Science & Society

OPEN ACCESS

EMBO reports

Facilitating open science without sacrificing IP rights

A novel tool for improving replicability of published research

Or Cohen-Sasson[1,*] & Ofer Tur-Sinai[2]

Replicability is a cornerstone of the scientific enterprise. Validating published scientific findings enhances their credibility and helps to build a self-correcting cumulative knowledge base. It also increases public trust in science (Wingen *et al* 2020). Unfortunately, the scientific community has been facing a considerable problem for at least two decades: the replication crisis (Ioannidis, 2005). Scientists in various disciplines have significant difficulties trying to verify published scientific findings (Baker, 2016). One prominent factor accounting for non-replicability is diminished access to research materials required for replication (*replication materials*).

> *"Scientists in various disciplines have significant difficulties trying to verify published scientific findings."*

This problem is particularly noticeable in computational studies: research that utilizes computational models, often with an immense amount of data. With the rise of powerful computers, machine learning and big data, computational studies are increasingly used in a variety of disciplines. This trend is evident in biology as well, including in systems biology, genomics, proteomics, and other areas (Markowetz, 2017). A famous example that demonstrates the importance of computational biology is the Human Genome Project. Developments in computational biology are crucial in advancing promising research prospects in areas such as vaccine antigen design and structural bioinformatics.

> *"The problem of diminished access to replication materials has been reported as a major stumbling block impeding the replicability of computational biology studies."*

A scientific paper alone would not typically enable others to replicate the study described therein (Merali, 2010). Replicating a computational study generally requires access to the code, software documentation, datasets, workflows, and other information regarding the methodology (Easterbrook, 2014). In most cases, however, authors do not publicly share these elements, which renders such studies impossible to replicate (Merali, 2010; Stodden *et al*, 2018). The problem of diminished access to replication materials has been reported as a major stumbling block impeding the replicability of computational biology studies (Crook *et al*, 2013; Miłkowski *et al*, 2018).

IP law and replicability

Various factors contribute to the restricted access to materials: avoiding criticism, fear of falsification and retraction, or a desire to stay ahead of peers. Commercial and proprietary concerns also play a significant role in the decision of scientists and organizations to conceal replication materials (Campbell & Bendavid, 2002; Hong & Walsh, 2009). Such motivations are more prominent as the line between academic and commercially oriented research becomes blurred. Nowadays, commercial firms commonly publish in scientific journals, whereas scientists, universities, and research institutions benefit from the commercialization of research findings and often seek patent protection. All of this cultivates an environment of secrecy, in contrast with the scientific tradition of openness and sharing (Merton, 1942).

> *"... limiting patenting practices might bear a detrimental impact on scientists beyond their commercial interests."*

Intellectual property (IP) law plays a significant role in encouraging this culture of secrecy. Both patent law and trade secret law—main IP regimes—motivate the concealment of certain information to qualify for legal protection. The following paragraphs elaborate on the mechanism by which each regime does so.

Patents are exclusive rights granted for a limited period to inventors in exchange for the disclosure of inventions. One of the conditions for patentability is that the invention be novel when the patent application is filed —that is, it was not publicly disclosed by the applicant or third parties. In addition, the invention must be non-obvious and

1 Zvi Meitar Center for Advanced Legal Studies, Faculty of Law, Tel Aviv University, Tel Aviv, Israel†
2 Ono Academic College, Kiryat Ono, Israel
*Corresponding author. E-mail: cohensasson@mail.tau.ac.il
†Correction added on 5 September 2022, after first online publication: This affiliation has been corrected.
DOI 10.15252/embr.202255841 | EMBO Reports (2022) 23: e55841 | Published online 16 August 2022

reflect more than a trivial step over prior art. A patent application that does not meet any of these conditions would be rejected. The unforgiving nature of the novelty and non-obviousness requirements encourages the concealment of technological information by prospective patent applicants.

A computational study may utilize a software invention that is eligible for patent protection. The study may also lead to the development of a patentable product or process, not in the field of software—for example, a computational immunology study that advances the development of a new vaccine. To maximize chances of obtaining patent protection over such inventions, individual scientists and organizations are likely to minimize public disclosure of any information that they are not obligated to share—including replication materials (Hong & Walsh, 2009).

