

# Unconditional data sharing, plus peer review transparency, is key to research reproducibility

Leonid Schneider<sup>1</sup>

<sup>1</sup><https://forbeterscience.wordpress.com>

April 17, 2023



# Unconditional data sharing, plus peer review transparency, is key to research reproducibility

LEONID SCHNEIDER<sup>1</sup>

1. <https://forbetterscience.wordpress.com>

READ REVIEWS

WRITE A REVIEW

CORRESPONDENCE:

[leonid.schneider@gmail.com](mailto:leonid.schneider@gmail.com)

DATE RECEIVED:

April 29, 2016

DOI:

10.15200/winn.146194.42516

ARCHIVED:

April 29, 2016

KEYWORDS:

#LJAFreproducibility, peer review, open access, open science, reproducibility

CITATION:

Leonid Schneider, Unconditional data sharing, plus peer review transparency, is key to research reproducibility, *The Winnower* 3:e146194.42516, 2016, DOI: [10.15200/winn.146194.42516](https://doi.org/10.15200/winn.146194.42516)

© Schneider This article is distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited.



*This is the shorted version of an article I published on 28.04.2016 on my site, [For Better Science](#).*

Open Science is these days largely about **mandatory** publishing in Open Access (OA), regardless of the costs to poorer scientists or the universities which already struggle to pay horrendous subscription fees.

Meanwhile, publishers openly declare that the so-called Gold (author-pays) OA will be much more expensive than even current subscription rates, yet wealthy western institutions like the Dutch university network **VSNU** or the German **Max Planck Society** do not seem troubled by this at all. They seriously expect the publishing oligopoly of Elsevier, SpringerNature and Wiley to lower the costs for Gold OA later on, out of the goodness of their hearts (as this winter's invitation-only **Berlin12 OA conference** suggests).

At the last major Open Science conference in Amsterdam on April 4-5 (**EU2016NL**) the EU Commissioner for Research, Science and Innovation, **Carlos Moedas** and EC Director-General for Research and Innovation, **Robert-Jan Smits**, announced to achieve the flip to **Gold OA by 2020**.

One side effect of the Golden OA flip will be that the scientists in developing countries will not be able to afford publishing in our prestigious OA journals (as the journalist **Richard Poynder** suggested). But of course they will finally be able to read our Western research for free, so I guess this is a fair deal (though I am not sure what benefit there is to finally be able to read for free the certain irreproducible or even manipulated papers in Cell, Nature or Science). In any case, these Eastern European, Latin American, Asian and African academics can still resort to "predatory" OA publishers like OMICS, which are very competitively priced, since they maintain no real editorial supervision or peer review. Maybe this consideration is the reason why predatory OA publishing is repeatedly seen as a non-issue in OA conferences or policies. It was apparently not even mentioned at the EU2016NL meeting.

Peer review transparency, often seen as a key ingredient of Open Science, was obviously not a priority of the EU2016NL conference. This is a pity, because most of the irreproducibility in published research is made possible due to intransparent peer review and lack of open discussion after the paper is published. Scientists, but also editors, abuse their networks and hidden conflicts of interests to covertly help each other placing substandard or unreliable papers in respectable journals, with predictable consequences. At a time, where more and more journals switch to **publishing peer review reports**, sometimes even signed ones, this topic somehow was deemed not important enough at EU2016NL to make it into the 12 goals of **Amsterdam Call for Action on Open Science**.

However, sharing of research data was discussed at length at EU2016NL, even politicians and funders seem to be demanding it, but apparently with certain restrictions, which may clip the wings of open science revolution before it even took off. Calls for data sharing got muddled by vested interests on its way into the 12 goals of EU2016NL. The Amsterdam Call for Action Goal 5 ( [Introduce FAIR and secure data principles](#)) turned out to be more about management of data and restrictive opt-out loopholes like “*legal (privacy) frameworks, and legitimate interests of the parties involved*”. Such “legitimate interests” can easily preclude the sharing of any research data which its authors unilaterally choose to declare as clinically or commercially sensitive.

**Thus, Open Data is about to end up where OA already is: a revolutionary ideal corrupted by grubby business interests combined with academic careerism and dishonesty.**

I believe it is the key to preserving science from the ever-growing threat of collapse of public trust and support. These are the main problems of academic research, and no open access, gold or otherwise, will provide much help in fixing them:

**Irreproducibility.** Its true extent is debatable and probably varies from field to field, but anyone who ever worked in science will know that too often bold claims in impactful publications are not to be entirely trusted. Junior researchers routinely waste months and years (never mind the monetary costs) attempting to reproduce some top-tier published results, only to give up and move on. Sometimes even published reagents are not reliable.

**Data manipulations,** from “minor” offences such as omission of proper controls or contradictory results over “p-hacking” and cherry-picking of “representative” images to wilful image and data forgery. The true extent of this research misconduct epidemic is unknown, but all observers agree what is detected on PubPeer and elsewhere is just a tip of a huge iceberg.

**Publishing intransparency.** Editors (many of whom are active scientists themselves) are known to assign inappropriate peer reviewers, by disregarding blatant conflicts of interests or lack of qualifications. At the same time, reviewers who report suspicious inconsistencies in the manuscripts they evaluate are sometimes overruled by the editors. Finally, whistle-blowers reporting data irregularities or plagiarism in published literature are too often ignored or even met with hostility by the journal’s editors and publishers.

