Open? The only way forward for science

Elena Giglia*
* *University of Turin, UNITO, Open Science Unit*

Abstract. This paper aims to set out the reasons underlying the need to foster as much as possible the sharing and re-use of research data as well as their FAIRness, taking into account the various interests at stake. COVID-19 showed that sharing is the only way to go and that to advance science we need data – and every bit of the research process -, not only the final synthesis of the research itself, i.e., the article on a scientific journal. Scientific journals are still at the core of research evaluation, which is being reformed to include any research output and to reward collaboration. To be openly shared, data needs to be FAIR, i.e., Findable, Accessible, Interoperable and Reusable, in order to streamline the workflow, to enable reproducibility, and to booster research integrity. The paper presents the FAIR open data as one of the components of the wider Open Science ecosystem, which we shall discuss here not with the usual "connecting block" approach but with an ecological one, where the web of interactions within the ecosystem defines its elements rather than the opposite and where the focus of Open Science is on co-creating knowledge instead of only disseminating it. The data sharing fostered by the Open Science approach is certainly not indiscriminate, but rather follows the principle "as open as possible, as closed as necessary". A balancing act is required that takes into account the conflicting interests at stake, such as the right to the protection of personal data, enshrined in Article 8 of the Charter of Fundamental Rights of the European Union.

Keywords: Open Science, FAIR principles, EOSC, European Open Science Cloud, data protection, privacy, Open Access

1. Introduction: Lessons learned from COVID-19

"Open data save lives". If there is one reason why we should all care about Open Science, that's it. This sentence opened the 2021 Forward to the State of Open data report, and of course referred to the role played by data sharing during the COVID-19 pandemic. If anything good came from the pandemic, it was this unprecedented common effort in collaborating and in bringing down barriers to speed up discovery. Let's go over some lessons learned from COVID-19 to substantiate the assertion that "Open is the only way to go".

COVID-19 made it clear that we can't rely only on the final synthesis of a research (i.e., the scientific publication or article), but we need data – FAIR by design – and we need them immediately. FAIR represents an acronym and stands for Findable, Accessibile, Interoperable and Reusable: all the features that research data needs to possess in order to be well-structured by a technical point of view and, in the end,

re-usable. The Council of the European Union perfectly summarises the lessons learned from COVID-19 in its Conclusions on Research assessment and implementation of Open Science (EU Council 2022):

> UNDERLINES that the COVID-19 crisis has highlighted the benefits of Open Science and of immediate open access to research publications, enabling quicker ways of vetting the quality thereof, and of further expanding open access modalities, which has been key to developing rapid responses to the crisis by permitting quick access to new research results to combat the disease; RECOGNISES that the crisis has also highlighted the benefits of increased access to research data based on the FAIR principles.

Nevertheless, our scholarly communication system is still too focused on the final research output, the journal article, which proved to be of no good during the pandemic. An average publication time of 9-18 months (Bjork 2013) makes no sense during a pandemic. Preprints, i.e. the article in its version immediately as completed by the author, prior to the peer review process, played a crucial role in timing dissemination of results: according to some scholars, preprints will likely become a mainstay of modern biomedical research (Yong 2021) as they accelerate research. The World Health Organization included in its Living Guidelines less than 25 percent materials coming from traditional scientific journals, which is to say, in Robert Terry's words, they failed us the moment we needed them more [1]. Nevertheless, we spend every year billions in journal subscriptions, feeding a dysfunctional system just for the sake of prestige (Brembs 2021). As Lizzie Gadd (Gadd 2020) put at the beginning of the pandemic,

> The virus is reminding us that the purpose of scholarly communication is not to allocate credit for career advancement, and neither is it to keep publishers afloat. Scholarly communication is about, well, scholars communicating with each other, to share insights for the benefit of humanity. [...]
> If we've created a generation of scholars who are just in it for the glory of papers in glamorous journals, and not to do good research that changes the world a little bit, then we really are in trouble.

When we apply the same perverse logic of prestige to paid hybrid Open Access, we get Nature asking for 11.500 dollars for a single

[1] Terry, R., Time to unlock the full potential of digital age, Keynote speech, Open Science Fair 2021, video, https://www.youtube.com/watch?v=TrZrRcCoQSo.

article to be published openly (Saltzberg 2020), or we see a sharp increase in the so-called "transformative agreements". Instead of being an innovative solution, they turned out to be not sustainable and not equitable, still linked to the idea of science as a commodity (Becerril Garcia, 2023). For publications, the only viable solution is developing a scholarly publishing infrastructure that is equitable, community-driven, academic-led and -owned, while respecting the cultural, multilingual, and disciplinary diversity: in other words, the so-called "Diamond" Open Access[2].

