**The hidden environmental impact of Open Science publishing**

In 2021 UNESCO released the Recommendation on Open Science. Signed by 193 countries, the Recommendation affirmed the importance of Open Science as “a vital tool to improve the quality and accessibility of both scientific outputs and scientific process, to bridge the science, technology and innovation gaps between and within countries and to fulfill the human right of access to science” (UNESCO, 2022).

The Recommendation strongly endorses the collective benefit of open research practices as a means of speeding up research for the Sustainable Development Goals (SDGs). Nonetheless, in light of the SDGs it becomes important to question how this collective benefit should be understood - from environmental as well as social perspectives. This, in turn, raises questions. In particular, what is the carbon footprint of the evolving practices of Open Science, and can the acceleration of climate-focused research be used as an excuse for rising carbon costs in research? This short commentary introduces some topics for consideration and advocates for critical self-reflection around the digital footprint of open research.

*Developments within academic publication*

Traditional academic publishing has been dominated by a small number of commercial companies for many decades. This dominance, together with incentive systems skewed towards academic publications, has led to widespread dissatisfaction with the dominant role of the traditional journal within academia. Business models of commercial academic publishers have significant implications for access to published research, control over intellectual property and over-reliance on (seemingly) free academic labour for reviewing and editing practices.

Over the last decade the Open Access movement has been extremely active in addressing these economic imbalances. This has included addressing the economics of participation and Open Access, the economics of transformative deals and publisher oligopolies. Creative initiatives have focused on the critical revision of traditional publication models, and the introduction of novel forms of dissemination such as diamond Open Access journals, preprint repositories, and open workflow publishing (Liverpool, 2023). Most recently, cOAlition S has advocated the publish-review-curate (PRC) model that suggests immediate preprint publication of articles followed by community review and curation into thematic areas (Stern *et al.*, 2023).

These new models of academic publication strongly reflect a growing value-driven approach to Open Science that is enshrined in the 2021 UNESCO Recommendation (UNESCO, 2022). This approach foregrounds access to research resources as fundamental to the achievement of the SDGs. The Recommendation positions open publishing in the drive for accelerating the reuse of research outputs and the global pace of research. Within this new approach to Open Science, it can be suggested, the global research community cannot afford not to be open.

*Expanding the “long tail” of research*

The UNESCO endorsement of increased publication transparency comes at a time of considerable digital transformation where the potential of Big Data and AI are starting to become both realised and actualised. Notwithstanding commitments to collective benefit, the contribution of open research objects to novel recombinations and reuse of data is widely recognised and has undoubtedly influenced discussions on publication openness. Making more publications - and corresponding data, methodologies and affiliated information - available, it is thought, can contribute to this data revolution.

Ultimately, such links have shaped the ways that “responsible publishing” is viewed. The process of publishing is increasingly seen to also include pre-registration of projects, pre-print deposition of articles, open workflows, and data sharing. All of these granular elements are linked through their DOIs in open workflows (Foster and Deardorff, 2017).

The value of a granular approach to publishing is thus variously justified as a safeguard of research reproducibility and transparency, a global good and as a resource for driving forward new technologies. While all are undoubtedly true, this does not necessarily mean that new approaches to publishing are unproblematic. Open and granular publishing has the potential to significantly expand what is known as the “long tail” of research. In place of one or two digital objects that traditional publishing would generate, the granular approach yields tens (or many tens) of separate digital objects. The article as a singular entity is thereby transformed into a collection of interlinked documents. This is a particularly significant development in the context of publication cultures and academic career systems that value nominal productivity - in the sense of sheer volume of publications - above else, even as there are critical discussions about a decrease in their relative disruptiveness (Park, Leahey and Funk, 2023).

Between 2014 and 2023, the number of academic articles published annually has risen from 1.95 million to 2.61 million. No such numbers exist for the number of digital objects relating to this published work, ranging from meeting minutes and raw data files to document drafts, pre-prints and presentations. Regardless of how well these objects are linked by graph databases or other means of structuring, this rapid rise in digital objects requires critical scrutiny (White 2019).

