RESEARCH · RESEARCH INTEGRITY

Open Science: Open Data, Open Materials and Open Access

What open science means in practice, covering open data, open materials and code, open access publishing, the FAIR principles, the evidence on benefits and costs, and how organisations can adopt open practices.

ModQu research article · 2,141 words · 22 references (APA 7th edition)

Research has always depended on sharing. Results that cannot be examined, checked or built upon add little to knowledge. Yet for much of the twentieth century, the data behind published studies stayed on researchers' own computers, the methods were described only briefly, and the articles themselves sat behind subscription paywalls. **Open science** is the movement to make the whole research process, not only its final conclusions, available for others to inspect and reuse. This article explains its main elements, the evidence for its benefits, the obstacles it faces and how to put it into practice.

An old idea in a new form

The principle behind open science is old. Merton (1973), describing the norms of modern science, identified **communalism** as one of them: the findings of science belong to the community, and secrecy is the antithesis of this norm. What is new is the scale and the technology. Digital repositories, persistent identifiers and the internet make it possible to share not only articles but data, code, materials and protocols at almost no cost.

UNESCO (2021) adopted a recommendation on open science that defines it broadly, as an inclusive approach combining open scientific knowledge, open infrastructure, open engagement with society and dialogue with other knowledge systems. Most practical discussions focus on three elements: open data, open materials and code, and open access to publications.

Elements of open science

What is shared, and what it makes possible

Element	What is shared	Main benefit
Open data	The data behind the results	Checking and reuse
Open materials and code	Questionnaires, protocols, scripts	Exact repetition
Open access	The published article	Anyone can read it
Preregistration	Hypotheses and analysis plan	Planned vs exploratory is clear
Open peer review	Reviewer reports or names	Accountable review

Drawing on UNESCO (2021) and Munafò et al. (2017) ModQu.com

Figure 1. The main elements of open science. Table by ModQu, drawing on UNESCO (2021) and Munafò et al. (2017).

Open data

Open data means making the data underlying a study available for others to examine and reuse, subject to legal and ethical limits. The case for it is straightforward. Shared data allow others to check that published results follow from the data, to test alternative analyses, to combine data from several studies and to answer new questions without collecting new data.

Before open data became expected, access was often poor. Wicherts et al. (2006) asked the authors of 141 psychology articles for their data, as the journals' policies required, and received data for only about a quarter of them. Vines et al. (2014) requested data from biology papers of different ages and found that the chance of obtaining them fell sharply as articles grew older, mainly because data had been lost or the authors could not be contacted. Data that are not deposited in a public repository at the time of publication are often lost for good.

Researchers' own views have been mixed. Tenopir et al. (2011) surveyed scientists and found broad support for data sharing in principle, alongside concerns about the time required, the lack of funding and infrastructure, and the risk that others would use the data without credit. These concerns have shaped how open data policies are designed.

The FAIR principles

Making data available is not enough if no one can find or understand them. Wilkinson et al. (2016) set out the **FAIR principles**: data should be **Findable**, through persistent identifiers and rich metadata; **Accessible**, through standard protocols; **Interoperable**, using shared formats and vocabularies; and **Reusable**, with clear licences and documentation of how the data were produced. FAIR does not require that all data be open. Sensitive data can be FAIR if their existence and the conditions for access are clearly described, even when access itself is restricted.

Open materials and code

Open materials include questionnaires, stimuli, protocols and instructions, which allow others to repeat a study exactly. **Open code** means sharing the scripts used to clean and analyse data. Because analysis decisions are often too detailed to describe fully in an article, code is frequently the only complete record of how results were produced.

Sharing materials and code supports both reproducibility, re-running the original analysis on the original data, and replicability, repeating the study with new data. It also helps other researchers to avoid mistakes and to build on existing work rather than starting again.

Open access

Open access means that research articles can be read online free of charge. The Budapest Open Access Initiative (2002) gave an influential definition: free availability on the public internet, permitting any user to read, download, copy, distribute, print, search or link to the full texts, with the only constraint on reproduction being the author's right to be properly acknowledged. Suber (2012) described the main routes by which this is achieved.