Publishing in a "prestigious" venue is strictly linked to research assessment criteria. Perverse effects of these criteria like adaptive behaviours, scientific misconduct, "gaming the system" are well known and investigated (Biagioli 2020, Brembs 2018, Casadevall 2011). The Council of the European Union recognized it in its Conclusions Research assessment and implementation of Open Science (EU Council 2022):

> that research assessment systems should focus on quality and impact, and RECALLS that the current research assessment systems are nowadays to a great extent too focused on the use of some quantitative journal- and publication-based indicators and the evaluation of a narrow range of research outputs; CONSIDERS that such an approach may lead to negative biases in terms of research quality, reproducibility and integrity; STRESSES that research assessment should include other research outcomes and processes and promote early knowledge sharing and collaboration to accelerate the implementation of Open Science policies and practices.

The Council Conclusions constitute the legal basis of the Coalition for Advancing Research Assessment (COARA) initiative[3], launched in 2022 with the aim of reforming the research assessment, avoiding the misuse of quantitative indicators, and rewarding openness and collaboration. Reforming the criteria of research assessment can be a game changer in fostering the adoption of Open Science practices – how many times did we hear in the past "Yes, fine, Open Science is nice, but we are still evaluated only on Impact Factor?"

COARA's principles are (i) quality, (ii) openness, (iii) transparency, (iv) reproducibility, (v)respect of diversity of carriers and disciplines,

[2] Diamond Open Access Plan, https://www.scienceeurope.org/our-resources/action-plan-for-diamond-open-access/.

[3] COARA, https://coara.eu/.

and (vi) valorisation of different contributions. In addition, the COARA initiative, in line with Open Science and Open FAIR data:

> Recognise the diversity of research activities and practices, with a diversity of outputs, and reward early sharing and open collaboration. Consider tasks like peer review, training, mentoring and supervision of Ph.D candidates, leadership roles, and, as appropriate, science communication and interaction with society, entrepreneurship, knowledge valorisation, and industry-academia cooperation. Consider also the full range of research outputs, such as scientific publications, data, software, models, methods, theories, algorithms, protocols, workflows, exhibitions, strategies, policy contributions, etc., and reward research behaviour underpinning open science practices such as early knowledge and data sharing as well as open collaboration within science and collaboration with societal actors where appropriate. Recognise that researchers should not excel in all types of tasks and provide for a framework that allows researchers to contribute to the definition of their research goals and aspirations.[4]

Data sharing being recognized and rewarded can be the first step in easing the resistances that still researchers have in opening up their entire workflow (Digital Science 2021, Gomes 2022). COARA may represent a game changer in a scenario in which some barriers to data sharing are still perceived. In particular, next section illustrates how the FAIR principles (Wilkinson et al. 2016) can be a fruitful tool in order to fully understand the limits, modalities and development of research data sharing.

2. The value of FAIR Open data

Why do we talk "FAIR Open data" and not simply "Open data"? Because opening data which are not FAIR (i.e., Findable, Accessible, Interoperable, and Reusable) can be at risk of misuse of misunderstanding – all the more so if there is any documentation or licence missing under the "R" side.

Data must be:

1. Properly managed, as "good research needs good data"[5].

[4] COARA Agreement, Principles for assessment criteria and processes, page 4, https://coara.eu/app/uploads/2022/09/$2022_07_19_ra_agreement_final.pdf$.

[5] Digital Curation Centre motto, https://www.dcc.ac.uk/.

2. Made FAIR by design, from the very beginning of the research project, as specifically envisaged to participate in the European research ecosystem, the so-called European Open Science Cloud (EOSC), a virtual environment where scientists, data producers, service producers and innovators come together, facilitating the use of Artificial Intelligence (AI) systems or analysis techniques such as text and data mining.

3. Open, whenever possible, according to the principle "as open as possible, as closed as necessary".

Open research data are like renewable energies: they can be used without diminishing their value, and their reuse creates new value (Digital Science 2017). Open data creates bridges between disciplines: that's their main value and that's one of the reasons why the UNESCO in its Recommendations identify Open Science and FAIR Open data as an accelerator in the achievement of the United Nations Sustainable Development Goals (UNESCO 2021). Global challenges like the climate change can be tackled and solved only by an inter- and cross-disciplinary approach, enabled by FAIR Open data.

