learn from the successes and failures and near-road-accidents of the last three years and see EOSC as the European contribution to a “Global Open Science Commons”, also known as the Internet of FAIR Data and Services, in full, open collaboration with the international organisations that are now joining forces in the Data Together initiative [16]. After all, the major challenges we face are global, so is the research needed to face them and so are the solutions we hope to find. I fully trust the current leadership of the association to make that vision reality.

REFERENCES

- [1] The EOSC strategic research and innovation agenda—Version 0.8. Available at: <https://www.eosc-portal.eu/news/eosc-strategic-research-and-innovation-agenda-version-08>. Accessed 29 December 2020
- [2] Opening the door to a world of FAIR research digital objects—On the threshold to the European Open Science Cloud. Available at: <https://www.bildung-forschung.digital/files/EOSC%20Declaration%20Opening%20the%20Doors%20to%20a%20FAIR%20World.pdf>. Accessed 29 December 2020
- [3] Wilkinson, M., et al.: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3, Article No. 160018 (2016)
- [4] Dunning Kruger effect. Available at: https://en.wikipedia.org/wiki/Dunning-Kruger_effect. Accessed 29 December 2020
- [5] Hendler, J., Berners-Lee, T. From the semantic web to social machines: A research challenge for AI on the World Wide Web. *Artificial Intelligence*, 174(2), 156–161. Available at: <https://www.sciencedirect.com/science/article/pii/S0004370209001404>. Accessed 29 December 2020
- [6] Directorate-General for Research and Innovation (European Commission). Turning FAIR into reality. Final report and action plan from the European Commission expert group on FAIR data. Available at: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf. Accessed 29 December 2020
- [7] De Smedt, K., Koureas, D., Wittenburg, P.: FAIR digital objects for science: From data pieces to actionable knowledge units. *Publications*, 8(2) 21, (2020)
- [8] Bonino da Silva Santos, L.O. (ed.): FAIR digital object framework documentation. Available at: <https://fairdigitalobjectframework.org/>. Accessed 29 December 2020
- [9] Magagna, B., et al.: Reusable FAIR implementation profiles as accelerators of FAIR convergence. Available at: <https://osf.io/2p85g>. Accessed 29 December 2020
- [10] The principle of subsidiarity. Available at: <https://www.europarl.europa.eu/factsheets/en/sheet/7/the-principle-of-subsidiarity>. Accessed 29 December 2020
- [11] Directorate-General for Research and Innovation (European Commission). Realising the European open science cloud. First report and recommendations of the Commission high level expert group on the European open science cloud. Available at: <https://op.europa.eu/en/publication-detail/-/publication/2ec2eced-9ac5-11e6-868c-01aa75ed71a1>. Accessed 29 December 2020
- [12] Germany and the Netherlands call for rapid action on the European Open Science Cloud. Available at: <https://www.government.nl/latest/news/2017/05/30/germany-and-the-netherlands-call-for-rapid-action-on-the-european-open-science-cloud>. Accessed 29 December 2020

- [13] Go FAIR. Available at: www.go-fair.org. Accessed 29 December 2020
- [14] The Internet layer protocol. Available at: <https://www.informit.com/articles/article.aspx?p=27864&seqNum=4>. Accessed 29 December 2020
- [15] Implementation networks. Available at: <https://www.go-fair.org/implementation-networks/>. Accessed 29 December 2020
- [16] Data together: Articulating the collaboration among the four major international data organizations to optimize the research data ecosystem. Available at: https://www.go-fair.org/wp-content/uploads/2020/03/Data-Together_March-2020.pdf. Accessed 29 December 2020

AUTHOR BIOGRAPHY



Barend Mons is Professor of BioSemantics at the Human Genetics Department of Leiden University Medical Center in the Netherlands and founder of the BioSemantics group. He was elected CODATA President in 2018. Next to his leading role in the research of the group, Barend plays a leading role in the international development of “data stewardship” for biomedical data. For instance, he was head-of-node of ELIXIR-NL at the Dutch Techcentre for Life Sciences (until 2015), is Integrator Life Sciences at the Netherlands eScience Center, and board member of the Leiden Center of Data Science. In 2014, Barend initiated the FAIR data initiative and in 2015, he was appointed Chair of the European Commission’s High Level Expert Group for the “European Open Science Cloud”, from which he retired by the end of 2016. Presently, Barend is co-leading the GO FAIR initiative, an initiative to kick start developments towards the Internet of FAIR Data and Services, which will also contribute to the implementation of components of the European Open Science Cloud. The focus of the contribution of the BioSemantics group is on developing an interoperability backbone for biomedical applications in general and rare disease in particular.

ORCID: 0000-0003-3934-0072

Annex 1 (from HLEG 1) CHALLENGES AND GENERAL OBSERVATIONS

- 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 reuse.
- 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 MSs and across different communities.
- There is no dedicated and mandated effort or instrument to coordinate EOSC-type activities across MSs.

KEY FACTORS FOR THE EFFECTIVE DEVELOPMENT OF THE EOSC AS PART OF OS

- New modes of scholarly communication (with emphasis on machine actionability) need to be implemented.
- Modern reward and recognition practices need to support data sharing and re-use.
- Core data experts need to be trained and their career perspective significantly improved.
- Innovative, fit for purpose funding schemes are needed to support sustainable underpinning infrastructures and core resources.
- A real stimulus of multi-disciplinary collaboration requires specific measures in terms of review, funding and infrastructure.
- The transition from scientific insights towards innovation needs a dedicated support policy.
- The EOSC needs to be developed as a data infrastructure commons, that is an eco-system of infrastructures.
- Where possible, the EOSC should enable automation of data processing and thus machine actionability is key.
- Lightweight but internationally effective guiding governance should be developed.
- Key performance indicators should be developed for the EOSC.

SPECIFIC RECOMMENDATIONS TO THE COMMISSION FOR A PREPARATORY PHASE

Policy recommendations

- P1: Take immediate, affirmative action on the EOSC in close concert with MSs.
- P2: Close discussions about the “perceived need”.
- P3: Build on existing capacity and expertise where possible.
- P4: Frame the EOSC as the EU contribution to an Internet of FAIR Data and Services underpinned with open protocols.

Governance recommendations

- G1: Aim at the lightest possible, internationally effective governance.
- G2: Guidance only where guidance is due (this relates to technical issues, best practices and social change).
- G3: Define Rules of Engagement for service provision in the EOSC.
- G4: Federate the gems and amplify good practice.

Implementation recommendations

- I1: Turn the HLEG report into a high-level guide to scope and guide the EOSC initiative.
- I2: Develop, endorse and implement the Rules of Engagement for the EOSC.
- I2.1: Set initial guiding principles to kick-start the initiative as quickly as possible.
- I3: Fund a concerted effort to develop core data expertise in Europe.
- I4: Develop a concrete plan for the architecture of data interoperability of the EOSC.
- I5: Install an innovative guided funding scheme for the preparatory phase.
- I6: Make adequate data stewardship mandatory for all research proposals.
- I7: Provide a clear operational timeline to deal with the early preparatory phase of the EOSC.