Piwowar et al. (2018) distinguished several forms of open access, shown below.

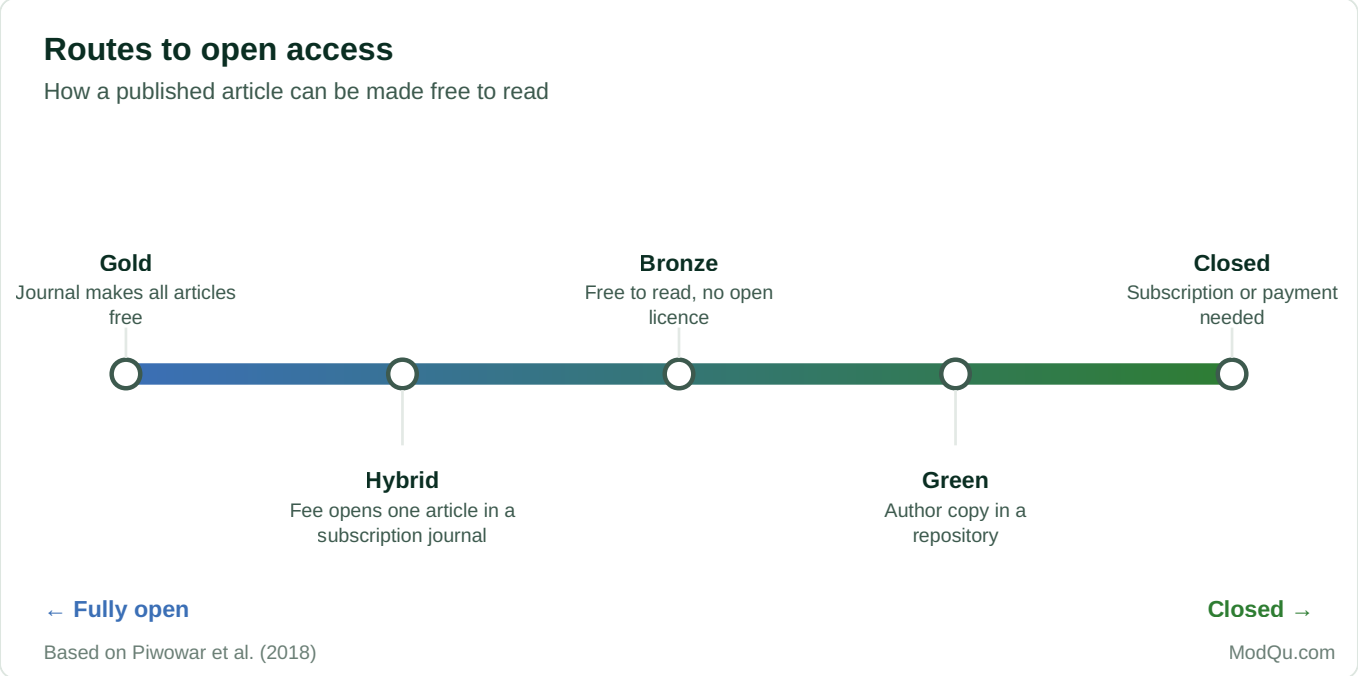


Figure 2. Routes to open access, from fully open journals to closed subscription access. Based on the categories used by Piwowar et al. (2018); diagram by ModQu.

Gold open access means publishing in a journal that makes all its articles free to read, often funded by a fee paid by the author or the author's institution. **Hybrid** means paying to make a single article free in a subscription journal. **Bronze** describes articles that a publisher makes free to read without a clear open licence. **Green** open access means depositing a

version of the article in a repository, such as a university or subject repository, where anyone can read it. Piwowar et al. (2018) estimated that at least 28 per cent of the scholarly literature was available in some open form, with the share higher among recent articles.

The case for open access rests partly on fairness. Much research is publicly funded, yet the public, practitioners, small businesses and researchers in poorer institutions often cannot read the results. Larivière et al. (2015) showed that a small number of commercial publishers had come to control a large share of journal articles, which concentrated the costs of access. Whether open access articles receive more citations has been debated: Eysenbach (2006) found that open access articles in one journal were cited more often than comparable articles that were not open, but later studies have disagreed about how much of the difference reflects access itself rather than authors choosing to make their best work open.