Opening research data benefits science in the first place, as transparency is a seal of integrity and methodology soundness, beside fostering reproducibility and avoiding frauds and retractions. FAIR Open data can then be reused, bringing benefits both to the researcher – increased visibility and credit – and the funders, avoiding duplications and waste of time and money and, in the end, maximising the return on investments. In biomedical research the issue is even more stringent: if data are not available and experiments are not reproducible, at best expensive research is of little or no value, at worst results of invalid research are put into clinical use[6].

Humanities and Social Sciences, far from being excluded, are an integral part of this process, even though particularly in the Humanities there might be still a low acceptance of the term "data". Research data can be defined broadly as "all materials and assets scholars collect, generate and use during all stages of the research cycle" (ALLEA 2020). For this reason,

> It is like in the case of Monsieur Jourdain, the title character of Molière's ***Le Bourgeois gentilhomme***, who learnt, to his great satisfaction, that unwittingly he had been speaking prose all his life. With research data in the humanities it is exactly the same: you are using it, even if you don't

[6] Weisteen Bjerde, Katrine (2023) Examples of benefits from investing in data infrastructures, presentation at the EOSC Symposium 2023, not available online.

> know it, and once you realise it, it will affect your research workflow forever (ALLEA 2020).

Is it simple and costless to FAIRify research data? No, of course not. There are costs – but think how much would it cost not to manage data and/or to lose them and think about how much time researcher spend to "clean" the data before using them (European Commission 2019). Allocating 5 per cent of research funds to data management – including hiring a data steward, which guarantees the maximum return on investment, is a precise responsibility of funders, as Barend Mons reminds us (Mons 2020):

> taking care of data is an ethical duty, and should be part of good research practice. Second, if data are treated properly, researchers will have significantly more time to do research. [...] Funders hold the stick: they should disburse no further funding without a properly reviewed and budgeted data-stewardship plan. The carrot is that FAIR data allow much more effective artificial intelligence (FAIR can also mean 'fully AI ready'), which will open up unprecedented research opportunities and increase reproducibility.

There are three important remarks that need to be considered here. First, if it's true that "good research needs good data" (as the UK Digital Curation Centre motto recites) it's even truer that AI should be trained on excellent data (on the role of the FAIR data principles for AI, see: (Paseri, 2022a).

Second, we are in the era of the European Open Science Cloud (EOSC), which is an environment to "unlock the full potential of research data to accelerate discovery and innovation"[7]. EOSC can be described as a web of FAIR data and services, in which data are easy to find and reuse within and across scientific disciplines[8]. As claimed above, global challenges like the COVID-19 pandemic or the climate change can only be solved by a cross-disciplinary, open approach. The new funding programme for research and innovation of the European Union, i.e., the Horizon Europe, embrace this approach. In particular, the Horizon Europe programme identifies a set of "Missions", each

[7] Luyben K.-Gunsenheimer U. (2022) Main Achievements of the Tripartite Collaboration Plans for 2023. Presentation at the EOSC Symposium in Prague, November 2022.https://symposium22.eoscfuture.eu/wp-content/uploads/2022/11/Copy-of-7.-Session$_{M}ain$ – $achievements$ – $tripartite_{L}uyben-Gunsenheimer20221115_{A}chievements_{E}OSC-A_{a}ll-slides.pdf$.

[8] EOSC on the European Commission webpage, https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc$_{e}n$.

one representing "a portfolio of actions – such as research projects, policy measures or even legislative initiatives – to achieve a measurable goal that could not be achieved through individual actions"[9], that fit perfectly with the Open Science approach. In addition, concerning the EOSC, the first of its objectives tree (EOSC Association 2022) is "Open Science practices and skills are rewarded and taught, becoming the new normal". Open Science is seen in the EOSC objectives tree as bringing benefits for science itself, in terms of trust and reproducibility, for industry, as a matter of innovation in services and products, and for society at large, widening the impact on real life.