*Identifying assumptions about open publishing*

*To c*ritically engage with and advance current discussions around open publishing, let us review a number of foundational assumptions. First, as in many areas of discourse surrounding Open Science, there is the assumption that open resources will be widely acknowledged, surveyed, and reused. Indeed, the UNESCO Recommendation points to “the significant available evidence for the economic benefits and substantial return on investment associated with open science practices and infrastructures, which enable innovation, dynamic research and economic partnerships” (UNESCO, 2022). Nonetheless, whether such reuse occurs remains an item for debate, and considerable further research is required to understand whatdigital objects are being reused and for what purpose.

Second, many commercial publishers are already engaged in discussions around environmental sustainability. In these discussions it is recognised that while “digital publishing has a lower footprint than print, it is not impact-free. Developing, hosting, and using digital content requires computing and server power, and most of this worldwide currently comes from burning fossil fuels” (Barsley, 2022, White, 2023). This recognition is present in a rising number of pledges by commercial digital providers. For instance, a recent pledge by publishers, *Publishing Declares (n.d.),* outlined commitments to offsetting the carbon impact of publication, while Google (2022) outlined commitments towards carbon neutrality. Nonetheless, the transition to action remains unclear, and the balance of centrally coordinated action with individual self-publishing remains debatable.

Third, while data retention/deletion strategies are common in many institutional repositories, the penetration of these policies beyond institutionally managed sites remains unclear. There has been little systematic assessment of the rate of deletion of cloud-stored digital objects outside of institutional settings. In particular, the rapid decrease in cost of cloud storage for individuals (and little training on digital hygiene) means that the number of digital objects associated with research held across the internet is difficult to comprehend.

*The environmental impact of new models of academic publishing*

Open Science discussions on academic publishing are largely silent on issues relating to environmental sustainability and impact. As the cost of digital publishing continues to decrease, discussions in open publishing have retained pre-existing misconceptions, namely that of information as non-material and therefore also devoid of *environmental* impact.

Such approaches contrast to a small, but growing, number of critical voices scrutinising the environmental impact of Big Data and AI. A recent study of data centres and high-performance computing facilities (HPCs) suggest that by 2030 they will account for up to 8% of worldwide emissions (Li *et al.*, 2023). This impact comes not just from power consumptions (and the energy sources used to power systems), but also the manufacturing and packaging of components. Such recognitions highlight a wicked problem, in that Big Data is needed to address the urgent needs of climate change, but at the same time may be contributing to the problems (Nature Computational Science, 2023).

While discussions regarding the carbon footprint of HPCs are timely, they tend to mask discussion of the carbon footprint of smaller digital transactions and objects. A 2020 assessment of global carbon emissions per minute by Monaghan drew attention to the significant environmental impact of daily digital habits. Furthermore, recent studies have highlighted the carbon footprint of the websites themselves (Nitnaware, 2020). While not directly related to academic publishing, the figures presented in figure 1 give rise to pause for thought.