Trade secrets may lead to a similar outcome. Generally speaking, any piece of information that is not widely known, and from which its holder derives a competitive edge while it is kept secret, can be protected as a trade secret. In connection with a computational study, a researcher or affiliated institution may choose to protect as trade secrets certain information that is not included in the scientific publication. All replication materials could qualify for trade secret protection, including the algorithm itself (if not disclosed in the publication), code, database, workflows, details regarding the computational environment, and any other information pertaining to the study's methodology. Once information is protected as a trade secret, the law prohibits its *misappropriation*. This term encompasses the acquisition, use or disclosure of another's trade secret by improper means, such as theft, or by a person who has confidentiality obligations towards the trade secret holder. Trade secret protection can last infinitely. To enjoy protection, however, the information must remain out of the public domain and its holder must continuously take reasonable measures to keep it confidential.

Given the high societal interest in replications, a natural inclination of policymakers could be to restrict IP rights, to remove incentives to conceal information. In our opinion, though, narrowing IP protection is not imperative to enable replicability. Perhaps, in a utopian world, the scientific enterprise would be free of commercial considerations, which could facilitate information sharing. However, the disposition of individual scientists and research institutions to utilize IP protection for retaining commercial benefits is understandable. Patents also hold an academic value by increasing the reputation of researchers and enhancing funding opportunities for their institutions (Stevens *et al*, 2011). Therefore, limiting patenting practices might bear a detrimental impact on scientists beyond their commercial interests. This is possibly why, despite widespread awareness of the replication crisis, most journals have not adopted a policy mandating the disclosure of replication materials (Rousi & Laakso, 2020).

Conditional access agreement

Instead of choosing between IP rights and replicability, we suggest an inclusive approach that facilitates replications without depriving scientists of IP rights. Our proposal is to implement a new policy tool: the *Conditional Access Agreement* (*CAA*). Recall that it is *public* access to replication materials that jeopardizes both the prospect of securing patent protection (as novelty and non-obviousness are examined vis-à-vis the public prior art) and trade secret protection (since the pertinent information must be kept out of the public domain). Access, however, does not have to be public. This is precisely the gist of the CAA mechanism—establishing a private, controlled channel of communication for the transfer of replication materials between authors and replicators.

The CAA mechanism would work as follows (Fig 1): When submitting a paper for publication, an author would execute an agreement vis-à-vis the journal, pledging to provide full access to replication materials upon demand. The agreement would specify that anyone requesting access to the materials can only obtain it upon signing a non-disclosure agreement (NDA). Under an NDA, the receiving party commits to use the information disclosed by the other party only for a limited purpose while keeping it confidential. In the context of the CAA, the NDA would prohibit the use of the replication materials delivered by the original authors for any purpose other than replication and restrict their further dissemination. Establishing a mechanism of private access to replication materials facilitates, on the one hand, on-demand replicability, while maintaining, on the other hand, the proprietary potential inherent in the study. This mechanism can also alleviate other concerns that a researcher may have with respect to sharing replication materials, including the fear of being scooped or plagiarism.

Admittedly, the CAA is not a risk-free solution. If replicators breach their contractual obligations under the NDA, the replication materials may become public. However,

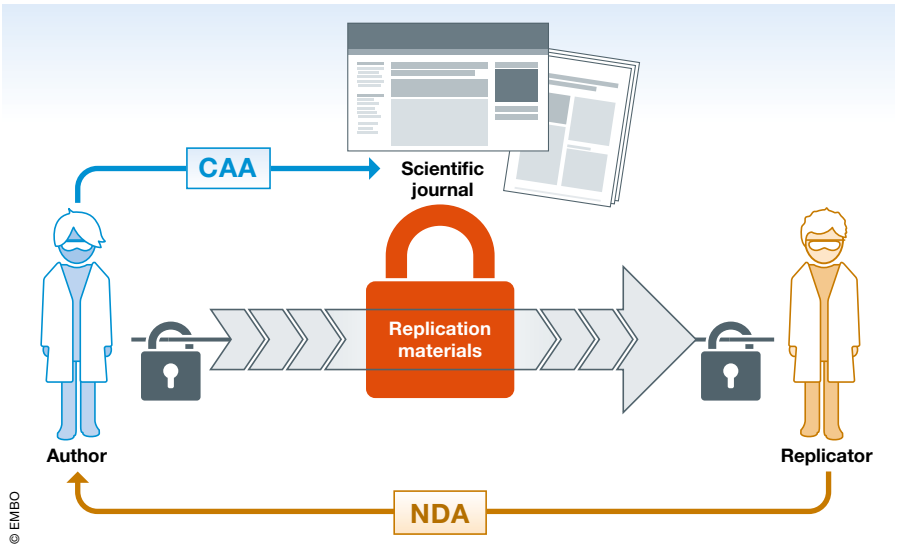

Figure 1. Mechanism of the *Conditional Access Agreement* to share research materials without public disclosure.

even in such a scenario, the CAA mechanism offers some relief. In some cases, a patent application may have already been filed by the time the information leaks. Even if this is not the case, the information disclosed by the replicator will not be considered prior art, provided that the patent application is filed within a specified time period of up to 12 months, depending on the jurisdiction (Merges, 2012). Of course, authors can also file a lawsuit for breach of the NDA and obtain compensation for their damages.