Licences and credit

Making material available is only part of openness. Users also need to know what they are allowed to do with it. Open licences, such as the Creative Commons licences widely used for articles and data, state the terms of reuse in a standard form. The most permissive require only that the original author be credited. Without a licence, legal uncertainty can stop others from reusing material that its authors intended to share.

Credit matters for a related reason. If data and code are to be shared, the people who produce them need recognition. Datasets and software can now be given persistent identifiers and cited in reference lists in the same way as articles, so that their reuse can be tracked and acknowledged.

Open science in practice: a workflow

Openness is easiest when it is planned from the start rather than added at the end. A typical open workflow has several stages:

1. **Plan.** Write a data management plan that says what data will be collected, how they will be stored and documented, and what can be shared.
2. **Preregister.** Record hypotheses, design and analysis plan in a public registry before collecting data.
3. **Document as you go.** Keep a clear record of data cleaning and analysis in scripts rather than manual steps, and describe every variable.
4. **Deposit.** On publication, place data, code and materials in a trusted repository with a persistent identifier and a clear licence.
5. **Publish openly.** Choose an open access route or deposit an accepted version of the article in a repository.
6. **Link everything.** Cite the data and code in the article and link the article from the repository record, so that each can be found from the other.

Following these steps adds some work during a project but saves time later, when results must be checked, questions answered or the analysis repeated by the original team.

Other open practices

Open science also includes practices that make the research process more transparent before and after publication.

Preregistration records hypotheses and analysis plans before data are collected, so that readers can distinguish planned tests from exploratory findings (Nosek et al., 2018). **Registered Reports** go further, with journals reviewing and accepting studies before results are known (Chambers, 2013).

Open peer review can mean several things: publishing reviewers' names, publishing review reports alongside articles, or allowing wider participation in review. Ross-Hellauer (2017) found that the term was used in many inconsistent ways and proposed a set of distinct traits that make up open peer review.

Preprints are versions of articles posted publicly before peer review, allowing rapid sharing and early feedback.

Why open science gained momentum

Open science attracted particular attention after the replication crisis. When large projects found that many published findings could not be repeated (Open Science Collaboration, 2015), it became clear that the lack of access to data, materials and analysis plans made errors and questionable practices hard to detect. Munafò et al. (2017) placed open practices at the centre of their manifesto for reproducible science.

Simmons et al. (2011) had already shown how undisclosed flexibility in collecting and analysing data could make almost any hypothesis appear to be supported. Their proposed remedy was disclosure: if researchers reported every decision, measure and exclusion, readers could judge how much flexibility had been used. Open materials, code and preregistration extend this logic by making the decisions visible as a matter of routine. There is early evidence that such measures work. Scheel et al. (2021) found that Registered Reports in psychology reported support for their first hypothesis far less often than standard articles, which suggests that publishing studies regardless of outcome gives a more accurate picture of the evidence.

Journals and funders responded. Nosek et al. (2015) proposed the **Transparency and Openness Promotion (TOP) guidelines**, a set of standards that journals can adopt at increasing levels of stringency, covering citation of data and code, data and materials transparency, preregistration and replication. Many funders now require data management plans and open access to the publications they pay for.

Benefits for researchers

Open practices are sometimes seen as a burden imposed on researchers, but there is evidence that they also bring benefits. Piwowar et al. (2007) found that clinical trial articles with publicly available data were cited more often than similar articles without them. McKiernan et al. (2016) reviewed the evidence and argued that open practices increase citations, media attention, potential collaborators and funding opportunities. Allen and Mehler (2019) noted that open practices can also benefit early-career researchers by making their work visible and their skills demonstrable, while acknowledging that the costs in time and the risks for those on short contracts are real.

Limits and challenges

Open science is not free of difficulties.

Privacy and consent. Data about people cannot always be shared. Participants may not have consented to sharing, and even data with names removed can sometimes be re-identified. Controlled access, anonymisation and synthetic data are partial solutions.