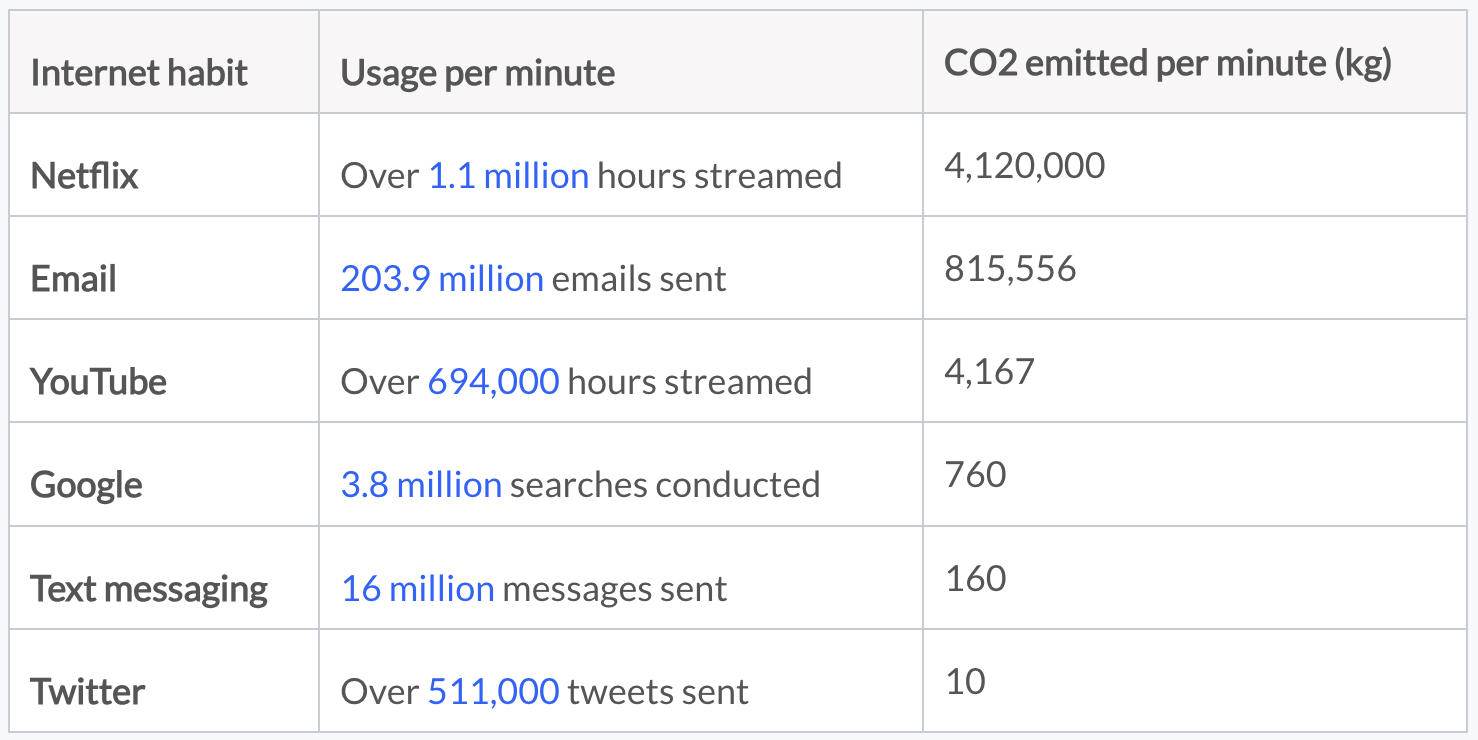


Figure 1: Monaghan’s 2020 estimation of global CO2 emissions per minute from key digital providers

As the number of digital objects associated with a single publication are set to rise exponentially through new publishing models, we must ask what environmental impact is being overlooked. These digital objects - many of which may be duplicates - will be stored on servers that all have a carbon, water and land footprint (European Commission, 2022). Continuing to render this cost invisible to the actors within academia we risk creating problems further down the line. Digital publishing, it must be recognised, is costly and has a significant ecological footprint in the sense that it requires energy and storage.

To demonstrate this, let us consider a collaborative research project involving six researchers from different universities. Together, these researchers have written an academic paper about the project on which they are all authors. Traditionally, this paper (with associated data) would be published in a journal yielding two or three digital objects. In contrast, by embracing Open Access publishing practices, the authors have pre-registered the study, embraced open workflow practices and added their methodology, results and analysis as separate digital objects, and published in a preprint repository prior to submitting to an Open Access journal. In addition, in compliance with their institutional requirements, the authors have independently added the paper to their institutional CRIS systems. Furthermore, many have added the paper to additional platforms such as Zenodo, ResearchGate, Figshare etc. As is evident from these activities, the two or three digital objects from the traditional publishing pathway have been replaced by tens of digital objects that will all require long term storage online.

As it is unlikely that these digital objects will be deleted, the single research output has an increased, persistent environmental impact through the storage required. It is also highly unlikely that these environmental costs will become visible to the individual researchers due to the compartmentalization dynamics through which these costs become invisible. Furthermore, the example highlights tensions within the narrative of Open Science, namely that the dual drivers of transparency and reproducibility strongly support the documentation of the underlying research process.

This commitment to transparency and reproducibility is not only seen as a good in its own right (to increase the reusability of research) but is also promoted to researchers as a benefit of engaging in open research practices (Majhi, Sahu and Behera, 2023). More documentation, it is suggested, leads to increased visibility and thus cite-ability of the publication to colleagues. Clearly, there is a link between the problem of the ecological footprint of digital publishing and a publication culture in which scholarly works are themselves treated as assets whose impact can be managed and optimized.