> *"One of the main advantages of the CAA solution is that it imposes minimal costs on replicators, researchers, and journals alike."*

The CAA policy is feasible thanks to an important intermediary: scientific journals. Journals are the gatekeepers of responsible science. As such, journals are continuously involved in a publication's lifecycle from submission to post-publication. In addition to their substantive discretion regarding publications, journals oversee the procedural aspects of the publication process and monitor actions such as uploading the manuscript, declaring conflicts of interests, and signing a copyright-related agreement. As most scientific publications go through the scientific journals portal, it is a convenient path to implement the CAA policy. Notably, mitigating the replication crisis and promoting openness align with journals' mission of promoting rigorous and credible science. Altogether, the CAA policy can be implemented best with the collaboration of scientific journals.

One of the main advantages of the CAA solution is that it imposes minimal costs on replicators, researchers, and journals alike. Replicators will address authors directly—after signing an NDA—to ask for materials without having to pay any fees for accessing the materials. For authors, the CAA would entail minimum transaction costs, as the paper submission process remains almost the same, and the journal will provide the applicable CAA documents. Costs are also low for journals; they only need to formulate a CAA once and embed it as part of their electronic submission platform.

Our proposed solution can enhance replicability and open science without compromising the ability of researchers and their affiliated institutions to benefit from IP rights. Indeed, the CAA mechanism is not a panacea for the replication crisis, which is caused by various factors, not only IP-related ones. Yet, we firmly believe that the CAA policy would reduce the portion of publications that cannot be replicated due to restricted access. Both the scientific community and the public at large would benefit from a greater level of replicability.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

References

Baker M (2016) 1,500 scientists lift the lid on reproducibility. *Nature* 533: 452 – 454

Campbell EG, Bendavid E (2002) Data-sharing and data-withholding in genetics and the life sciences: results of a national survey of technology transfer officers. *J Health Care Law Policy* 6: 241

Crook SM, Davison AP, Plesser HE (2013) Learning from the past: approaches for reproducibility in computational neuroscience. In *20 years of computational neuroscience*, JM Bower (ed), pp 73 – 102. New York, NY: Springer

Easterbrook S (2014) Open code for open science. *Nat Geosci* 7: 779 – 781

Hong W, Walsh JP (2009) For money or glory? Commercialization, competition, and secrecy in the entrepreneurial university. *Sociol Q* 50: 145 – 171

Ioannidis JP (2005) Why most published research findings are false. *PLoS Med* 2: e124

Markowetz F (2017) All biology is computational biology. *PLoS Biol* 15: e2002050

Merali Z (2010) Computational science:. . . error. *Nature* 467: 775 – 777

Merges RP (2012) Priority and novelty under the AIA. *Berkeley Technol Law J* 27: 1023 – 1046

Merton RK (1942) The ethos of science. *J Legal Polit Sociol* 1: 115 – 126

Miłkowski M, Hensel WM, Hohol M (2018) Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail. *J Comput Neurosci* 45: 163 – 172

Rousi AM, Laakso M (2020) Journal research data sharing policies: a study of highly-cited journals in neuroscience, physics, and operations research. *Scientometrics* 124: 131 – 152

Stevens AJ, Johnson GA, Sanberg PR (2011) The role of patents and commercialization in the tenure and promotion process. *Technol Innov* 13: 241 – 248

Stodden V, Krafczyk MS, Bhaskar A (2018) Enabling the verification of computational results: an empirical evaluation of computational reproducibility. In *Proceedings of the first international workshop on practical reproducible evaluation of computer systems, Tempe, AZ*, pp 1 – 5. New York, NY: Association for Computing Machinery

Wingen T, Berkessel JB, Englich B (2020) No replication, no trust? How low replicability influences trust in psychology. *Soc Psychol Personal Sci* 11: 454 – 463