Commercial confidentiality. Research funded by companies may involve data or methods that cannot be made public without harming the business or breaching contracts.

Costs. Preparing data and code for others takes time and skill. Publishing fees for gold open access can be high and may shift costs from readers to authors, disadvantaging those without funding.

Qualitative research. Interview transcripts and field notes are rich in identifying detail and depend on context. Sharing them raises distinct ethical questions, and some qualitative researchers argue that the reproducibility rationale for open data does not fit their work.

Credit. Researchers are still rewarded mainly for publications. Without credit for shared data and code, the effort of sharing goes largely unrecognised.

These challenges do not argue against openness, but they show why open science is better understood as a principle, "as open as possible, as closed as necessary", than as a single rule.

Open practices in business research

Commercial research cannot always be fully open, but many of the same practices improve its quality. Documenting methods in full, keeping clean and well-described data files, recording analysis code, and setting out the questions and analysis plan before fieldwork all make results easier to check and reuse within an organisation. Clients and readers can trust findings more when they can see how they were produced. Where data can be released, even in summary or anonymised form, doing so allows others to build on the work and demonstrates confidence in it.

Conclusion

Open science applies an old scientific norm with modern tools. Open data, open materials and code, and open access to publications allow results to be checked, combined and built upon, and they make research available to those who fund and use it. The evidence suggests that openness benefits both research and researchers, although it raises real questions about privacy, cost and credit. The practical goal is not openness at any price but transparency wherever it is possible, and clear reasons wherever it is not.

References

Allen, C., & Mehler, D. M. A. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biology, 17*(5), Article e3000246. https://doi.org/10.1371/journal.pbio.3000246

Budapest Open Access Initiative. (2002). *Read the Budapest Open Access Initiative*. https://www.budapestopenaccessinitiative.org/read/

Chambers, C. D. (2013). Registered Reports: A new publishing initiative at *Cortex*. *Cortex, 49*(3), 609–610. https://doi.org/10.1016/j.cortex.2012.12.016

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology, 4*(5), Article e157. https://doi.org/10.1371/journal.pbio.0040157

Larivière, V., Haustein, S., & Mongeon, P. (2015). The oligopoly of academic publishers in the digital era. *PLoS ONE, 10*(6), Article e0127502. https://doi.org/10.1371/journal.pone.0127502

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *eLife, 5*, Article e16800. https://doi.org/10.7554/eLife.16800

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, Article 0021. https://doi.org/10.1038/s41562-016-0021

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article aac4716. https://doi.org/10.1126/science.aac4716

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE, 2*(3), Article e308. https://doi.org/10.1371/journal.pone.0000308

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ, 6*, Article e4375. https://doi.org/10.7717/peerj.4375

Ross-Hellauer, T. (2017). What is open peer review? A systematic review. *F1000Research, 6*, Article 588. https://doi.org/10.12688/f1000research.11369.2

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with Registered Reports. *Advances in Methods and Practices in Psychological Science, 4*(2). https://doi.org/10.1177/25152459211007467

Simmons et al. (2011) had already shown how undisclosed flexibility in collecting and analysing data could make almost any hypothesis appear to be supported. Their proposed remedy was disclosure: if researchers reported every decision, measure and exclusion, readers could judge how much flexibility had been used. Open materials, code and preregistration extend this logic by making the decisions visible as a matter of routine. There is early evidence that such measures work. Scheel et al. (2021) found that Registered Reports in psychology reported support for their first hypothesis far less often than standard articles, which suggests that publishing studies regardless of outcome gives a more accurate picture of the evidence.

Suber, P. (2012). *Open access*. MIT Press. https://doi.org/10.7551/mitpress/9286.001.0001

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE, 6*(6), Article e21101. https://doi.org/10.1371/journal.pone.0021101

UNESCO. (2021). *UNESCO recommendation on open science*. https://doi.org/10.54677/MNMH8546

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J.-S.,
Renaut, S., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology, 24*(1), 94–97. https://doi.org/10.1016/j.cub.2013.11.014

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*(7), 726–728. https://doi.org/10.1037/0003-066X.61.7.726

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*, Article 160018. https://doi.org/10.1038/sdata.2016.18