*Difficult questions for next generation publishing*

Our argument raises the question how we should balance the completeness of documenting the research process with a need to lower the number of digital objects. We propose four starting points for tackling this problem. First, the culture of **rewards and recognition of openness should be balanced by a drive to lower the number of digital objects.** The modern Open Science movement strongly advocates that researchers be rewarded for any activities that increase the accessibility, transparency and equity of their research. Nonetheless, there remains little consensus on how these recognition structures should be set up and monitored. To date, many of the proposals focus on counting digital objects through Open Science dashboards and researcher Open Science profiles (PLOS, 2024). Associating the Open Science movement with strong incentives to increase the number of digital objects associated with research will contribute to the Open Science movement being a driver - rather than a solution - for environmental impact.

Second, **we, as an academic community, should develop timely policies for data deletion**. Discussions within Open Science have long recognised that not all data are equally important. Nonetheless, the advent of the Big Data movement has subtly changed this position through the demonstration of unexpected data reuse. Moreover, the advent of Large Language Models (LLMs) and AI has significantly increased the utility of seemingly less-useful digital objects for model training purposes. Recognition of the environmental impact of these “long tail” digital objects thus places the academic community in a pivotal moment in which the commitment and future value of these LLM and AI tools must be weighed against the potential benefits of data deletion and digital frugality (Jungblut, 2024). While many institutional repositories maintain data deletion policies for their own data collections, the same cannot be said for many generalist repositories, cloud storage facilities and preprint servers. Further discussion is urgently needed.

Third, to mitigate the ecological footprint of digital publishing, **we need to further scrutinise the question of digital sovereignty**. In particular, it becomes important to question whether the *geographic location* of data should retain a position of primacy in these discussions. While all data storage centers are significant consumers of power, water and physical space, it is important to recognise that their environmental impacts vary considerably. In countries where the national power grid continues to rely on coal and gas (European Commission, 2022), or in water-scarce countries (Gonzalez Monserrate, 2022) their environmental impact is significantly larger. The recognition of this variability has led to a number of practical suggestions, such as the relocation of data centers to cold climates to save on power consumption or the development of green data centers that use less energy-intensive technologies (European Commission, 2022).

Discussions surrounding data sovereignty center on the importance of nations’ ability to control and access the data that is generated in their territories and often points to the importance of the geographical location of data. Many low/middle-income countries, however, are in water-scarce regions and do not currently have the resources to invest in green data centers. Thus, while relocating the physical hosting of the data may make sense from an environmental perspective, it directly conflicts with data sovereignty issues whereby the countries/researchers/individuals have the right and ability to control their own digital data. This suggests a need for developing legislation to better protect digital sovereignty, so as to reconcile the pressures of environmental sustainability and sovereignty/right to control.

Fourth, and finally, **we need to develop new collective models of curating academic work that consider the hidden ecological footprint of digital publishing**. There was a time when visibility for research could more easily be created simply through publishing in established journals and participating in conferences. Yet increasing degrees of competition for funding and employment across many research fields, combined with growth and fragmentation into highly differentiated specialties, have made it ever more important for researchers to proactively manage the attention invested in their activities by relevant others. In addition, new models of academic publishing, such as the Publish-Review-Curate model, emphasize the importance of preprint publishing, which would seem to increase the pressure especially for younger and thus less visible researchers to actively publicise their pre-prints and connect their work to other digital objects. Such activities - on social media, through direct communication, and through recorded presentations etc. - all increase the footprint associated with an individual article. Discussions about new publishing models should therefore include strategies for curation of academic publications, to prevent researchers substituting the filtering function of journal prestige with a practice of ensuring the maximal number of digital objects all over the internet.

*Shifting to a dialogue of “mindful publishing”*

The recent developments in academic publishing offer exciting ways through which to increase equitable access to research outputs and to broaden societal engagement. Nonetheless, these developments are not without costs. It is timely to start conversations about the ecological dynamics of novel publishing mechanisms, recognising that it will be human practices and choices that determine the future impact, not simply the design limits of the digital infrastructures. Such conversations are starting, such as in the Climate Neutral Data Centre initiative (Climate Data Neutral Data Center, n. d.), but far more is needed. It is vital that we, as the scholarly community, start to raise awareness, critically scrutinise the infrastructures that we employ and apply pressure to organisations involved in the publishing landscape to decarbonise their supply chains and infrastructures.