The third consideration has to do with the scope of openness. The Open Science approach does not aim at an indiscriminate openness, but an openness that is always the result of a balancing of multiple interests (and rights) at stake. First and foremost, this balancing act takes place between the right to science, as enshrined in Article 27 of the Universal Declaration of Human Rights (and reaffirmed by UN General Comment 25/2020, see: Perrone, 2020), and the right to the protection of personal data (the risks of which have been analysed in: Dennis et al. 2019; Erb et al. 2021). However, scientific progress and the advancement of knowledge is not intended to harm individuals by encroaching on their personal sphere. Indeed, the balance of Open Science is embodied in the formula "as open as possible as closed as necessary". The first fundamental purpose is to ensure good management of research data, something that is now often lacking. This is why the FAIR data principles are so relevant: they guarantee good data management from a technical point of view (Paseri, 2022b) and promote awareness on the part of researchers.

The stress in "FAIR" is on the "R" element: by enabling reuse, new value is created, as

> It is all the more compelling to seize the opportunity presented by data for social and economic good, as data – unlike most economic resources – can be replicated at close to zero cost and its use by one person or organisation does not prevent the simultaneous use by another person or organisation (European Commission 2020).

In thinking about "reuse" we should never forget that what is "noise" for one researcher can be a "signal" for another one. This is the reason why anything should be open, as "We don't know which research papers [and I would add: research output, including data] that today

[9] https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe_en.

remain largely inaccessible could inspire solutions and bright ideas for tomorrow's challenges" (Roorick 2020).

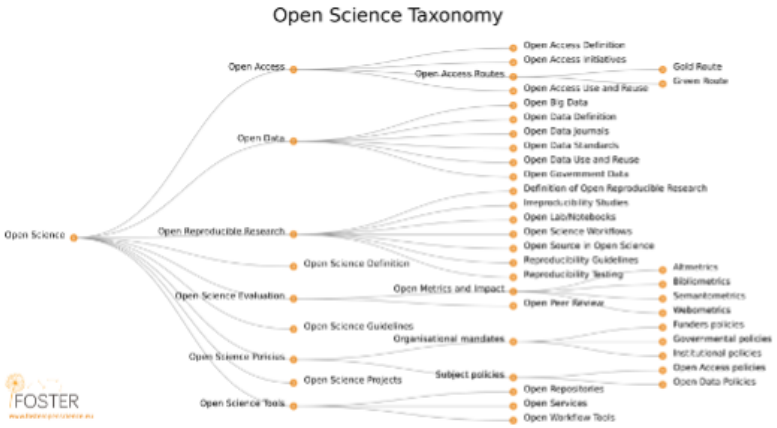
3. Conclusions: FAIR Open data in the Open ecosystem

FAIR Open research data are just one of the components of Open Science. A broad definition of Open Science is

> the transition to a new, more open and participatory way of conducting, publishing and evaluating scholarly research. Central to this concept is the goal of increasing cooperation and transparency in all research stages. This is achieved, among other ways, by sharing research data, publications, tools and results as early and open as possible. Open Science leads to more robust scientific results, to more efficient research and (faster) access to scientific results for everyone. This results in turn in greater societal and economic impact.[10]

As Open Science is an umbrella concept, the usual approach is to describe its components, even in a graphical way as the FOSTER taxonomy did (Pontika 2015).

Figure 1. FOSTER taxonomy.[11]



How are these components connected? Pierre Mounier suggests to adopt a holistic approach and start considering Open Science as an

[10] Open science definition, Qeios, https://doi.org/10.32388/838962.
[11] FOSTER Open Science taxonomy, https://www.fosteropenscience.eu/resources.

ecosystem producing a "milieu" of knowledge and supported by a vibrant community gathered around shared values (Mounier 2022). This leads to an ecological approach to Open Science where the web of interactions within the ecosystem defines its elements rather than the opposite. In this perspective, what matters is not "how" to connect these building blocks, but "why". Interoperability and interconnections are necessary but not sufficient, as they often result in a plethora of platforms and a kaleidoscope of interfaces "which leads to questioning if this supports the production of knowledge, or leads to a fragmentation of cognition" (Mounier 2022). This is important,

> because the small "crystals of knowledge" (Stern et al., 2015) that flows freely in the new digital open environment are mere meaningless pieces of information as long as they don't find their place in an organised cognitive environment that allows for a controlled activity of interpretation and the organisation of a scientific discussion that leads to a better and steadier constructed knowledge. This seamless cognitive environment researchers probably need in the open science context is still to come. "Connect the building blocks" of open science starts with driving the places of knowledge such as platforms, tools and services to work together, adapt to each other, up to a point where they can develop some sort of symbiotic relations to offer, all together, a fertile and meaningful "milieu of knowledge" (Dumas Primbault et al., 2021), rather than chaos (Mounier 2022).