This is why I actually propose to make sharing of research data mandatory, instead of forcing scientists to publish OA, regardless of the costs.

Though I [am fully behind Open Access](#), I do not think that the simple flipping of the current corrupt system of subscription-based publishing to OA is anything worth paying even more public money for. In fact, it will be highly dangerous, by generating a soothing illusion of openness and transparency in science which does not actually exist. Those established academics, publishers and policy makers who benefited from the dishonesty and unaccountability of the current system, are actually the ones who will profit once again from the fake openness façade of the Gold OA.

**Mandating Open Data will actually deliver both- increased reproducibility and accountability in science as well as OA together with reduction of publishing costs. How so?**

There is no logical reason NOT to share published research data. The data-generating researchers will always receive their due credit, in fact they can greatly boost the citation index of their papers by sharing their original data with the community. These counter-arguments, often brought against data sharing (like at EU2016NL), are actually vacuous and misleading:

**Protection of intellectual property against “research parasitism”.** Scientists are expected to share their published reagents with the community. Some do it with Material Transfer Agreements, some deposit their reagents with biobanks which distribute the samples for a fee to anyone who asks, without questioning. The recipients always acknowledge the source and

cite the appropriate paper. Most scientists still happily share their published (and sometimes even unpublished) reagents, and those who don't: there are often good reasons for that. Sometimes authors know that their reagents are not what they were claimed to be, and sometimes these reagents do not exist. Some papers have been retracted due to authors' refusal to share reagents. Therefore, why must reagents be shared, but data has to be kept locked up and shown only to select collaborators?

**Commercial interests.** Patenting of discoveries and technologies must happen before their publication, otherwise it is too late for it. However, pharma companies will be probably reluctant to release their original research data for their competitors to see and use. There is however no point of submitting commercially confidential material for academic peer review anyway. Neither reviewer invited by the journal nor future post-publication peer reviewer should be expected to evaluate any research which original data cannot be made available.

**Patient privacy in clinical research.** This argument is being invoked quite regularly, but can easily be avoided with proper design of patient consent forms, which will imply sharing of anonymised trial data with the research community. With such anonymisation in place, no identifiable patient information will be released and patients' privacy will be assured even when the trial data is distributed to non-collaborating researchers (who might have to sign some kind of data sharing agreement though).

**Publisher's copyright.** Subscription publishers require academic authors to surrender the copyright for their publication. This is what makes Green OA so difficult, because publisher's embargoes do not allow institutional deposition of such publications, at least not until some time has passed. However, while the publishers may obtain the copyright on the final paper, they surely cannot get it for its content and certainly not the original research data. Otherwise, it would be Elsevier patenting all the inventions, and not the researchers who made them. The original research data belongs only to the scientists and their research institutions.

The latter point also implies why mandatory data sharing can succeed where Green OA failed. Scientists are afraid to anger the publishers by infringing on their copyright: after all, they need journals' benevolence when submitting their works for publication.

**With open data, a uniquely bizarre constellation would take place: scientists all over the world will be able to obtain the original data of a paywalled paper of which they can only read the abstract.**

This might be enough for researchers of the same field to procure all the information they need: interpret and re-analyse data, reproduce the results and engage authors into discussion or even collaboration, and all this without the need to actually buy their paper. Scientists, who are proud of their research and who stand behind every bit of their data, will surely not oppose reaching a much wider audience without paying huge sums for OA. But where will it leave the subscription publishing oligarchs? Looking very stupid, that's where. They will have it much tougher to convince university libraries to pay horrendous sums for subscriptions. In fact, they might all by themselves abandon their subscription models and beg for OA negotiations, before universities choose to dismiss their services altogether in favour of alternative publishing models.

**The benefits of data sharing for research reproducibility are obvious. What this approach would need, is a mandate from the side of research institutions, funders and state governments to deposit original data of each and every publication they supported.**

Sharing data is certainly a magnitude cheaper than publishing in Gold OA, and the scientists can retain their autonomy as to which journal they want to submit their works to. They can laudably opt for OA, or decide to publish in some traditional journals. It wouldn't really matter, their original research data would be in any case available for free download to anyone interested. The data repositories should be

best independently operated, publicly or even commercially. Non-complying scientists or those who wilfully submit only unreadable or incomplete data would face the dangers of negative evaluation by their research institutions or funding withdrawal, or even see their publications recommended for retraction.

The peculiarity of this method would be: it leaves the mighty journals and publishers once again out of the loop. The re-evaluation and post-publication peer review of the research they published will happen completely outside of their control. The wider academic community will take charge of the quality control and publish their reports on social networks, personal blogs or, indeed in other peer-reviewed journals. It would be therefore in every journal's best interest to promote editorial and peer review transparency, as well as data sharing, if they wish to avoid being publicly associated with science which others have exposed as faulty or manipulated. The desired effect of peer review openness might come by itself.

When every single paper can be easily scrutinised and re-evaluated, dishonest or negligent scientists would be playing with fire if they were to publish unreliable results. No friendly journal editor could cover up for them, while funders and institutions would have a direct tool at hand to evaluate these scientists' true productivity. It is time for governments and funders to stop listening to peddlers of vested interests and start acting on behalf of science itself.

**Only mandatory Open Data, not Gold Open Access, will lead to more honest and more reproducible science.**