**References**

Barsley, Z. C. (2022) *Is publishing sustainable*? Available at: <https://blog.oup.com/2022/11/is-publishing-sustainable/> (Accessed: 10 February 2025).

Climate Neutral Data Center (n.d). Available at: <https://www.climateneutraldatacentre.net/signatories/> (Accessed: 10 February 2025).

European Commission (2022) *Digitalising the energy system - EU action plan*. European Commission. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022DC0552&qid=1666369684560 (Accessed: 25 June 2024).

Foster, E.D. and Deardorff, A. (2017) ‘Open Science Framework (OSF)’, *Journal of the Medical Library Association : JMLA*, 105(2), pp. 203–206. Available at: https://doi.org/10.5195/jmla.2017.88.

Gonzalez Monserrate, S. (2022) ‘The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage’, *MIT Case Studies in Social and Ethical Responsibilities of Computing* [Preprint], (Winter 2022). Available at: https://doi.org/10.21428/2c646de5.031d4553.

Jungblut, S.-I. (2024) ‘Our Digital Carbon Footprint: What’s the Environmental Impact of the Online World?’, *Digital for Good | RESET.ORG*, 30 March. Available at: https://en.reset.org/our-digital-carbon-footprint-environmental-impact-living-life-online-12272019/ (Accessed: 25 June 2024).

Google (2022) *Net-zero carbon*. Available at: https://sustainability.google/operating-sustainably/net-zero-carbon/ (Accessed: 9 February 2025).

Li, B. *et al.* (2023) ‘Toward Sustainable HPC: Carbon Footprint Estimation and Environmental Implications of HPC Systems’, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–15. Available at: https://doi.org/10.1145/3581784.3607035.

Liverpool, L. (2023) ‘Open-access reformers launch next bold publishing plan’, *Nature*, 623(7986), pp. 238–240. Available at: https://doi.org/10.1038/d41586-023-03342-6.

Majhi, S., Sahu, L. and Behera, K. (2023) ‘Practices for enhancing research visibility, citations and impact: review of literature’, *Aslib Journal of Information Management*, 75(6), pp. 1280–1305. Available at: <https://doi.org/10.1108/AJIM-11-2023-532>.

Monaghan, M. (2020). *The world’s digital carbon emissions per minute.* Available at: <https://www.websitebuilderexpert.com/building-websites/global-digital-carbon-emissions/> (Accessed: 10 February 2025).

Nature Computational Science (2023) ‘The carbon footprint of computational research’, *Nature Computational Science*, 3(8), pp. 659–659. Available at: <https://doi.org/10.1038/s43588-023-00506-2>.

# Nitnaware, H. (2020). *Low-carbon websites to cut emissions.* Available at: <https://india.mongabay.com/2020/08/low-carbon-websites-to-cut-emissions/> (Accessed: 10 February 2025).

Park, M., Leahey, E. and Funk, R.J. (2023) ‘Papers and patents are becoming less disruptive over time’, *Nature*, 613(7942), pp. 138–144. Available at: https://doi.org/10.1038/s41586-022-05543-x.

PLOS (2024) ‘Driving Open Science adoption with a global framework: the Open Science Monitoring Initiative’, *The Official PLOS Blog*, 20 May. Available at: https://theplosblog.plos.org/2024/05/driving-open-science-adoption-with-a-global-framewo/ (Accessed: 25 June 2024).

Publishing Declares (n.d.) *PUBLISHING DECLARES CLIMATE ACTION*. Available at: https://publishingdeclares.com/home (Accessed: 8 February 2025).

Stern, B. *et al.* (2023) ‘Towards Responsible Publishing: a proposal from cOAlition S’. Zenodo. Available at: https://zenodo.org/doi/10.5281/zenodo.8398480 (Accessed: 25 June 2024).

UNESCO (2022) *UNESCO Recommendation on Open Science*. Paris: UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000379949>.

## White, K. (2019) *Publications Output: U.S. Trends and International Comparisons*. Washington DC: National Science Foundation.

# White, D. (2023) *Current and Future Trends of the Academic Publishing Industry’s Environmental Effects.* Available at: <https://blog.cabells.com/2023/04/12/current-and-future-trends-of-the-academic-publishing-industrys-environmental-effects/> (Accessed: 10 February 2025).