The idea here is to develop a "seamless" research environment that encompasses FAIR open data, publications, or any other research output. The reality is that the fragmentation in several tools and platforms is coupled with fragmentation in policies, stakeholders, assessment criteria which creates a jungle for researchers. Competition over collaboration is another tension in science, but Mounier underlines how Open Science should be a rallying cry around the core values of science, and a way to build "the collective of all those who concur to knowledge creation, primarily as a community, and not only as competitors (Neylon et al., 2019). It is a way to reassert the centrality of knowledge as a common good for the benefit of all and not as an asset for the exclusive advantage of some" (Mounier 2022). Knowledge as a commons is the idea underpinning the UNESCO Recommendations, coupled with the idea of open dialogue with other knowledge systems. Dialogue is a two ways, bidirectional conversation. Reducing Open Science at its end point of "sharing" research outputs is still the old academia graciously

make gift of its wisdom – and diminishing its co-creation potential. We should focus more on the process of creating knowledge than on disseminating it: what is important is co-creating science, as well as opening up every step of the research process, for a science in real dialogue with society. In this perspective, "'Connecting the building blocks' of open science is thus much more than just creating connections: it is more than ensuring technical interoperability between different systems, more than coordinating various stakeholders, more than disseminating science in society: it is to create a milieu of knowledge, to build the community that supports it and to open it beyond the limits of academia" (Mounier 2022). If we start considering Open Science as an ecosystem supporting the creation of open knowledge, the connections became crucial. As Pierre Mounier (Mounier 2022) clearly deemed:

> The objective is no more to "connect the building blocks" of open science, as bricks are assembled in a wall, but to support symbiotic systems of relations between initiatives, platforms, tools, communities and practices that thrive for and by open knowledge. Which means, when considering or even evaluating open science initiatives, projects, services and tools, to flip the order or priorities and to pay attention first to the way they move in their ecosystem: how do they nurture from it, how do they fertilise it, how do they cooperate with others, rather than other criteria that are usually considered as more important; such as innovation, efficiency, excellence. And then, when we have a comprehensive representation of the full web of interactions and interdependencies, maybe we could start asking the right questions: is it sustainable? Is it inclusive? Is it creative? Is it alive?

Acknowledgements

In this paper I am particularly indebted to Pierre Mounier and his profound reasoning. I am also so grateful for all the marvellous Open Science people who crossed their paths with mine in the last 15 years, sharing ideas and enthusiasm – or a good glass of wine - and teaching me a lot. I tried to do my best to pass on their lessons and inspiring views, so I am grateful, too, to all the people who attended my training courses, sharing new insights and enlightening thoughts - again teaching me a lot.

References

ALLEA (2020) Sustainable and FAIR Data Sharing in the Humanities. Report. https://doi.org/10.7486/DRI.tq582c863.

Becerril-Garcia, A. (2023) The unique opportunity to advance Science as a Global Public Good: Open Science in a world of contrasts, keynote, 3rd United Nations Open Science conference.

Biagioli M. and Lippman A. ed. (2020) Gaming the Metrics: Misconduct and Manipulation in Academic Research, The MIT Press https://doi.org/10.7551/mitpress/11087.001.0001.

Bo-Christer Björk, David Solomon, The publishing delay in scholarly peer-reviewed journals, Journal of Informetrics, Volume 7, Issue 4, 2013, Pages 914-923,https://doi.org/10.1016/j.joi.2013.09.001.

Brembs, B. (2018) Prestigious Science Journals Struggle to Reach Even Average Reliability, Frontiers in Human Neuroscience, 12, https://www.frontiersin.org/articles/10.3389/fnhum.2018.00037.

Björn Brembs, Philippe Huneman, Felix Schönbrodt, Gustav Nilsonne, Toma Susi, Renke Siems, Pandelis Perakakis, Varvara Trachana, Lai Ma, Sara Rodriguez-Cuadrado. (2021). Replacing academic journals. https://doi.org/10.5281/zenodo.5526635

Dennis, S., et al. (2019) Privacy versus open science. Behavior research methods 51, pp. 1839-1848.

Digital Science (2021) The State of Open Data 2021. Report. https://doi.org/10.6084/m9.figshare.17061347.v1

Digital Science (2017): The State of Open Data Report Report. https://doi.org/10.6084/m9.figshare.5481187.v1

Dumas Primbault, S., Tortosa, P.-A., Vailly, M. (2021). Introduction—Milieux, media, écologie des savoirs. Cahiers François Viète, III 10, Art. III 10.https://doi.org/10.4000/cahierscfv.381.

EOSC Association (2022) Strategic Research and Innovation Agenda, https://eosc.eu/sria-mar

Erb, B., et al. (2021) Emerging Privacy Issues in Times of Open Science. Psyarxiv, pp. 1-5.

European Commission (2020) Communication to the European Parliament, the Council, the European Economic tnd Social Committee and the Committee of the Regions A European strategy for data COM/2020/66 final a https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066.

European Commission, Directorate-General for Research and Innovation, (2019) Cost-benefit analysis for FAIR research data : cost of not having FAIR research data, Publications Office, https://data.europa.eu/doi/10.2777/02999.

European Council (2022) Conclusions on Research assessment and implementation of Open Science", adopted by the

Council at its 3877th meeting held on 10 June 2022. https://www.consilium.europa.eu/media/56958/st10126-en22.pdf.

Fang, F. C., and Casadevall, A. (2011). Retracted science and the retraction index. Infectious Immunology 79, 3855–3859. doi: 10.1128/IAI.05661-11 https://journals.asm.org/doi/10.1128/IAI.05661-11.

Gadd, Lizzie The purpose of publications in a pandemic and beyond, WONKHE, 20 April 2020.

Gomes Dylan G. E., Pottier Patrice, Crystal-Ornelas Robert, Hudgins Emma J., Foroughirad Vivienne, Sánchez-Reyes Luna L., Turba Rachel, Martinez Paula

Andrea, Moreau David, Bertram Michael G., Smout Cooper A. and Gaynor Kaitlyn M. (2022) Why don't we share data and code? Perceived barriers and benefits to public archiving practices Proc. R. Soc. B.2892022111320221113.

Mons, B. Invest 5 % of research funds in ensuring data are reusable (2020) Nature 578, 491 doi: https://doi.org/10.1038/d41586-020-00505-7.

Mounier, P. (2022). Connecting the building blocks of Open Science: an ecological approach. Septentrio Conference Series, (1). https://doi.org/10.7557/5.6772.

Paseri, L. (2022a), *Il ruolo delle istituzioni in design, sviluppo e applicazione dell'IA per il settore della ricerca scientifica*, In Queste istituzioni 4, pp. 213-226.

Paseri, L. (2022b), *Accessibilità al sapere*, In Durante, M. and Pagallo, U. (eds.), "La Politica dei dati, Il governo delle nuove tecnologie tra diritto, economia e società", Mimesis Editore, Torino-Udine, pp. 141-162.

Perrone, G. (2020) Scienza e diritti economici, sociali e culturali: Il Commento generale n. 25 del Comitato dei diritti economici, sociali e culturali. Diritti umani e diritto internazionale 3, pp. 786-795.

Pontika, N., Knoth, P., Cancellieri, M., Pearce, S. (2015). Fostering open science to research using a taxonomy and an eLearning portal. Proceedings of the 15th International Conference on Knowledge Technologies and Data-Driven Business, 1 8. https://doi.org/10.1145/2809563.2809571.

Roorick, Johan Open Access lessons during Covid-19: No lockdown for research results!. Blog post PlanS blog, June 8, 2020 https://www.coalition-s.org/blog/open-access-lessons-during-covid-19-no-lockdown-for-research-results/

Saltzberg, S. (2020) Prestige Journal Publisher, Nature, Slaps Scientists In The Face, Forbes, Dec. 7, https://www.forbes.com/sites/stevensalzberg/2020/12/07/prestige-journal-publisher-nature-slaps-scientists-in-the-face/?sh=17547b9e5ae4.

Stern, N., Guédon, J.-C., Jensen, T. W. (2015). Crystals of Knowledge Production. An Intercontinental Conversation about Open Science and the Humanities. Nordic Perspectives on Open Science, 1(0), 1 24.https://doi.org/10.7557/11.3619.

UNESCO Recommendations on Open Science (2021), https://en.unesco.org/science-sustainable-future/open-science/recommendation.

Yong, E. (2021) How science beat the virus and what it lost in the process, The Atlantic, Jan./Feb., https://www.theatlantic.com/magazine/archive/2021/01/science-covid-19-manhattan-project/617262/ Wilkinson, M. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 https://doi.org/10.1038/sdata.2016.18.